Handbook of Technical Communication

Edited by
Alexander Mehler &
Laurent Romary

In cooperation with
Dafydd Gibbon

De Gruyter Mouton

HAL 8

Handbook of Technical Communication
HAL 8

Handbooks of Applied Linguistics

Communication Competence
Language and Communication Problems
Practical Solutions

H A L

Editors
Karlfried Knapp and Gerd Antos

Volume 8

De Gruyter Mouton

Handbook of Technical Communication

Edited by
Alexander Mehler and Laurent Romary

In cooperation with
Dafydd Gibbon

De Gruyter Mouton

ISBN 978-3-11-018834-9
e-ISBN 978-3-11-022494-8

Library of Congress Cataloging-in-Publication Data

A CIP catalog record for this book has been applied for at the Library of Congress.

Bibliographic information published by the Deutsche Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.dnb.de.

Cover design: Martin Zech, Bremen
Typesetting: Dörlemann Satz GmbH & Co. KG, Lemförde
Printing: Hubert & Co. GmbH & Co. KG, Göttingen
♾ Printed on acid-free paper

Printed in Germany

www.degruyter.com

MIX
Papier aus verantwortungsvollen Quellen
FSC® C016439

Introduction to the handbook series

Linguistics for problem solving

Karlfried Knapp and Gerd Antos

1. Science and application at the turn of the millennium

The distinction between "pure" and "applied" sciences is an old one. According to Meinel (2000), it was introduced by the Swedish chemist Wallerius in 1751, as part of the dispute of that time between the scholastic disciplines and the then emerging epistemic sciences. However, although the concept of "Applied Science" gained currency rapidly since that time, it has remained problematic.

Until recently, the distinction between "pure" and "applied" mirrored the distinction between "theory and "practice". The latter ran all the way through Western history of science since its beginnings in antique times. At first, it was only philosophy that was regarded as a scholarly and, hence, theoretical discipline. Later it was followed by other leading disciplines, as e.g., the sciences. However, as academic disciplines, all of them remained theoretical. In fact, the process of achieving independence of theory was essential for the academic disciplines to become independent from political, religious or other contingencies and to establish themselves at universities and academies. This also implied a process of emancipation from practical concerns – an at times painful development which manifested (and occasionally still manifests) itself in the discrediting of and disdain for practice and practitioners. To some, already the very meaning of the notion "applied" carries a negative connotation, as is suggested by the contrast between the widely used synonym for "theoretical", i.e. "pure" (as used, e.g. in the distinction between "Pure" and "Applied Mathematics") and its natural antonym "impure". On a different level, a lower academic status sometimes is attributed to applied disciplines because of their alleged lack of originality – they are perceived as simply and one-directionally applying insights gained in basic research and watering them down by neglecting the limiting conditions under which these insights were achieved.

Today, however, the academic system is confronted with a new understanding of science. In politics, in society and, above all, in economy a new concept of science has gained acceptance which questions traditional views. In recent philosophy of science, this is labelled as "science under the pressure to succeed" – i.e. as science whose theoretical structure and criteria of evaluation are increasingly conditioned by the pressure of application (Carrier, Stöltzner, and Wette 2004):

> Whenever the public is interested in a particular subject, e.g. when a new disease develops that cannot be cured by conventional medication, the public requests science to provide new insights in this area as quickly as possible. In doing so, the public is less interested in whether these new insights fit seamlessly into an existing theoretical framework, but rather whether they make new methods of treatment and curing possible. (Institut für Wirtschafts- und Technikforschung 2004, our translation).

With most of the practical problems like these, sciences cannot rely on knowledge that is already available, simply because such knowledge does not yet exist. Very often, the problems at hand do not fit neatly into the theoretical framework of one particular "pure science", and there is competition among disciplines with respect to which one provides the best theoretical and methodological resources for potential solutions. And more often than not the problems can be tackled only by adopting an interdisciplinary approach.

As a result, the traditional "Cascade Model", where insights were applied top-down from basic research to practice, no longer works in many cases. Instead, a kind of "application oriented basic research" is needed, where disciplines – conditioned by the pressure of application – take up a certain still diffuse practical issue, define it as a problem against the background of their respective theoretical and methodological paradigms, study this problem and finally develop various application oriented suggestions for solutions. In this sense, applied science, on the one hand, has to be conceived of as a scientific strategy for problem solving – a strategy that starts from mundane practical problems and ultimately aims at solving them. On the other hand, despite the dominance of application that applied sciences are subjected to, as sciences they can do nothing but develop such solutions in a theoretically reflected and methodologically well founded manner. The latter, of course, may lead to the well-known fact that even applied sciences often tend to concentrate on "application oriented basic research" only and thus appear to lose sight of the original practical problem. But despite such shifts in focus: Both the boundaries between disciplines and between pure and applied research are getting more and more blurred.

Today, after the turn of the millennium, it is obvious that sciences are requested to provide more and something different than just theory, basic research or pure knowledge. Rather, sciences are increasingly being regarded as partners in a more comprehensive social and economic context of problem solving and are evaluated against expectations to be practically relevant. This also implies that sciences are expected to be critical, reflecting their impact on society. This new "applied" type of science is confronted with the question: Which role can the sciences play in solving individual, interpersonal, social, intercultural, political or technical problems? This question is typical of a conception of science that was especially developed and propagated by the influential philosopher Sir Karl Popper – a conception that also this handbook series is based on.

2. “Applied Linguistics”: Concepts and controversies

The concept of “Applied Linguistics” is not as old as the notion of “Applied Science”, but it has also been problematical in its relation to theoretical linguistics since its beginning. There seems to be a widespread consensus that the notion “Applied Linguistics” emerged in 1948 with the first issue of the journal *Language Learning* which used this compound in its subtitle *A Quarterly Journal of Applied Linguistics*. This history of its origin certainly explains why even today “Applied Linguistics” still tends to be predominantly associated with foreign language teaching and learning in the Anglophone literature in particular, as can bee seen e.g. from Johnson and Johnson (1998), whose *Encyclopedic Dictionary of Applied Linguistics* is explicitly subtitled *A Handbook for Language Teaching*. However, this theory of origin is historically wrong. As is pointed out by Back (1970), the concept of applying linguistics can be traced back to the early 19th century in Europe, and the very notion “Applied Linguistics” was used in the early 20th already.

2.1. Theoretically Applied vs. Practically Applied Linguistics

As with the relation between “Pure” and “Applied” sciences pointed out above, also with “Applied Linguistics” the first question to be asked is what makes it different from “Pure” or “Theoretical Linguistics”. It is not surprising, then, that the terminologist Back takes this difference as the point of departure for his discussion of what constitutes “Applied Linguistics”. In the light of recent controversies about this concept it is no doubt useful to remind us of his terminological distinctions.

Back (1970) distinguishes between “Theoretical Linguistics” – which aims at achieving knowledge for its own sake, without considering any other value –, “Practice” – i.e. any kind of activity that serves to achieve any purpose in life in the widest sense, apart from the striving for knowledge for its own sake – and “Applied Linguistics”, as a being based on “Theoretical Linguistics” on the one hand and as aiming at usability in “Practice” on the other. In addition, he makes a difference between “Theoretical Applied Linguistics” and “Practical Applied Linguistics”, which is of particular interest here. The former is defined as the use of insights and methods of “Theoretical Linguistics” for gaining knowledge in another, non-linguistic discipline, such as ethnology, sociology, law or literary studies, the latter as the application of insights from linguistics in a practical field related to language, such as language teaching, translation, and the like. For Back, the contribution of applied linguistics is to be seen in the planning of practical action. Language teaching, for example, is practical action done by practitioners, and what applied linguistics can contribute to this is, e.g., to provide contrastive descriptions of the languages involved as a foundation for teaching methods. These contrastive descriptions in turn have to be based on the descriptive methods developed in theoretical linguistics.

However, in the light of the recent epistemological developments outlined above, it may be useful to reinterpret Back's notion of "Theoretically Applied Linguistics". As he himself points out, dealing with practical problems can have repercussions on the development of the theoretical field. Often new approaches, new theoretical concepts and new methods are a prerequisite for dealing with a particular type of practical problems, which may lead to an – at least in the beginning – "application oriented basic research" in applied linguistics itself, which with some justification could also be labelled "theoretically applied", as many such problems require the transgression of disciplinary boundaries. It is not rare that a domain of "Theoretically Applied Linguistics" or "application oriented basic research" takes on a life of its own, and that also something which is labelled as "Applied Linguistics" might in fact be rather remote from the mundane practical problems that originally initiated the respective subject area. But as long as a relation to the original practical problem can be established, it may be justified to count a particular field or discussion as belonging to applied linguistics, even if only "theoretically applied".

2.2. Applied linguistics as a response to structuralism and generativism

As mentioned before, in the Anglophone world in particular the view still appears to be widespread that the primary concerns of the subject area of applied linguistics should be restricted to second language acquisition and language instruction in the first place (see, e.g., Davies 1999 or Schmitt and Celce-Murcia 2002). However, in other parts of the world, and above all in Europe, there has been a development away from aspects of language learning to a wider focus on more general issues of language and communication.

This broadening of scope was in part a reaction to the narrowing down the focus in linguistics that resulted from self-imposed methodological constraints which, as Ehlich (1999) points out, began with Saussurean structuralism and culminated in generative linguistics. For almost three decades since the late 1950s, these developments made "language" in a comprehensive sense, as related to the everyday experience of its users, vanish in favour of an idealised and basically artificial entity. This led in "Core" or theoretical linguistics to a neglect of almost all everyday problems with language and communication encountered by individuals and societies and made it necessary for those interested in socially accountable research into language and communication to draw on a wider range of disciplines, thus giving rise to a flourishing of interdisciplinary areas that have come to be referred to as hyphenated variants of linguistics, such as sociolinguistics, ethnolinguistics, psycholinguistics, conversation analysis, pragmatics, and so on (Davies and Elder 2004).

That these hyphenated variants of linguistics can be said to have originated from dealing with problems may lead to the impression that they fall completely into the scope of applied linguistics. This the more so as their original thematic focus is in line with a frequently quoted definition of applied linguistics as "the theoretical and empirical investigation of real world problems in which language is a central issue" (Brumfit 1997: 93). However, in the recent past much of the work done in these fields has itself been rather "theoretically applied" in the sense introduced above and ultimately even become mainstream in linguistics. Also, in view of the current epistemological developments that see all sciences under the pressure of application, one might even wonder if there is anything distinctive about applied linguistics at all.

Indeed it would be difficult if not impossible to delimit applied linguistics with respect to the practical problems studied and the disciplinary approaches used: Real-world problems with language (to which, for greater clarity, should be added: "with communication") are unlimited in principle. Also, many problems of this kind are unique and require quite different approaches. Some might be tackled successfully by applying already available linguistic theories and methods. Others might require for their solution the development of new methods and even new theories. Following a frequently used distinction first proposed by Widdowson (1980), one might label these approaches as "Linguistics Applied" or "Applied Linguistics". In addition, language is a trans-disciplinary subject par excellence, with the result that problems do not come labelled and may require for their solution the cooperation of various disciplines.

2.3. Conceptualisations and communities

The questions of what should be its reference discipline and which themes, areas of research and sub-disciplines it should deal with, have been discussed constantly and were also the subject of an intensive debate (e.g. Seidlhofer 2003). In the recent past, a number of edited volumes on applied linguistics have appeared which in their respective introductory chapters attempt at giving a definition of "Applied Linguistics". As can be seen from the existence of the Association Internationale de Linguistique Appliquée (AILA) and its numerous national affiliates, from the number of congresses held or books and journals published with the label "Applied Linguistics", applied linguistics appears to be a well-established and flourishing enterprise. Therefore, the collective need felt by authors and editors to introduce their publication with a definition of the subject area it is supposed to be about is astonishing at first sight. Quite obviously, what Ehlich (2006) has termed "the struggle for the object of inquiry" appears to be characteristic of linguistics – both of linguistics at large and applied linguistics. Its seems then, that the meaning and scope of "Applied Linguistics" cannot be taken for granted, and this is why a wide variety of controversial conceptualisations exist.

For example, in addition to the dichotomy mentioned above with respect to whether approaches to applied linguistics should in their theoretical foundations and methods be autonomous from theoretical linguistics or not, and apart from other controversies, there are diverging views on whether applied linguistics is an independent academic discipline (e.g. Kaplan and Grabe 2000) or not (e.g. Davies and Elder 2004), whether its scope should be mainly restricted to language teaching related topics (e.g. Schmitt and Celce-Murcia 2002) or not (e.g. Knapp 2006), or whether applied linguistics is a field of interdisciplinary synthesis where theories with their own integrity develop in close interaction with language users and professionals (e.g. Rampton 1997/2003) or whether this view should be rejected, as a true interdisciplinary approach is ultimately impossible (e.g. Widdowson 2005).

In contrast to such controversies Candlin and Sarangi (2004) point out that applied linguistics should be defined in the first place by the actions of those who practically *do* applied linguistics:

> […] we see no especial purpose in reopening what has become a somewhat sterile debate on what applied linguistics is, or whether it is a distinctive and coherent discipline. […] we see applied linguistics as a many centered and interdisciplinary endeavour whose coherence is achieved in purposeful, mediated action by its practitioners. […]
> What we want to ask of applied linguistics is less what it is and more what it does, or rather what its practitioners do. (Candlin/Sarangi 2004:1–2)

Against this background, they see applied linguistics as less characterised by its thematic scope – which indeed is hard to delimit – but rather by the two aspects of "relevance" and "reflexivity". Relevance refers to the purpose applied linguistic activities have for the targeted audience and to the degree that these activities in their collaborative practices meet the background and needs of those addressed – which, as matter of comprehensibility, also includes taking their conceptual and language level into account. Reflexivity means the contextualisation of the intellectual principles and practices, which is at the core of what characterises a professional community, and which is achieved by asking leading questions like "What kinds of purposes underlie what is done?", "Who is involved in their determination?", "By whom, and in what ways, is their achievement appraised?", "Who owns the outcomes?".

We agree with these authors that applied linguistics in dealing with real world problems is determined by disciplinary givens – such as e.g. theories, methods or standards of linguistics or any other discipline – but that it is determined at least as much by the social and situational givens of the practices of life. These do not only include the concrete practical problems themselves but also the theoretical and methodological standards of cooperating experts from other disciplines, as well as the conceptual and practical standards of the practitioners who are confronted with the practical problems in the first place. Thus, as Sarangi and van Leeuwen (2003) point out, applied linguists have to become part of the respective "community of practice".

If, however, applied linguists have to regard themselves as part of a community of practice, it is obvious that it is the entire community which determines what the respective subject matter is that the applied linguist deals with and how. In particular, it is the respective community of practice which determines which problems of the practitioners have to be considered. The consequence of this is that applied linguistics can be understood from very comprehensive to very specific, depending on what kind of problems are considered relevant by the respective community. Of course, following this participative understanding of applied linguistics also has consequences for the Handbooks of Applied Linguistics both with respect to the subjects covered and the way they are theoretically and practically treated.

3. Applied linguistics for problem solving

Against this background, it seems reasonable not to define applied linguistics as an autonomous discipline or even only to delimit it by specifying a set of subjects it is supposed to study and typical disciplinary approaches it should use. Rather, in line with the collaborative and participatory perspective of the communities of practice applied linguists are involved in, this handbook series is based on the assumption that applied linguistics is a specific, problem-oriented way of "doing linguistics" related to the real-life world. In other words: applied linguistics is conceived of here as "linguistics for problem solving".

To outline what we think is distinctive about this area of inquiry: Entirely in line with Popper's conception of science, we take it that applied linguistics starts from the assumption of an imperfect world in the areas of language and communication. This means, firstly, that linguistic and communicative competence in individuals, like other forms of human knowledge, is fragmentary and defective – if it exists at all. To express it more pointedly: Human linguistic and communicative behaviour is not "perfect". And on a different level, this imperfection also applies to the use and status of language and communication in and among groups or societies.

Secondly, we take it that applied linguists are convinced that the imperfection both of individual linguistic and communicative behaviour and language based relations between groups and societies can be clarified, understood and to some extent resolved by their intervention, e.g. by means of education, training or consultancy.

Thirdly, we take it that applied linguistics proceeds by a specific mode of inquiry in that it mediates between the way language and communication is expertly studied in the linguistic disciplines and the way it is directly experienced in different domains of use. This implies that applied linguists are able to demonstrate that their findings – be they of a "Linguistics Applied" or "Applied Linguistics" nature – are not just "application oriented basic research" but can be made relevant to the real-life world.

Fourthly, we take it that applied linguistics is socially accountable. To the extent that the imperfections initiating applied linguistic activity involve both social actors and social structures, we take it that applied linguistics has to be critical and reflexive with respect to the results of its suggestions and solutions.

These assumptions yield the following questions which at the same time define objectives for applied linguistics:

1. Which linguistic problems are typical of which areas of language competence and language use?
2. How can linguistics define and describe these problems?
3. How can linguistics suggest, develop, or achieve solutions of these problems?
4. Which solutions result in which improvements in speakers' linguistic and communicative abilities or in the use and status of languages in and between groups?
5. What are additional effects of the linguistic intervention?

4. Objectives of this handbook series

These questions also determine the objectives of this book series. However, in view of the present boom in handbooks of linguistics and applied linguistics, one should ask what is specific about this series of nine thematically different volumes.

To begin with, it is important to emphasise what it is not aiming at:

– The handbook series does not want to take a snapshot view or even a "hit list" of fashionable topics, theories, debates or fields of study.
– Nor does it aim at a comprehensive coverage of linguistics because some selectivity with regard to the subject areas is both inevitable in a book series of this kind and part of its specific profile.

Instead, the book series will try

– to show that applied linguistics can offer a comprehensive, trustworthy and scientifically well-founded understanding of a wide range of problems,
– to show that applied linguistics can provide or develop instruments for solving new, still unpredictable problems,

– to show that applied linguistics is not confined to a restricted number of topics such as, e.g. foreign language learning, but that it successfully deals with a wide range of both everyday problems and areas of linguistics,
– to provide a state-of-the-art description of applied linguistics against the background of the ability of this area of academic inquiry to provide descriptions, analyses, explanations and, if possible, solutions of everyday problems. On the one hand, this criterion is the link to trans-disciplinary co-operation. On the other, it is crucial in assessing to what extent linguistics can in fact be made relevant.

In short, it is by no means the intention of this series to duplicate the present state of knowledge about linguistics as represented in other publications with the supposed aim of providing a comprehensive survey. Rather, the intention is to present the knowledge available in applied linguistics today firstly from an explicitly problem solving perspective and secondly, in a non-technical, easily comprehensible way. Also it is intended with this publication to build bridges to neighbouring disciplines and to critically discuss which impact the solutions discussed do in fact have on practice. This is particularly necessary in areas like language teaching and learning – where for years there has been a tendency to fashionable solutions without sufficient consideration of their actual impact on the reality in schools.

5. Criteria for the selection of topics

Based on the arguments outlined above, the handbook series has the following structure: Findings and applications of linguistics will be presented in concentric circles, as it were, starting out from the communication competence of the individual, proceeding via aspects of interpersonal and inter-group communication to technical communication and, ultimately, to the more general level of society. Thus, the topics of the nine volumes are as follows:

1. Handbook of Individual Communication Competence
2. Handbook of Interpersonal Communication
3. Handbook of Communication in Organisations and Professions
4. Handbook of Communication in the Public Sphere
5. Handbook of Multilingualism and Multilingual Communication
6. Handbook of Foreign Language Communication and Learning
7. Handbook of Intercultural Communication
8. Handbook of Technical Communication
9. Handbook of Language and Communication: Diversity and Change

This thematic structure can be said to follow the sequence of experience with problems related to language and communication a human passes through in the course of his or her personal biographical development. This is why the topic areas of applied linguistics are structured here in ever-increasing concentric circles: in line with biographical development, the first circle starts with the communicative competence of the individual and also includes interpersonal communication as belonging to a person's private sphere. The second circle proceeds to the everyday environment and includes the professional and public sphere. The third circle extends to the experience of foreign languages and cultures, which at least in officially monolingual societies, is not made by everybody and if so, only later in life. Technical communication as the fourth circle is even more exclusive and restricted to a more special professional clientele. The final volume extends this process to focus on more general, supra-individual national and international issues.

For almost all of these topics, there already exist introductions, handbooks or other types of survey literature. However, what makes the present volumes unique is their explicit claim to focus on topics in language and communication as areas of everyday problems and their emphasis on pointing out the relevance of applied linguistics in dealing with them.

Bibliography

Back, Otto
1970 Was bedeutet und was bezeichnet der Begriff 'angewandte Sprachwissenschaft'? *Die Sprache* 16: 21–53.

Brumfit, Christopher
1997 How applied linguistics is the same as any other science. *International Journal of Applied Linguistics* 7(1): 86–94.

Candlin, Chris N. and Srikant Sarangi
2004 Making applied linguistics matter. *Journal of Applied Linguistics* 1(1): 1–8.

Carrier, Michael, Martin Stöltzner, and Jeanette Wette
2004 *Theorienstruktur und Beurteilungsmaßstäbe unter den Bedingungen der Anwendungsdominanz.* Universität Bielefeld: Institut für Wissenschafts- und Technikforschung [http://www.uni-bielefeld.de/iwt/projekte/wissen/anwendungsdominanz.html, accessed Jan 5, 2007].

Davies, Alan
1999 *Introduction to Applied Linguistics. From Practice to Theory.* Edinburgh: Edinburgh University Press.

Davies, Alan and Catherine Elder
2004 General introduction – Applied linguistics: Subject to discipline? In: Alan Davies and Catherine Elder (eds.), *The Handbook of Applied Linguistics*, 1–16. Malden etc.: Blackwell.

Ehlich, Konrad
1999 Vom Nutzen der „Funktionalen Pragmatik“ für die angewandte Linguistik. In: Michael Becker-Mrotzek und Christine Doppler (eds.), *Medium Sprache im Beruf. Eine Aufgabe für* die *Linguistik*, 23–36. Tübingen: Narr.

Ehlich, Konrad
2006 Mehrsprachigkeit für Europa – öffentliches Schweigen, linguistische Distanzen. In: Sergio Cigada, Jean-Francois de Pietro, Daniel Elmiger, and Markus Nussbaumer (eds.), *Öffentliche Sprachdebatten – linguistische Positionen. Bulletin Suisse de Linguistique Appliquée/VALS-ASLA-Bulletin* 83/1: 11–28.

Grabe, William
2002 Applied linguistics: An emerging discipline for the twenty-first century. In: Robert B. Kaplan (ed.), *The Oxford Handbook of Applied Linguistics*, 3–12. Oxford: Oxford University Press.

Johnson, Keith and Helen Johnson (eds.)
1998 *Encyclopedic Dictionary of Applied Linguistics. A Handbook for Language Teaching.* Oxford: Blackwell.

Kaplan, Robert B. and William Grabe
2000 Applied linguistics and the Annual Review of Applied Linguistics. In: W. Grabe (ed.), *Applied Linguistics as an Emerging Discipline. Annual Review of Applied Linguistics* 20: 3–17.

Knapp, Karlfried
2006 Vorwort. In: Karlfried Knapp, Gerd Antos, Michael Becker-Mrotzek, Arnulf Deppermann, Susanne Göpferich, Joachim Gabowski, Michael Klemm und Claudia Villiger (eds.), *Angewandte Linguistik. Ein Lehrbuch*. 2nd ed., xix–xxiii. Tübingen: Francke – UTB.

Meinel, Christoph
2000 Reine und angewandte Wissenschaft. In: *Das Magazin*. Ed. Wissenschaftszentrum Nordrhein-Westfalen 11(1): 10–11.

Rampton, Ben
1997 [2003] Retuning in applied linguistics. *International Journal of Applied Linguistics* 7 (1): 3–25, quoted from Seidlhofer (2003), 273–295.

Sarangi, Srikant and Theo van Leeuwen
2003 Applied linguistics and communities of practice: Gaining communality or losing disciplinary autonomy? In: Srikant Sarangi and Theo van Leeuwen (eds.), *Applied Linguistics and Communities of Practice*, 1–8. London: Continuum.

Schmitt, Norbert and Marianne Celce-Murcia
2002 An overview of applied linguistics. In: Norbert Schmitt (ed.), *An Introduction to Applied Linguistics.* London: Arnold.

Seidlhofer, Barbara (ed.)
2003 *Controversies in Applied Linguistics.* Oxford: Oxford University Press.

Widdowson, Henry
1984 [1980] Model and fictions. In: Henry Widdowson (1984) *Explorations in Applied Linguistics 2*, 21–27. Oxford: Oxford University Press.

Widdowson, Henry
2005 Applied linguistics, interdisciplinarity, and disparate realities. In: Paul Bruthiaux, Dwight Atkinson, William G. Egginton, William Grabe, and Vaidehi Ramanathan (eds.), *Directions in Applied Linguistics. Essays in Honor of Robert B. Kaplan*, 12–25. Clevedon: Multilingual Matters.

Contents

Introduction: Framing Technical Communication

Alexander Mehler, Laurent Romary, and Dafydd Gibbon

1. On the technical side of communication

The *Handbook of Technical Communication* brings together a selection of topics that range from technical media in human communication to the linguistic, multimodal enhancement of present-day technologies. It covers computer-mediated text, voice and multimedia communication. In doing so, the handbook takes professional and private communication into account. Special emphasis is put on multimodal communication, its semiotic underpinning, standardized representation, automatic processing and evaluation in the process of system development. Throughout the handbook, this is primarily done from a technological point of view that puts emphasis on the underlying processing technologies as well as on standards and formats of data representation in technical communication. Thus, basic communication technologies and infrastructures play a major role in this handbook. In summary, the handbook deals with theoretical issues of technical communication, technology-enhanced communication and its practical impact on the development and usage of text and speech as well as of multimodal technologies.

In order to give a blueprint of how the various fields of research on technical communication collected by this handbook hang together, we start from a notion of technical communication that integrates the notions of *mediality, codality, modality, temporality* and *social complexity* (see Figure 1). More specifically, we refer to five constituents in our notion of technical communication as defined below:

1. *Mediality in the range of mono- and multimediality:* with the rise of hypertext and hypermedia (Bush, 1945; Nelson, 1995; Engelbart, 1995; Berners-Lee, Hendler, and Lassila, 2001), multimedia has been a longstanding topic in computer science and related disciplines (only recently being complemented or even replaced by the topic of multimodality (Gibbon, Mertins, and Moore, 2000)). Generally speaking, the term medium is used in two respects: as a substance of information transmission and as a device of information storage (cf. Hess-Lüttich and Schmauks, 2004).[1] However, what is stored is designated to be received in the future and, thus, storage media can be seen to serve as transmission media where the border between both readings parallels the distinction between ephemeral and persistent (synchronous, quasi-synchronous or asynchronous) transmission. An asynchronous, non-face-to-face transmission necessarily requires the use of a

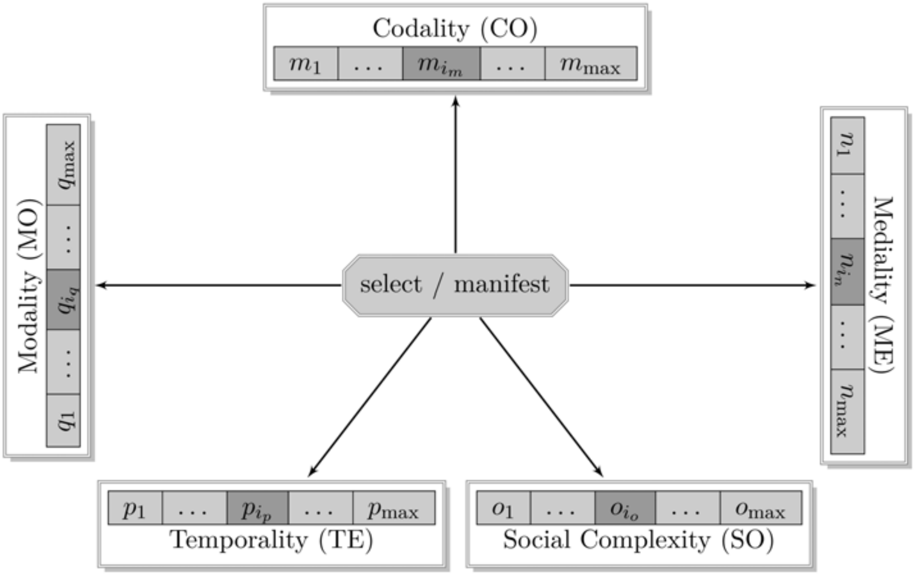


Figure 1. A five dimensional decision matrix together with a candidate manifestation in the form of the vector ($CO = m_{i_m}$, $ME = n_{i_n}$, $SO = o_{i_o}$, $TE = p_{i_p}$, $MO = q_{i_q}$). As an example consider wiki-based technical communication that is characterized by the signature m_{i_m} = multicodal, n_{i_n} = multimedial, o_{i_o} = *n:m*, p_{i_p} = {quasi-synchronous, asynchronous}, n_{i_n} = {monomodal production, multimodal reception}.

technical device that renders the respective communication *technical*. In any event, storage media do not necessarily determine the mode by which the information being transferred is processed, nor do they necessarily require a single code for encoding or decoding, respectively. Thus, we have to distinguish two more reference points.

2. *Codality in the range of mono- and multicodality:* traditionally, linguistics has been concerned with monocodal signs that are encoded according to a single underlying language (or code). However, not only with the advent of web-based communication and globalization we face multilingual documents that are obviously encoded by means of different codes and, thus, are multicodal. By the same token, a crossmodal supersign (Kendon, 2004) or a so called multimodal ensemble (Lücking, Mehler, and Menke, 2008) as, for example, a supersign that consists of a linguistic sign and a gesture, is multicodal (though not multilingual) where the underlying codes require at least two modes for producing and receiving its constituents. Thus, multicodal signs of this sort inevitably fall into the realm of multimodality. Another example of distinguishing mono- and multicodal signs relates to the auto-

matic annotation of documents based on information extraction technologies (Wilks, 1997; Feldman and Sanger, 2007) that are used, for example, for automatic forecasting and algorithmic trading (Dang, Brabazon, Edelman, and O'Neill, 2009). In these cases, an algorithm based on some language model implements an additional code that, if being successful, is used to enrich the input documents informationally. It may be a question at issue whether annotated documents are multicodal or just based on multiple authors that are partially artificial. In any event, one has to consider that algorithms do not necessarily use the same code as human authors even if the output of the algorithms is input to algorithmic trading and, finally, to decision making by humans.

3. *Modality in the range of mono- and multimodality:* the modality of a communication concerns the sensory interfaces that the communicators use to perform this communication. In this sense, a communication is called multimodal, if at least one communicator uses at least two different sensory interfaces for sign production or reception.[2] One may argue that any communication is multimodal since one never uses the same interface for both sign production and reception. However, if we make an additional distinction it is evident that monomodal *production* and monomodal *reception* are quite common.

   Note that we call a *communication* (i) mono- or multimodal, (ii) mono- or multimedial, (iii) mono- or multicodal where the reference point of such attributions is (i) the sign producer or recipient, (ii) the transmission or storage medium being used and (iii) the sign system(s) being manifested (cf. Weidenmann, 1997). Note also that we may speak of multimodal, multimedial or multicodal *signs*: on the token level this means to select one of the latter reference points in order to stress the singularity or multiplicity of the codes, media or modes manifested by the tokens under consideration.

The three reference points of classifying communications introduced so far are, so to speak, orthogonal. This is demonstrated, for example, by multimedial documents that consist of graphically manifested texts and sound tracks that can be received in a monomodal fashion if the texts are read aloud by means of some speech synthesis. Further, a multimodal discourse in which gesture accompanies speech can be digitized and transferred by means of a single medium that captures both the visual and acoustical data stream. Thirdly, a multilingual instruction manual can be monomedially stored and monomodally received. It is evident how we may extend this list of examples that keep the communicators, transmission media and codes apart as different reference points of classifying communications. We continue this classification by means of two additional reference points:

4. *Temporality in the range of synchronicity, quasi-synchronicity, and asynchronicity:* none of the reference points distinguished so far *explicitly* refers to the notion of time as being constitutive for any act of communication. However, in reviewing the classical model of conceptual speech and writing as introduced by Koch and Oesterreicher (cf. Koch and Oesterreicher, 2001), Dürscheid (2003) stresses the need to distinguish what she calls mediality, that is, what we call temporality. The reason for this notional add-on is to distinguish communication situations along the synchronicity or asynchronicity of turn-taking. Along this line of reasoning, Dürscheid distinguishes between *synchronous* communications (of immediate or even overlapping turn-taking in shared communication situations), *quasi-synchronous* communications (in which turn-taking is marginally stretched so that feedback among communicators is slightly restricted), and *asynchronous* communications (where the communication situation is significantly stretched so that a common frame of communication can hardly be assumed). Below we shall see how this criterion of temporality combines with those of mediality, modality and codality. However, before we do that, we need to make a final distinction.
5. *Social complexity in the range of monades, dyads, triads, and large social groups (aka communities):* Traditionally, linguistics had to do with written documents that were mainly produced for mass communication (Wehner, 1997) where single text producers addressed large social groups of communicators (Brinker, Antos, Heinemann, and Sager, 2000). During the 1970's, dialogical communication came into the fore of linguistic research, consequently focusing on dyads of interlocutors (Brinker and Sager, 2006). Even more recently, linguistics started looking at triads or tiny groups of persons that together perform so called multilogues (Ginzburg and Fernandez, 2005). Thus, we observe a rise in social complexity that starts from monologues and dialogues to finally reach the level of so called multilogues. What is missed in these approaches is a more comprehensive account of the social complexity of communication in terms of one-to-one, one-to-many and many-to-many relations as they are prototypically observed in face-to-face communication (1:1), mass communication (1:*n*), and web-based collaboration (*n:m*). Any account of technical communication has to consider this social dimension (cf. Zerubavel, 2003) – otherwise it fails to clarify the peculiarities of technological enhancements of, say, collaborative writing as currently provided by wiki-based systems (Leuf and Cunningham, 2001; Stegbauer, 2009). With the rise of web-based communication, we observe an increase of the social complexity of communication that has to be reflected by any account of technical communication.

One may ask why such a high-dimensional, parameter-rich space is needed to clarify the notion of technical communication. The reason is that technological means of communication have pervaded human communication, modifying it in various, but insufficiently classified, ways. By addressing one of these five reference points we can classify different approaches to technical communication by their semiotic *differentia specifica*. In order to see this, we consider three combinations of these reference points that are accompanied by specific technologies that render the corresponding communication *technical* in a very specific sense:

1. *The case of collaborative writing:* wikis (e.g., Wikipedia) allow for a form of collaborative writing that uses distributed computer-based media for producing, storing, transferring and receiving the contributions of their "wikilocutors". From the point of view of reception, this sort of collaborative writing is primarily monomodal since contributions of wikilocutors are mainly textual. Exceptions occur, for example, in the form of sound tracks or when speech synthesis is used for reading written contributions aloud so that in principle we face a situation of multimodal reception. Further, in large Wikipedias (e.g., the English or German ones) that contain citations in many languages, multicodality is the norm. In terms of temporality, wikis allow for quasi-synchronous as well as asynchronous communication: in the former case an author's contribution may be removed or further processed quickly by her collaborators, while in the latter case a textual contribution may stay unchanged for years before being modified by other wikilocutors. This temporal peculiarity of wikis relates to their dual nature as storage and transmission media (as, e.g., in the case of highly disputed discussions (Wattenberg, Viégas, and Hollenbach, 2007; Laniado, Tasso, Volkovich, and Kaltenbrunner, 2011; Gómez, Kappen, and Kaltenbrunner, 2011)). However, what makes wikis an exceptional case of technical communication is their *social complexity*, which ranges from manifesting mass communication (where articles are authored by single authors, but received by large audiences (Mehler and Sutter, 2008)) to collaborative writing and edit wars (where articles are authored by many authors in the role of content providers, removers or defenders (Brandes, Kenis, Lerner, and Raaij, 2009)). Here, technology plays the role of enabling a sort of collaborative writing where, in principle, communicators can work on the same texts at different points in time at different locations – irrespective of how distant they are in terms of space and time. That is, wiki technologies are not just a means for externalizing our social memory of texts and their intertextual relations (Sager, 1997). They also allow for suspending the time-space continuity of face-to-face communication (Zerubavel, 2003) in terms of multilogues. They do so by enabling a multitude of communicators to adopt both the role

of text recipients and text producers concerning the same texts irrespective of their time-space discontinuity. Thus, we may speak of an externalization of collaborative sign processing in social groups that is enabled by wiki technologies and that extends immediate peer-to-peer learning (Rieger, 2003) to *mediate collaborative learning in groups*. This example shows that beyond the trias of codality, mediality and modality, temporality and social complexity are indispensable reference points of characterizing the peculiarity of technical communication as enabled by wikis. Of course, collaborative story-telling and social memorizing existed before the advent of wikis; however, wikis allow for externalizing these processes in terms of quasi-synchronous and asynchronous collaborative writing. Obviously, this is a step forward in the development of technical communication.

2. *The case of ubiquitous computing:* The term 'ubiquitous computing' means, literally, 'computing anywhere', sometimes also 'computing everywhere' or 'omnipresent computing' (see Chapter 21, 'Ubiquitous Computing' for a comprehensive overview of this topic). The field of ubiquitous computing arose with the emergence of small mobile computing devices such as laptop computers, PDAs (Personal Digital Assistants) and multifunction cellphones in the 1990s. In the meantime the field has developed into a very broad spectrum of interactive wireless systems with internet-connected smartphones, cameras and audio players, GPS-based location coordinating software applications, automatic RFID (Radio Frequency Identification) shopping checkout, and industrial, military and domestic 'smart environment' control by means of proximity detection of moving agents. Each of these different kinds of ubiquitous computing uses a different specification of the 5-dimensional decision space of classifying technical communication as presented in Figure 1. For example, collaborative game players using smartphones will typically work with the following (simplified) specification: they preferably communicate synchronously (temporal complexity) (of course, smartphones also allow for quasi-synchronous and even asynchronous communication) using touch-activated screens (multimediality) in order to produce and process signs in different (audio, visual and tactile) modes (multimodality) in a competitive, at least dialogical (1:1) or even multilogical (*n:m*) manner (social complexity). What is characteristic for this example of ubiquitous computing is the enhancement that it provides for *multimodal computing* especially in situations of quasi-synchronous communication. In these situations, the space continuity of the interlocutors is suspended. Ubiquitous computing implies more than just particular technologies, however. In the technical coding dimension, ubiquitous computing is dependent on standardization of interoperable interfaces and inter-device as well as human-machine communication protocols (see the discussion of aircraft and train control in Chapter 12, 'Verbal Communication Proto-

cols in Safety-Critical System Operations'). Further, in the social dimension, the 'invisible' or 'implicit' identification procedures which are often involved in ubiquitous communication raise problematic ethical issues of technology-driven changes in privacy concepts, which are the topic of continuing controversy. In any event, ubiquitous computing stresses the multimodal variety of communication thereby opening up many new areas of application of technical communication.

3. *The case of adaptable hypermedia:* The third example covers so called adaptable hypermedia (Brusilovsky, 1996).[3] These have been proposed as a class of dynamic hypertexts that allow for their continual adaptation subject to their usage. An example of such a system has been given by O'Donnell, Mellish, Oberlander, and Knott (2001) who argue for hypermedia systems that adapt even the wording of single hypertext modules (Storrer, 2002) to the needs, usage habits or interests of their recipients (so that, for example, beginners are provided text that is accessible at a high level of readability). As a result, different hypertexts are generated for different readers that may reflect the social, cognitive or situational specifics of their usage – with respect to form *and* content. Even though this is a historical example (since this technology has been replaced by a completely different development of the web), it is worth considering the kind of complexity of this sort of technical communication according to Figure 1. Evidently, adaptable hypermedia are multimedial and require multimodal reception by definition. However, they are simple as far as their social complexity is concerned: although single users can influence the generation of hypertexts being presented to them, this mainly works in terms of a filter mechanism for which the system specifies the space of possible states it can take up in advance. As a consequence, an adaptable system of this sort hardly enhances communication *among* different users. If at all, this relates to modeling types of users (e.g., beginners, experts etc.) and, thus, to a sort of typological interrelation of users who are supposed to "interact" in similar ways with the system (Mehler, 2009). From this perspective, it is no surprise that the web took a different development toward collaborative media as exemplified by Wikipedia. In any event, there are recent developments that aim at unifying the semantic web (Berners-Lee, Hendler, and Lassila, 2001) (see also Waltinger and Breuing (2012) in this volume) and wikis (Leuf and Cunningham, 2001) in terms of so called *semantic wikis* in order to allow for a kind of collaborative writing under the control of a semantic model as provided by semantic web technologies (Schaffert, Bischof, Bürger, Gruber, Hilzensauer, and Schaffert, 2006; Völkel, Krötzsch, Vrandecic, Haller, and Studer, 2006; Hammwöhner, 2007; Baer, Erbs, Zesch, and Gurevych, 2011). This may end up in a revival of notions of adaptive hypermedia under the header of the semantic web and Web 2.0 technologies.

These three examples stress different reference points of the complexity of technical communication as shown in Figure 1: wikis point to the social complexity of collaborative writing while ubiquitous computing relates to the complexity of multimodal interfaces. Finally, adaptive hypermedia and semantic wikis emphasize the potential for development of technical writing even in settings with a focus on classical, monomedial texts. Any such development raises the question for the combinability of manifestations along the reference points distinguished in Figure 1. Equivalently, any such development challenges the limits of multimodal fusion and fission as well as of the codal, medial, temporal and social complexity of communication. The present handbook's chapters provide insights into such a drawing of boundaries.

Note that technologies are a means to resolving constraints of our embodiment (cf. Eco's notion of prostheses in Eco, 1988). Technologies are also a prerequisite for making available even the ephemeral traces that communication leaves behind – regarding any of the reference points considered so far. That is, without technical media for transferring or storing these traces, research, for example, on the social complexity of collaborative writing, on the conversational unfolding of microblogging or on the alignment (Pickering and Garrod, 2004; Rickheit, 2005) of interlocutors in multimodal face-to-face communication (Ruiter, Rossignol, Vuurpijl, Cunningham, and Levelt, 2003) would hardly be possible. Because technical communication is an indispensable channel still under development, it is now possible to study communication to its full semiotic extent – as described and foreseen in the comprehensive handbook on semiotics edited by Posner, Robering, and Seboek (1996).

In order to systematize our examples just given look again at Figure 1. It shows the reference points of classifying communication situations as considered so far. Our basic idea is to call a communication *technical* if any of the attribute values selected along these reference points is affected by some technology (e.g., enhanced, replaced, complemented, augmented, accelerated, controlled, managed or otherwise modified).[4] That is, the technologies in use may primarily focus on enhancing the reception or the production of signs (as in the case of barrier-free communication – see Kubina and Lücking (2012) in this handbook) (i.e., the aspect of modality). Alternatively, they may focus on enhancing the storage and transmission of data (as in the case of semantic information retrieval – see Paaß (2012) in this handbook) (i.e., the aspect of mediality in Figure 1). Further, technologies may essentially address the problem of handling documents that contain signs of multiple languages or to adopt monocodal texts to the needs of different language communities (as in the area of multilingual computing – see Sasaki and Lommel (2012) in this volume) (i.e., the aspect of codality). Further, we have to think of digital curation (see Lee (2012) in this handbook) that seeks technologies that allow for using present-day digital artifacts (e.g., computer games, wikis, or blogs) by future generations – even under comparable conditions subject to which they are currently used (i.e., the aspect of temporality). Obviously, this is an extreme example of the temporal complexity of technical communication, where the addressee of the communicative act is yet not alive. Finally, the main focus of technologies in technical communication may relate to the social complexity of communication (as in decentralized social networks and wiki media – see Diewald (2012) and Waltinger and Breuing (2012) both in this volume) (i.e., the aspect of social complexity).

Of course, technologies may provide enhancements along different dimensions at the same time and hardly realize a single such focus. Such a multiple classification is reflected in Figure 1 by example of wikis that manifest a certain selection of attribute values along our list of five reference points (see above). This results in a signature of the peculiarity of wikis in terms of technical communication. The idea is to perform the same classificatory procedure for any other sort of technical communication. As this is an open development, no one can foresee what combinations will become relevant and what further reference points will be needed to account for the rapid enhancements and modifications of our every-day technical communication. In any event, this handbook intends to provide a snapshot of this ongoing development. It collects contributions to the following areas of technical communication:

- technical foundations of technical communication
- standardization
- linguistic computing
- multilingual computing
- multimodal computing
- talking computers (speech recognition, speech synthesis)
- language and speaker identification
- artificial interactivity
- barrier-free communication
- assistive technologies for motor and sensory communication impairments
- distributed social networks
- peer-to-peer communication

As any non-wiki-based handbook necessarily makes a selection of certain topics, we hope the reader will find our selections rich and thought-provoking, even if we have not included every relevant theme. We believe in collaborative writing and ask the reader to follow the references given at the end of each chapter and to consult Wikipedia as a continually enhanced encyclopedia that because of its temporal and social scale can better document recent developments in the rapidly changing field of technical communication. In any event, the present handbook may help to get first insights and references for some of the main fields of research in this area.

2. Overview of the book

The handbook on technical communication is divided into three parts and 22 chapters. While Part I focuses on technologies, the management of communication and its technology-based enhancement are dealt with in Part II and III respectively.

Part I starts with overviewing basic technologies that are used in various fields of technological communication. This includes technologies for document representation and document analysis as well as technologies for handling multimodal data and speech data in technical systems that accompany, augment or even enhance oral and other kinds of face-to-face communication. Part I distinguishes between the generation (Chapter 1), the representation (Chapter 2, 3 and 4), the classification (Chapter 5) as well as the management and retrieval of mono- and multimodal documents (Chapter 5 and Chapter 6). It mirrors the two-fold nature of data in technical communication, which, historically, started with text-based systems in order to enter the full range of multimodal communication between humans and their artificial companions. Special emphasis is put in Part I on resources – such as corpora and machine readable dictionaries – for training, testing and evaluating technical communication systems (Chapter 7). The thorough evaluation of these systems by means of automatic, semi-automatic or purely intellectual procedures is dealt with by a special chapter (Chapter 8). Part I is the most technical part of the handbook: some of its chapters can be read as thorough introductions to the respective field of research with a focus on representation formats or on state-of-the-art technologies for processing corresponding data. All other chapters of the handbook take profit of these introductions by referring to these selected chapters as references for the respective technologies. As it stands, no handbook can give comprehensive overviews of *all* relevant fields of research or application. In any event, the interested reader is asked to consult the technical literature cited in these chapters to get further information.

Part II of the handbook covers seven chapters that focus on the management of technical communication in different application areas. These areas range from lexicography (including terminology and ontology management) via multilingual communication, and scientific communication to the usage of controlled languages in risk and security management. Ontologies have been invented as representation formats that enable automatic, semantic (inference-driven) data processing. Hence, ontologies play an important role in any area of technical communication that has to do with processing linguistic data (Chapter 9). In any event, processing linguistic data has to face the variety of languages due to the globalization of technical communication. Thus, Part 2 puts special emphasis on multilingual computing (Chapter 9). This includes the localization of technical systems, that is, procedures by which content and software systems are provided in multiple languages. Beyond multilinguality, multimodality is one of the areas of current development of technical communication whose semiotic underpinning is described in Chapter 13. This chapter complements the more technological basing point of multimodality as provided by Chapter 6 in that it looks at multimodal communication strategies in human-machine communication. Chapter 11, Chapter 12 and Chapter 14 give overviews of three areas of the management of technical communication. This tripartite overview starts with scholarly communication (Chapter 11) with a special emphasis on digital libraries and current developments in scientific communication. It is complemented by an example-driven survey of technical communication in safety-critical systems (Chapter 12). With the advent of the web, ever new technical media for communication arise. Most recently, this includes so called decentralized online social networks. The handbook reflects this very important and still upcoming area of technical communication by means of Chapter 14. Last but not least, technical media can be seen as devices not only for externalizing the human memory, but also for archiving semiotic data for future generations (see above). Chapter 15 is about this aspect of technical "communication" between the past, the present and the future. It does so from the point of view of digital curation, that is, the technical management of artifacts for future use.

Part III complements the handbook by surveying 7 areas of technology-enhanced communication. The first of these chapters focuses on Internet-based communication and Web 2.0 technologies together with their impact on technical communication and documentation (Chapter 16). Technology-enhanced learning and especially computer-assisted language learning is the topic of Chapter 17. A central aspect of technology-enhanced communication relates to the communication of human beings with artificial agents. This is the topic of Chapter 18 and 20. Both of these chapters continue the handbook's subtopic *multimodal communication*, however, with an emphasis on multimodal fusion (in the course of understanding multimodal data) and multimodal fission (in the course of producing multimodal data). A classical area of technology-enhanced communication is barrier-free communication that is dealt with by Chapter 19. Starting from the comprehensive overview on multimodal communication given in this handbook, this chapter distinguishes handicaps and impairments of human beings together with corresponding technologies for bridging them in communication. Two novel developmental branches of technical communication are given with ubiquitous computing (Chapter 21) and P2P-computing (Chapter 22). Whereas the former accounts for the pervasion of computing technologies in everyday activities, the latter utilizes the notion of agent or peer networks for the distributed processing of data, say, in information retrieval. The handbook ends with these two application areas. They exemplify upcoming areas of human communication that once started from non-technology-enhanced face-to-face communication in order to reach the area of cloud computing, artificial interactivity and related fields of application where even machines can serve as turn-takers. Obviously, this is an ongoing development whose destination is not foreseeable.

In the following sections, the 22 chapters of this handbook are summarized in detail.

2.1. Part I: basic communication technologies & infrastructures

In **Chapter 1 ("Document authoring: from word processing to text generation")**, Aurélien Max presents the various methods that have been progressively developed to accompany the authoring of text in technical domains. Articulating the topics around the document creation workflow, the chapter tackles all aspects of document control, from purely formatting (or structural) aspects to the linguistic checking of content. A last section addresses the issue of natural language generation whether fully automated or interactive. Together with Chapter 2, 3, 4 and 5 of this handbook, this chapter is mainly about written communication with a focus on monomedial documents.

Chapter 2 by Sebastian Rahtz **("Representation of Documents in Technical Documentation")** covers the various aspects of the representation of documents in digital formats. Starting with an analysis of the various levels at which a document can be construed, and identifying the correlation between the syntactic (and layout) representation of a document with its underlying semantic, Sebastian Rahtz surveys the various existing formats and standards currently deployed in the industrial and the academic communities (TeX and XML based formats such as the HTML, TEI, DocBook, OpenXML or ODF).

Complementing the preceding chapter, **Chapter 3 ("Foundations of markup languages")** by Maik Stührenberg provides an in-depth presentation of XML technology, as used in a variety of representational contexts. Starting with the notion of a semi-structured document and eliciting the class of digital objects that can be thus represented, the author shows in detail the theoretical and technical background of the XML recommendation, addressing in particular issues related to expressivity (overlapping, pointing) and parsing (schema languages).

In **Chapter 4 ("Controlled language structures in technical communication")**, Thorsten Trippel reflects on the various domains where controlled languages maybe used to enforce unambiguous meaning in technical communication. After an overview of the traditional domain, e.g. technical writing, where controlled languages have been in use for years, the author covers more recent aspects where digital content management requires the handling of controlled languages to ascertain the semantics of digital content. Covering issues related to controlled vocabularies, metadata languages and data categories for data modeling, the chapter provides a systematic link with the corresponding standardization activities. Although this chapter is mainly about written communication, controlled languages are also more and more important for enhancing multimodal communication by technical means, a topic that is thoroughly treated by several chapters in this handbook.

In **Chapter 5 ("Document Classification, Information Retrieval, Text & Web Mining")**, Gerhard Paaß surveys state-of-the-art technologies of data mining and information retrieval with a focus on textual units. The chapter deals with quantitative models of documents in written communication and their utilization for various tasks such as text classification, topic detection and tracking, topic clustering and labeling, word sense disambiguation, co-reference resolution, relation extraction and semantic retrieval. The latter technologies are called *semantic* as they do not require that search queries and retrieved documents share search terms; rather, they are required to contain semantically related or at least similar terms where semantic similarity is computed in terms of a distributional approach (Harris, 1954; Landauer and Dumais, 1997; Turney and Pantel, 2010). In such a way, the chapter touches a deep problem in technology-enhanced communication where the technology needs to "understand" the semiotic output of human beings in order to better serve their needs. As it stands, monomodal, monomedial written communication (including multicodal documents) will be the first area of real progress in solving this task. Consequently, the chapter includes the rising area of ontology mining. In so doing, it provides a reference for the interested reader for any related chapter on lexical resources, controlled languages, terminologies and lexicography within this handbook.[5] A central part of the chapter relates to the distinction of supervised and unsupervised machine learning. In this regard, the chapter surveys standard methods in text and web mining. Starting from the automatic preprocessing of documents, this includes, amongst others, clustering methods, kernel methods and conditional random fields that are used, for example, for automatically annotating text constituents (e.g., as named entities) in written communication. This sort of semantic annotation is an indispensable step towards advanced technologies, for example, in technical documentation, multilingual computing or, more generally, in computational humanities. It is called semantic as the information being annotated concerns the semantic interpretation of text constituents. Advancements in technical communication will critically depend on achievements in automatizing such interpretations that heretofore constituted the domain of human expertise.[6]

The topic of multimodal communication, i.e. communication with more than one human output-input channel such as speech and gesture, is pervasive in the handbook, and addressed in **Chapter 6 ("Multimodal and Speech Technology")** from the technological perspective by Jean-Claude Martin and Tanja Schultz, combining the issue of multimodality with that of multilinguality. Language specific problems such as typological differences in sound and writing systems and word types, and technological problems such as the lack of data and missing orthographic standards are discussed and technical strategies for solving these problems are presented. Spoken dialogue systems are addressed in detail, as well as new areas such as the recognition of emotion, personality, workload stress and multimodal speaker recognition and tracking. Finally, approaches to and applications of multimodal interfaces are discussed.

Technical communication systems require resources for development in the form of raw data, systematized data, tools for data processing and standardized archiving. In **Chapter 7 ("Resources for Technical Communication Systems")**, Dafydd Gibbon defines technical communication systems as devices or device networks which intervene in the communication channel between speaker/writer and addressee. The chapter also details resource problems from the point of view of standardization requirements, from names for languages to *intellectual property rights* (IPR). Typical technical communication systems are listed, as well as the kinds of data required for their development and intellectual resources such as notations and transcription systems and storage formats. Two areas are selected for detailed exemplification: resources for lexicon development and resources for developing spoken language systems and multimodal systems.

Part I is concluded by **Chapter 8 ("Evaluation of Technical Communication")** by Peter Menke. It provides an overview of related topics and methods with a focus on technical communication. In this way, Chapter 8 can be read as a complement to any of the handbook's chapters (in any of the areas distinguished by Figure 1) since evaluation is indispensable whenever it comes to assessing the quality of technical enhancements of communication. Starting from a notion of communication that distinguishes agents, medium and content as reference points of its technical enhancement, this chapter provides a typology of different objects of evaluation in technical communication. Further, with the help of model theory, the chapter describes evaluation as a measurement operation based on which notions (like objectivity, reliability and validity) can be explained (for more details on these topics from the point of view of linguistics see Altmann and Grotjahn, 1988; Altmann, 1993). This is the background of the third part of the chapter, where qualitative and quantitative methods of evaluation are described. On the one hand, this includes questionnaires, interviews and expert reviews, and on the other hand, statistical measures. Special emphasis is given to measures of effectiveness as used in automatic classification (cf. Chapter 5). The chapter concludes with classification scenarios showing that more often than not evaluation is rather an ongoing process than a point measurement. Since the transparency and acceptability of technically enhanced communication become more and more important not least with the advent of decentralized social networks (cf. Chapter 14), evaluation will have a strong impact on future developments in any of the areas dealt with in this handbook.

2.2. Part II: technical communication management

Serge Sharoff and Anthony Hartley address in **Chapter 9 ("Lexicography, terminology and ontologies")** the creation and representation of lexical structures based upon a semasiological (concept to term) principle. Starting with the basic concepts issued from the semiotic domain and the early work of E. Wüster, they provide an overview of a) the characteristics of semiological structures seen as ontological systems and b) of the various technological answers to the representation of such data, in particular within ISO committee TC 37 (with TMF and TBX). In the last section of the chapter, the authors present the state of the art in terminology extraction from text and show the importance of corpora as a basis for the construction of termbanks.

The issue of multilingualism (i.e., multicodality in terms of Figure 1) in scientific communication is presented in **Chapter 10 ("Multilingual Computing")**, where Felix Sasaki and Arle Lommel cover the various technologies, together with the corresponding international standards that have accompanied the development of multilingual computing over the years. From character encoding (ISO 10646 and Unicode) to the various methods of representing multilingual information and segments (terminologies with ISO 16642 and TBX, translation memories with TMX, localization documents with XLIFF), the authors clarify how these methods compare with the actual processes that are required when dealing with multilingual content.

Chapter 11 by Laurent Romary **("Scholarly Communication")** tackles the role of scholarly publication in the research process and looks at the consequences of new information technologies in the organization of the scholarly communication ecology. Covering issues related to document acquisition and open access models, the chapter broadens to include research data, and ends up by presenting a possible scenario for future scientific communication based on virtual research environments. In this sense, the chapter addresses both the social complexity of collaborative scientific communication and the dissemination and retrieval of multimedia scientific data.

Central issues for technical communication systems concern the issues of fitness for purpose, including reliability, risk assessment and safety in deployment environments. In **Chapter 12 ("Verbal Communication Protocols in Safety-Critical System Operations")**, Peter Ladkin treats the specific domain of verbal communication protocols in this context, with particular reference to communication involving decisions taken before airline crashes, and to train control protocols. For each of these domains case studies are provided. In the aircraft domain, communication preceding an aircraft collision (the 2002 Überlingen mid-air collision), involving conflicting decision criteria from humans and technical systems and a case where technical information was interpreted incorrectly by the crew (Cali) are discussed in detail. In the train domain, the logic of two-way train operation on single lines and the consequences of failure to follow this (Warngau) are analyzed. Solution strategies for handling these communication issues are discussed in terms of formal grammars for protocol implementation. Safety-critical situations as considered in Chapter 12 are extremely important also from the point of view of multimodal communication – this topic bridges to the next chapter.

In **Chapter 13 (“Multimodal Communication”)**, the topic of multimodal communication is addressed from a linguistic point of view by Jens Allwood and Elisabeth Ahlsén, starting with a clarification of the production, reception and function of multimodal signs in interactive contexts. Differences between human-human and human-machine communication are discussed in terms of the relative flexibility of human adaptation and the relative inflexibility of machine adaptation, cultural variation and the complex multiple channels of human communication, including gesture, posture, gaze and facial expressions. Communication control and monitoring issues such as turn management are outlined, and examples of machines involved in multimodal man-machine communication are given, ranging from computers and smartphones to robot interaction and tutoring systems. Finally, the ethical issue of apparent naturalness in machine communication systems leading to over-confidence in the abilities of the system, and responsibility for consequences of actions of the system are emphasized.

In **Chapter 14 (“Decentralized Online Social Networks”)**, Nils Diewald deals with *Social Network Sites* (SNS) as one of the most recent developments of online technical communication with a great influence on social inclusion and exclusion. SNSs provide a technology to build *Online Social Networks* (OSN). These are social networks that are spanned among agents according to their communication relations as manifested by the underlying SNS. In this sense, we may speak of a social community as an OSN that is basically networked by means of the communication acts of its members as performed, for example, with the help of an SNS such as Facebook. From this point of view, Chapter 14 focuses on the social complexity of (quasi-synchronous or even asynchronous) technical communication that may include cases as simple as 1:1 communication up to the level of situations in which many members of the same community react upon the tweets or posts of many other members of the same community. However, since this sort of communication happens online, aspects of multimediality are also at the core of this chapter. The chapter is divided into three parts. It starts with discussing sociological aspects of SNSs and OSNs. This mainly includes quantitative models based on social network analysis (Wasserman and Faust, 1999; Newman, 2010). Note that this section also includes a discussion of commercial aspects of SNSs. The second section is devoted to the architecture of SNSs and its impact on issues of (de-)centralization, participation, privacy and legacy of communication via SNSs. The chapter continues with a description of technologies that underly SNSs mainly in terms of protocols and formats used to technically manifest online communication via SNSs. A last section is devoted to recent projects on SNSs that certainly will attract more and more attention with the rise of SNSs and OSNs as scientific objects of communication theory.

With **Chapter 15 ("Digital curation as communication mediation")**, Christopher A. Lee introduces the specific aspect of the management of digital content from the point of view of maintenance and preservation. Digital curation, seen as the selection, representation and storing of digital traces, is contemplated here from both communicative and technical activity. On the basis of the OSI (ISO 1994) model, the author covers the various levels of representation that have to be considered in digital curation, ranging from bit stream processes up to complex clusters of digital objects. Interestingly, digital curation addresses the preservation of ultimately any aspect of the complexity of technical communication as distinguished by Figure 1. Therefore, it can be seen as an outstanding research area of technical communication.[7]

2.3. Part III: communication by means of technology

Part III starts with **Chapter 16 ("Internet-based Communication")** by Ulli Waltinger and Alexa Breuing. It does so with an emphasis on semantic web standards and technologies (e.g., RDF, OWL) and the Web 2.0. This also relates to the notion of crowdsourcing (according to which efforts in data annotation and semantic tagging are delegated to web communities) that may include the integration of automatic and human computation (e.g., in the framework of games with a purpose (Ahn, 2008)). An example of crowdsourcing is collaborative writing as not only exemplified by Wikipedia (Leuf and Cunningham, 2001; Stegbauer, 2009) but also by numerous wikis that serve, for example, scientific communication, leisure communication or enterprize communication. Starting from an overview of the proliferation of technologies of web-based communication (including, amongst others, e-mail-based communication, instant messaging, and voice over IP), the chapter discusses several challenges of this field of research. This includes the notion of the web of data (cf. the *linked data paradigm* (Chiarcos, Nordhoff, and Hellmann, 2012)) as well as models of social and emergent semantics (Steels and Hanappe, 2006; Gabrilovich and Markovitch, 2009; Egozi, Markovitch, and Gabrilovich, 2011). In this sense, Chapter 16 can be read as an overview of the development of social complexity in web-based communication starting from mass communication (where users tend to be passive recipients of static web content), and advancing now to the area of true *n:m*-communication as exemplified by collaborative writing and microblogging (where users perform both roles as authors and recipients).

Chapter 17 ("Tutoring Systems and Computer-Assisted Language Learning (CALL)") by Henning Lobin and Dietmar Rösler deals with applying technical communication as a means to assist and enhance processes of learning. It starts with enumerating aspects under which language learning (as a special sort of communication between teacher and students) is altered due to the use of assistive computer-based systems. From the point of view of Figure 1, this mainly relates to the (multi-)mediality and (multi-)modality of communication processes. The chapter also bridges to the social complexity of communication as computer-assisted learning aims at autonomous learners who communicate not only more effectively with their teachers but also with each other. This is exemplified by peer-assisted learning, cooperative learning and learning in tandem that all are enhanced by technical means. The chapter provides an overview of the development of this field of research as well as a thorough exemplification in terms of computer-assisted language learning. The second part of the chapter is devoted to data driven learning and teaching thereby bridging to corpus-based approaches. It concludes with an overview of intelligent systems of CALL that integrate this area of application with intelligent systems as studied and built in computer science.

The topic of multimodal technical communication recurs in **Chapter 18 ("Framing Multimodal Technical Communication")**, in which Andy Lücking and Thies Pfeiffer concentrate on complexity issues in communication with speech and gesture, and the problem of integrating information from these two channels. Different forms of non-verbal behaviour are addressed in detail and the development of studies in this field is surveyed. The theoretical semiotic background is addressed in terms of sign types such as icon, index and symbol, and the interaction of modalities with media is discussed. A detailed overview of applications of technical communication is given, focussing specifically on devices for communicating gestural movement, such as eye-trackers and tracking gloves. Finally, issues of integration or fusion of multimodal data from different channels, as well as the fission or splitting of communication channels are discussed from technical and logical points of view.

Barrier-free communication is a field which has been seen as a small niche mainly for the sight and hearing impaired and those with serious illnesses and injuries, but in recent years the field has entered the mainstream of technical communication with human-machine interfaces. Petra Kubina and Andy Lücking discuss the state of the art in this area in **Chapter 19 ("Barrier-free Communication")**, pointing out areas such as the personal element of barriers to communication, with technical solutions for augmentative and alternative communication as well as social barriers to communication. Contexts for barrier-free communication such as the World-Wide Web are characterized in terms of perceivability, operability, understandability and robustness, and markup systems for structuring barrier-free communication data in the conventional web, the semantic web and the social web are discussed, including special issues like PDF readers. Finally a wide range of research directions, applications and tools are discussed, including examples such as the handheld Simputer developed in Bangalore for pre-literate communities, with icon input and speech synthesis output. Special techniques are also outlined. This includes tools for sonification and haptic input, through the treatment of sign languages to cochlear implants.

Recent developments in multimodal technical communication are treated by Stefan Kopp and Ipke Wachsmuth in **Chapter 20 ("Artificial Interactivity")**. Constraints on communication with embodied agents and avatars, also related to issues featured in Chapter 13 such as turn handling, timing and feedback are discussed in the context of interaction with a virtual human performing a construction task with instruction from a human interlocutor. The different levels of processing and integration of forms and meanings in multimodal behavior are discussed, and the specific issue of generating multimodal behavior is handled in detail, with special reference to the architecture of behavior production, from sensory data through feature extraction, segmentation and pattern recognition to interpretation of the behavior and its integration into existing knowledge representations.

Chapter 21 ("Ubiquitous Computing") by Codrina Lauth, Bettina Berendt, Bastian Pfleging & Albrecht Schmidt extends the book's perspective on multimedial and multimodal technical communication. It deals with one of the most recent developments in this field of research that relates to the pervasiveness of computing in physically situated areas of human communication – that is, not in the virtual world of web-based communication (cf. Chapter 16) or, more specifically, social network sites (cf. Chapter 14). Instead, this chapter is about technical communication in situations of everyday communication in which visual, acoustic, haptic or even olfactoric channels are in use as transmission media. In principle, ubiquitous computing challenges the classical notion of human-computer interaction in that sensors can make the body as a whole an interface in technically enhanced everyday actions at any point in time at any location. This opens technical communication to new areas of application that have been widely closed for traditional approaches to text-based communication. With this development, we may enthusiastically state that we are at the beginning of a new area of technical communication (as explained and exemplified in this chapter) that covers the whole branch of semiosis, that is, of human sign formation based on whatever modality in whatever living conditions. In any event, any such assessment has to keep track with the current stage of development of these technologies. Chapter 21 does exactly this: it provides a definition of ubiquitous computing in conjunction with an overview of its current trends, exemplifications by means of short case studies as well as an overview of its underlying technologies and future challenges. In this way, it can be read as an entry point to one of the major developmental branches of technical communication.

The last chapter of Part III and also the last chapter of this handbook is **Chapter 22 ("P2P-based Communication")** by Gerhard Heyer, Florian Holz & Sven Teresniak. It extends our series on future developments in technical communication. On the one hand, the chapter focuses on one of the traditional application areas of technical communication, that is, information retrieval (Baeza-Yates and Ribeiro-Neto, 1999; Melucci, Baeza-Yates, and Croft, 2011). However, it does so from the point of view of recent advancements in the theory of complex networks (Steyvers and Tenenbaum, 2005). The basic idea of this field of research is to profit from the peculiarities of social and semiotic networks in which neighbors tend to be linked according to similar or related interests, meanings or functions. Using this model, information retrieval in P2P networks means giving up the notion of a centralized retrieval engine in order to rely instead on a network of semantically interlinked peers (software agents) whose networking is optimized to serve the retrieval needs. From this point of view, Chapter 22 mainly contributes to the social complexity of technically enhanced communication, however, it does so with a focus on monomodal written communication. Chapter 22 closes the handbook's thematic circle as it deals with an area of application that, at the beginning of this handbook, is described in terms of document authoring and representation.

Acknowledgment

We thank everyone who helped in realizing this handbook. First and foremost, this includes our authors who spent a lot of their academic time supporting this scientific endeavor. Our acknowledgement also includes the series editors as well as Barbara Karlson who accompanied this project with great patience and understanding.

Notes

1. See Lücking and Pfeiffer (2012) for a comprehensive consideration of these terms in the context of the notion of multimodal communication.
2. See Gibbon (2012) for a notion of modality that starts from a notion of communication as a pair of motor-sensory (output-input) organs, rather than separating production and reception.
3. Here one may also think of so called intelligent documents (Lobin, 1999).
4. See Menke (2012) for a similar approach using a classical model of signal transmission.
5. Ontologies concern the inference-driven processing of textual data and as such go beyond purely distributional approaches. The area of human-machine interaction based on artificial agents will profit especially from this research (Ferrucci, Brown,

Chu-Carroll, Fan, Gondek, Kalyanpur, Lally, Murdock, Nyberg, Prager, Schlaefer, and Welty, 2010; Waltinger, Breuing, and Wachsmuth, 2011; Ovchinnikova, 2012).

6. Note that wikis currently evolve as a preferred information source of these approaches (Gabrilovich and Markovitch, 2009) so that the social complexity of wiki-based communication becomes an upcoming topic of text and web mining.
7. This is reflected, for example, by a recent Dagstuhl seminar on automation in digital preservation (Chanod, Dobreva, Rauber, and Ross, 2010).

Bibliography

Ahn, Luis von
2008 "Human Computation." In: *IEEE 24th International Conference on Data Engineering (ICDE 2008)*. IEEE, pp. 1–2.

Altmann, Gabriel
1993 "Science and Linguistics." In: *Contributions to Quantitative Linguistics*. Ed. by Reinhard Köhler and Burghard B. Rieger. Dordrecht: Kluwer, pp. 3–10.

Altmann, Gabriel and Rüdiger Grotjahn
1988 "Linguistische Meßverfahren." In: *Sociolinguistics. Soziolinguistik. An International Handbook of the Science of Language and Society*. Ed. by U. Ammon, N. Dittmar, and K. J. Mattheier. Vol. 2. Berlin/New York: De Gruyter, pp. 1026–1039.

Baer, Daniel, Nicolai Erbs, Torsten Zesch, and Iryna Gurevych
2011 "Wikulu: An Extensible Architecture for Integrating Natural Language Processing Techniques with Wikis." In: *Proceedings of the System Demonstrations of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*.

Baeza-Yates, Ricardo and Berthier Ribeiro-Neto, eds.
1999 *Modern Information Retrieval*. Reading, Massachusetts: Addison-Wesley.

Berners-Lee, Tim, James Hendler, and Ora Lassila
2001 "The Semantic Web." In: *Scientific American*, pp. 34–43.

Brandes, Ulrik, Patrick Kenis, Jürgen Lerner, and Denise van Raaij
2009 "Network analysis of collaboration structure in Wikipedia." In: *Proceedings of the 18th international conference on World wide web*. WWW '09. Madrid, Spain: ACM, pp. 731–740.

Brinker, Klaus, Gerd Antos, Wolfgang Heinemann, and Sven F. Sager, eds.
2000 *Text- und Gesprächslinguistik. Linguistics of Text and Conversation*. Berlin/New York: De Gruyter.

Brinker, Klaus and Sven F. Sager
2006 *Linguistische Gesprächsanalyse*. Berlin: Erich Schmidt Verlag.

Brusilovsky, Peter
1996 "Methods and Techniques of Adaptive Hypermedia." In: *User Modeling and User-Adapted Interaction* 6.2–3, pp. 87–129.

Bush, Vannevar
1945 "As We May Think." In: *The Atlantic Monthly* 176.1, pp. 101–108.

Chanod, Jean-Pierre, Milena Dobreva, Andreas Rauber, and Seamus Ross, eds.
2010 *Automation in Digital Preservation*. Dagstuhl Seminar Proceedings 10291. Dagstuhl, Germany: Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, Germany.

Chiarcos, Christian, Sebastian Nordhoff, and Sebastian Hellmann, eds.
2012 *Linked Data in Linguistics*. Berlin: Springer, pp. 15–23.

Dang, Jing, Anthony Brabazon, David Edelman, and Michael O'Neill
2009 "An Introduction to Natural Computing in Finance." In: *Applications of Evolutionary Computing*. Ed. by M. Giacobini, A. Brabazon, S. Cagnoni, G. A. Di Caro, R. Drechsler, A. Ekart, A. I. Esparcia-Alcazar, M. Farooq, A. Fink, J. McCormack, M. O'Neill, J. Romero, F. Rothlauf, G. Squillero, S. Uyar, and S. Yang. Vol. 5484. Lecture Notes in Computer Science. Springer, pp. 182–192.

Diewald, Nils
2012 "Decentralized social network sites." In: *Handbook of Technical Communication*. Ed. by Alexander Mehler, Laurent Romary, and Dafydd Gibbon. Berlin/Boston: De Gruyter.

Dürscheid, Christa
2003 "Medienkommunikation im Kontinuum von Mündlichkeit und Schriftlichkeit. Theoretische und empirische Probleme." In: *Zeitschrift für angewandte Linguistik* 38, pp. 37–56.

Eco, Umberto
1988 *Über Spiegel und andere Phänomene*. München: Hanser.

Egozi, Ofer, Shaul Markovitch, and Evgeniy Gabrilovich
2011 "Concept-Based Information Retrieval using Explicit Semantic Analysis." In: *ACM Transactions on Information Systems* 29.2.

Engelbart, Douglas C.
1995 "Toward augmenting the human intellect and boosting our collective IQ." In: *Communications of the ACM* 38 (8), pp. 30–32.

Feldman, Ronen and James Sanger
2007 *The Text Mining Handbook. Advanced Approaches in Analyzing Unstructured Data*. Cambridge: Cambridge University Press.

Ferrucci, David, Eric Brown, Jennifer Chu-Carroll, James Fan, David Gondek, Aditya A. Kalyanpur, Adam Lally, J. William Murdock, Eric Nyberg, John Prager, Nico Schlaefer, and Chris Welty
2010 "Building Watson: An Overview of the DeepQA Project." In: *AI Magazine* 31.3, pp. 59–79.

Gabrilovich, Evgeniy and Shaul Markovitch
2009 "Wikipedia-based Semantic Interpretation for Natural Language Processing." In: *Journal of Artificial Intelligence Research*, pp. 443–498.

Gibbon, Dafydd
2012 "Resources for technical communication systems." In: *Handbook of Technical Communication*. Ed. by Alexander Mehler, Laurent Romary, and Dafydd Gibbon. Berlin/Boston: De Gruyter.

Gibbon, Dafydd, Inge Mertins, and Roger K. Moore, eds.
2000 *Handbook of Multimodal and Spoken Dialogue Systems*. Boston/Dordrecht/London: Kluwer.

Ginzburg, Jonathan and Raquel Fernandez
2005 “Action at a distance: the difference between dialogue and multilogue.” In: *Proceedings of the 9th Workshop on the Semantics and Pragmatics of Dialogue (Dialor’05), Nancy, France*.

Gómez, Vicenç, Hilbert J. Kappen, and Andreas Kaltenbrunner
2011 “Modeling the structure and evolution of discussion cascades.” In: *Proceedings of the 22nd ACM conference on Hypertext and hypermedia*. HT ’11. Eindhoven, The Netherlands: ACM, pp. 181–190.

Hammwöhner, Rainer
2007 *Semantic Wikipedia – Checking the Premises*. http://epub.uni-regensburg.de/15593/.

Harris, Zellig S.
1954 “Distributional Structure.” In: *Word* 10, pp. 146–162.

Hess-Lüttich, Ernst W. B. and Dagmar Schmauks
2004 “Multimediale Kommunikation.” In: *Semiotik: ein Handbuch zu den zeichentheoretischen Grundlagen von Natur und Kultur*. Ed. by Roland Posner, Klaus Robering, and Thomas A. Seboek. Vol. 4. Handbücher zur Sprach- und Kommunikationswissenschaft. Berlin: de Gruyter. Chap. XV. Ausgewählte Gegenstände der Semiotik/Selected Topics of Semiotics, pp. 3487–3503.

Kendon, Adam
2004 *Gesture: Visible Action as Utterance*. Cambridge, MA: Cambridge University Press.

Koch, Peter and Wulf Oesterreicher
2001 “Langage écrit et langage parlé.” In: *Lexikon der romanistischen Linguistik*. Ed. by Günter Holtus, Michael Metzeltin, and Christian Schmitt. Vol. I/1. Tübingen, Germany: Niemeyer, pp. 584–627.

Kubina, Petra and Andy Lücking
2012 “Barrier-free communication.” In: *Handbook of Technical Communication*. Ed. by Alexander Mehler, Laurent Romary, and Dafydd Gibbon. Berlin/Boston: De Gruyter.

Ladkin, Peter
2012 “Verbal communication protocols in safety-critical system operations.” In: *Handbook of Technical Communication*. Ed. by Alexander Mehler, Laurent Romary, and Dafydd Gibbon. Berlin/Boston: De Gruyter.

Landauer, Thomas K. and Susan T. Dumais
1997 “A Solution to Plato’s Problem: The Latent Semantic Analysis Theory of Acquisition, Induction, and Representation of Knowledge.” In: *Psychological Review* 104.2, pp. 211–240.

Laniado, D., R. Tasso, Y. Volkovich, and A. Kaltenbrunner
2011 “When the Wikipedians Talk: Network and Tree Structure of Wikipedia Discussion Pages.” In: *Proceedings of the 5th International AAAI Conference on Weblogs and Social Media*.

Lauth, Codrina, Bettina Berendt, Albrecht Schmidt, and Bastian Pfleging
2012 “Ubiquitous computing.” In: *Handbook of Technical Communication*. Ed. by Alexander Mehler, Laurent Romary, and Dafydd Gibbon. Berlin/Boston: De Gruyter.

Lee, Christopher A.
2012 "Digital curation as communication." In: *Handbook of Technical Communication*. Ed. by Alexander Mehler, Laurent Romary, and Dafydd Gibbon. Berlin/Boston: De Gruyter.

Leuf, Bo and Ward Cunningham
2001 *The Wiki Way. Quick Collaboration on the Web*. Boston: Addison Wesley.

Lobin, Henning
1999 "Intelligente Dokumente. Linguistische Repräsentation komplexer Inhalte für die hypermediale Wissensvermittlung." In: *Text im digitalen Medium. Linguistische Aspekte von Textdesign, Texttechnologie und Hypertext Engineering*. Ed. by Henning Lobin. Opladen: Westdeutscher Verlag, pp. 89–119.

Lücking, Andy, Alexander Mehler, and Peter Menke
2008 "Taking Fingerprints of Speech-and-Gesture Ensembles: Approaching Empirical Evidence of Intrapersonal Alignment in Multimodal Communication." In: *LONDIAL 2008: Proceedings of the 12th Workshop on the Semantics and Pragmatics of Dialogue (SEMDIAL)*. King's College London, pp. 157–164.

Lücking, Andy and Thies Pfeiffer
2012 "Framing multi-modal technical communication. With Focal Points in Speech-Gesture-Integration and Gaze Recognition." In: *Handbook of Technical Communication*. Ed. by Alexander Mehler, Laurent Romary, and Dafydd Gibbon. Berlin/Boston: De Gruyter.

Mehler, Alexander
2009 "Artifizielle Interaktivität. Eine semiotische Betrachtung." In: *Medienwandel als Wandel von Interaktionsformen – von frühen Medienkulturen zum Web 2.0*. Ed. by Tilmann Sutter and Alexander Mehler. Wiesbaden: VS.

Mehler, Alexander and Tilmann Sutter
2008 "Interaktive Textproduktion in Wiki-basierten Kommunikationssystemen." In: *Kommunikation, Partizipation und Wirkungen im Social Web – Weblogs, Wikis, Podcasts und Communities aus interdisziplinärer Sicht*. Ed. by Ansgar Zerfaß, Martin Welker, and Jan Schmidt. Köln: Herbert von Halem, pp. 267–300.

Melucci, Massimo, Ricardo Baeza-Yates, and W. Bruce Croft, eds.
2011 *Advanced Topics in Information Retrieval*. Vol. 33. The Information Retrieval Series. Berlin: Springer.

Menke, Peter
2012 "Evaluation of technical communication." In: *Handbook of Technical Communication*. Ed. by Alexander Mehler, Laurent Romary, and Dafydd Gibbon. Berlin/Boston: De Gruyter.

Nelson, Theodor Holm
1995 "The heart of connection: hypermedia unified by transclusion." In: *Communications of the ACM* 38 (8), pp. 31–33.

Newman, Mark E. J.
2010 *Networks: An Introduction*. Oxford: Oxford University Press.

O'Donnell, Michael, Chris Mellish, Jon Oberlander, and Alisair Knott
2001 "ILEX: An Architecture for a Dynamic Hypertext Generation System." In: *Natural Language Engineering* 7, pp. 225–250.

Ovchinnikova, Ekaterina
2012 *Integration of World Knowledge for Natural Language Understanding*. Paris: Atlantis Press.

Paaß, Gerhard
2012 "Document classification, information retrieval, text and web mining." In: *Handbook of Technical Communication*. Ed. by Alexander Mehler, Laurent Romary, and Dafydd Gibbon. Berlin/Boston: De Gruyter.

Pickering, Martin J. and Simon Garrod
2004 "Toward a mechanistic psychology of dialogue." In: *Behavioral and Brain Sciences* 27, pp. 169–226.

Posner, Roland, Klaus Robering, and Thomas A. Seboek, eds.
1996 *Semiotik: ein Handbuch zu den zeichentheoretischen Grundlagen von Natur und Kultur*. Vol. 1. Handbücher zur Sprach- und Kommunikationswissenschaft. Berlin: de Gruyter.

Rickheit, Gert
2005 "Alignment und Aushandlung im Dialog." In: *Zeitschrift für Psychologie* 213.3, pp. 159–166.

Rieger, Burghard B.
2003 "Semiotic Cognitive Information Processing: Learning to Understand Discourse. A Systemic Model of Meaning Constitution." In: *Adaptivity and Learning. An Interdisciplinary Debate*. Ed. by R. Kühn, R. Menzel, W. Menzel, U. Ratsch, M. M. Richter, and I. O. Stamatescu. Berlin: Springer, pp. 347–403.

Ruiter, Jan Peter de, Stéphane Rossignol, Louis Vuurpijl, Douglas Cunningham, and Willem J. M. Levelt
2003 "SLOT: A research platform for investigating multimodal communication." In: *Behavior Research Methods* 35 (3), pp. 408–419.

Sager, Sven F.
1997 "Intertextualität und die Interaktivität von Hypertexten." In: *Textbeziehungen: linguistische und literaturwissenschaftliche Beiträge zur Intertextualität*. Ed. by Josef Klein and Ulla Fix. Tübingen: Stauffenburg, pp. 109–123.

Sasaki, Felix and Arle Lommel
2012 "Multilingual computing." In: *Handbook of Technical Communication*. Ed. by Alexander Mehler, Laurent Romary, and Dafydd Gibbon. Berlin/Boston: De Gruyter.

Schaffert, Sebastian, Diana Bischof, Tobias Bürger, Andreas Gruber, Wolf Hilzensauer, and Sandra Schaffert
2006 "Learning with Semantic Wikis." In: *Proceedings of SemWiki2006 Workshop "From Wiki to Semantics"*, pp. 109–123.

Steels, Luc and Peter Hanappe
2006 "Interoperability through Emergent Semantics. A Semiotic Dynamics Approach." In: *Journal on Data Semantics* VI, pp. 143–167.

Stegbauer, Christian
2009 *Wikipedia: Das Rätsel der Kooperation*. Wiesbaden: VS.

Steyvers, Mark and Josh Tenenbaum
2005 "The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth." In: *Cognitive Science* 29.1, pp. 41–78.

Storrer, Angelika
2002 "Coherence in Text and Hypertext." In: *Document Design* 3.2, pp. 156–168.
Turney, Peter D. and Patrick Pantel
2010 "From Frequency to Meaning: Vector Space Models of Semantics." In: *Journal of Artificial Intelligence Research* 37, pp. 141–188.
Völkel, M., M. Krötzsch, D. Vrandecic, H. Haller, and R. Studer
2006 "Semantic Wikipedia." In: *Proceedings of the 15th International Conference on World Wide Web (WWW '06), Edinburgh, Scotland, May 23–26*. New York: ACM Press, pp. 585–594.
Waltinger, Ulli and Alexa Breuing
2012 "Internet-based communication." In: *Handbook of Technical Communication*. Ed. by Alexander Mehler, Laurent Romary, and Dafydd Gibbon. Berlin/Boston: De Gruyter.
Waltinger, Ulli, Alexa Breuing, and Ipke Wachsmuth
2011 "Interfacing Virtual Agents With Collaborative Knowledge: Open Domain Question Answering Using Wikipedia-based Topic Models." In: *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI 2011), July, 16–22, 2011, Barcelona, Spain*.
Wasserman, Stanley and Katherine Faust
1999 *Social Network Analysis. Methods and Applications*. Cambridge: Cambridge University Press.
Wattenberg, Martin, Fernanda B. Viégas, and Katherine J. Hollenbach
2007 "Visualizing Activity on Wikipedia with Chromograms." In: *Human-Computer Interaction – INTERACT 2007, 11th IFIP TC 13 International Conference, Rio de Janeiro, Brazil, September 10–14, 2007, Proceedings, Part II*. Ed. by Maria Cecília Calani Baranauskas, Philippe A. Palanque, Julio Abascal, and Simone Diniz Junqueira Barbosa. Vol. 4663. LNCS. Springer, pp. 272–287.
Wehner, Josef
1997 "Interaktive Medien – Ende der Massenkommunikation." In: *Zeitschrift für Soziologie* 26.4, pp. 96–11.
Weidenmann, Bernd
1997 "Multicodierung und Multimodalität im Lernprozess." In: *Information und Lernen mit Multimedia*. Ed. by Ludwig J. Issing and Paul Klimsa. Weinheim: Beltz, pp. 45–62.
Wilks, Yorick
1997 "Information Extraction as a Core Language Technology." In: *Information Extraction: a Multidisciplinary Approach to an Emerging Information Technology. Proceedings of the International Summer School (SCIE-97), Frascati, Italy*. Ed. by Maria Terese Pazienza. Lecture notes in computer science 1299. Berlin: Springer, pp. 1–9.
Zerubavel, Eviatar
2003 *Time Maps: Collective Memory and the Social Shape of the Past*. Berkley: University of California Press.

I. Basic communication technologies & infrastructures

1. Document Authoring: from Word Processing to Text Generation

Aurélien Max

1. Introduction

This chapter discusses document authoring with the help of computers. Many activities that are naturally carried out with computers are closely related to the processes involved during different stages of text writing. So much that Barrett (1991) argues that "*there exists a natural affinity between writing and computers, almost a genetic relationship. They are used as tools for representing and discovering knowledge and as instruments for communication between people*". As computers have imposed themselves for several decades now in writing practices, studying the current state of the art in using computers for producing text and the major evolutions that led to it may help better understand how industrial initiatives and research efforts will take writers to the next generation of authoring practices.

We will first review the domain of computer-assisted authoring for free and constrained text authoring, where the human writer takes the main role. We will then discuss the other end of the spectrum with document generation and transformation by computer, and visit intermediary areas. Future directions drawn from research in Natural Language Processing will conclude the chapter.

2. Computer-assisted Authoring

This section is concerned with the implications and possibilities offered by using computers for creating textual documents. We start by discussing the main steps of a typical cycle of document creation, and review the main ways in which computers can assist humans and discuss their impact on document creation and usage. This section then moves on to computer software that can be used for supporting free authoring activities. We first consider the opposition between rendered editing and declarative formatting. We then turn to the description of tools for structuring documents, tools for searching information during writing, and tools for language analysis. The last part of this section considers the domain of constrained text authoring, which is of significant importance for technical communication. We first discuss the constraints that can be put on language use with controlled languages, and then the constraints that can be put on document content and structure.

2.1. Document management and text processing

2.1.1. Document lifecycle and the production of documents

The lifecycle of a high-quality document is quite complex: from its initial conception to one of its actual uses, under some form and by some profile of readers, many processes had to take place. A typical document lifecycle, which usually contains inner loops, has the following steps (Wilkinson et al. 1998): *document description*, where the content, structure and metadata of a document are decided upon; *document production*, where a document management system is used to produce a new and unique form of a document; *document delivery*, where information is delivered to the relevant recipients; *document publication*, where a document is registered and stored for later retrieval; and *document discovery*, where a document is proposed to a user in response to some expressed information need. There is an important body of literature on all these subjects, and the advent of the Internet and of digital librairies has only increased the activity of the corresponding research areas. This chapter is primarily concerned with *document production*, and in particular how machines can be of assistance in the process of producing the textual parts of documents. We first review existing technologies that are typically used for *free text authoring* in section 2.2, and we then move on to the case of *constrained authoring* and its associated technologies in section 2.3. Documents are created using *Document Management Systems* (DMS), which, while being potentially quite complex if covering important parts of the document lifecycle, include or call some *text authoring software*, whose role is to assist authors in creating written electronic documents.[1] We review such programs in section 2.2.

The most natural way to create new documents is by writing their content entirely, for example from a verbal specification, from written notes or from ideas in an author's mind. However, in many cases, notably in technical and scientific communication, documents can also be created by reusing existing text. *Document integration* denotes cases where document fragments are integrated to form a coherent whole, as is the case for edited books such as the present volume. Various processes are needed, including the construction of global tables of contents, indexes and bibliographies, as well as the creation and resolution of appropriate cross-references. Such processes can be easily automated using standard DMS functionalities provided the document integration was carefully planned in advance. Ensuring coherence of content (e.g. no overlapping or missing content) and linguistic usage (e.g. approved terminology or language dialect) is a much more difficult task. Again, careful *a priori* description of the document and communication of adequate guidelines to authors should limit the need for manual editing. When importing documents which were not initially considered for integration, complete restructuring or even rewriting may be required.

Document modification corresponds to cases where some existing version of a document is reused to produce a new version, which contains the same content presented in a way tailored to its intended readers, or contains only a summary of the original information, or adds new information. Whether the original documents were initially considered for such uses may again greatly facilitate or, on the contrary, hinder the creation of new documents by means of adequate software. We will discuss document reuse later in the context of structured authoring in section 2.3, and in the context of automatic and interactive text transformation in section 3.3.

2.1.2. Types of assistance by the machine for producing documents

Machines certainly have an important role in increasing *productivity* in document creation. Perhaps not less importantly, machines should also participate in increasing the *quality* and *usability* of the created documents. The issue at hand is therefore not only to create documents globally *faster*, but also to create *better* documents that will be easier to understand and use by their readers as well as easier to find and reuse. More generally, one could also expect technology to assist in producing documents for which all potential future uses (e.g. translation, adaptation, summarization) will be significantly facilitated, if not direct by-products of the authoring process (see e.g. Hartley and Paris 1997).

Such requirements however rely on the integration of mature and advanced language technologies in the software supporting the document lifecycle. In a state-of-the-art on language technologies published fifteen years ago, the assistance by the computer in the preparation of the content and linguistic expression was described as holding "*tremendous potential*" (Halvorsen 1997); although some language technology solutions have emerged and are used in practice, as we will show in the next sections, major evolutions are yet to come to make a radical change in how the machine effectively assists in text authoring. For most computer users, text editors or word processors are now considered standard tools; yet, only a fraction of them use integrated *spell checkers*, and a minority would use advanced tools such as *grammatical* and *style checkers*[2]. For the general public, most of the assistance is concentrated on the rendering capabilities of the authoring software, and the ability to edit documents in their final printed form appears to be the most common expectation from such users (see Section 2.2). For obvious reasons, professional writers have adopted some technologies ahead of the general public, which broadly operate in two areas: controlling *linguistic complexity* (discussed in Section 2.3) and guiding *document structuring* (discussed in Section 2.3). A critical view on some of the solutions proposed would describe them as *constraints* imposed on the human author rather than effective assistance from the machine. Additional complexity on the work of the document creator has been indeed the key to producing more structured and usable documents. Nonetheless, the positive impact of these solutions should not be underestimated, and their wide adoption by professional document providers reflects that they have filled important positions when they were introduced.

Information technology applied to text authoring can provide assistance of three types (Dale 1997): an existing task can be performed faster by a person with the help of the machine (*accelerative technology*); a task previously performed by a person can now be carried out by the machine (*delegative technology*); and the machine can assist a person in carrying out an existing task (*augmentative technology*). At the time of the previously mentioned state-of-the-art on language technology, Dale predicted that the next major developments in computer assistance for text creation and editing were to be of an augmentative nature, with an increased complementarity between people and machines. *Interactive processes* (e.g. interactive grammar checking, interactive text generation) indeed appear as the most promising viable technology solutions for the short term, as they allow a fine-grained control by a person while efficiently delegating to the machine what it performs flawlessly, such as performing Subject-Verb agreement in automatic sentence generation when both the subject and its verb have been identified. However, such technologies are still not well known to the general public and are not deployed in the general office. This state of affairs is certainly attributable to both the current level of performance of the involved language technology software (as illustrated, for example, by the numerous *false alarms* raised by current grammar checking tools) and to the sometimes radical changes in uses and practices that most users or institutions are not ready to adopt. This section needs not however end on a negative tone: as we will see in section 4, research in language technology is very active in several areas and is likely to produce in the years to come some desirable changes in the way in which the machine can assist in text authoring.

2.1.3. Assessing the impact on the document creation and usage processes

An important consideration to which we now turn is how assistance by the machine in creating documents may impact the activities of the people who create and those who use them. Broadly speaking, the concept of *productivity* encompasses most desirable objectives: by using machines, fewer resources (time, money, human contributors) should be necessary to produce new documents or versions of documents, as well as fewer resources for carrying out the tasks associated with using these documents. It can also be understood in terms of *productivity in subsequent document uses*, as a significant proportion of documents are likely to undergo further processing: for instance, translation of a document should be made as easy and fast as possible.

Writing is a very demanding task that requires a lot of executive attention because of the heavy demands it places on working memory (Flower and Hayes

1980). Psychocognitive studies reveal the overall complexity of the task, and bring empirical evidence of what cognitive capabilities and working conditions are necessary. The types of writing processes involved during text composition seem to follow a typical pattern of temporal organization (Olive et al. 2002): the *planning* activity declines throughout writing, the *translation* of contents into text activity remains constant, while the *revision* activity increases after the middle of a writing session. The verbal and visual components of the working memory have been shown to be highly sollicited during text composition (Olive et al. 2008). The verbal working memory may support formulation of the text at the phonological or orthographic levels of representation, and is also involved in the reading that frequently interrupts the writer during text composition and revision. The visual working memory would be mostly used during planning, while the spatial working memory may be mainly sollicited by specific drafting strategies.

Researchers in writing studies resort to various measures to measure *writing performance*, including number of words per time unit and human assessed quality of the produced text in terms of use of language and number of arguments, and compute appropriate interjudge reliability scores. A significant finding of the study of Olive et al. (2008) is that writers compose their text faster (in terms of the number of arguments produced in a fixed amount of time) without any concurrent task involving any of the working memory components, but that text quality seems not to be affected by any of these conditions.

Text quality can be linked to the ease with which readers can understand text content. *Automatic evaluation of text quality* has long been an object of research, with a focus on either determining the appropriateness of a text for some types of readers, or ranking text by decreasing quality for competent language users. The central issue concerns identifying and combining complexity factors that are good predictors of the perceived quality of a text. In a recent study, Pitler and Nenkova (2008) show that the presence of rare or complex vocabulary and of incoherent discourse relations are the most useful factors. Other factors include general text characteristics (such as average sentence length), syntactic characteristics of sentences, entity coherence and the form of referring expressions. See (Menke 2012) in this volume for a detailed presentation of the issue of the evaluation of technical communication.

2.2. Tools for free text authoring

2.2.1. Rendered editing vs. declarative formatting

Document creation almost always involves the writing of some new text. *Text editors*[3] are basic tools found on every modern computer which allow the creation and modification of any set of data that can be expressed using human-readable characters. Characters for any language script available from the user's keyboard or a virtual keyboard can be entered and assembled into words and larger linguistic units. Documents are saved using some appropriate *encoding* that ensures that all characters can be recognized unambiguously and associated to their appropriate glyph. The advent of the *Unicode standard*[4] now offers generic solutions for describing, representing and interchanging electronic text in most of the world's writing systems, although many other encodings developed for specific languages or language families are still in use today, in particular for legacy documents (see also Sasaki and Lommel (2012) in this volume).

Text editors in their most simple form are agnostic as to the type of data they handle, but most of them support *syntax modes*. For instance, such a mode for source code for a programming language assigns colors to tokens depending on their type (e.g. instructions, variables, strings). However, parts of documents expressed in natural language are merely considered as sequences of characters. Some text editors have basic *tokenization* capabilities to segment a text into individual words and other signs, so as to allow, for example, extending text selections up to word boundaries or counting the number of words. In addition to the basic editing operations of inserting or deleting individual characters, functions are provided for searching text and performing replacements, as well as for operating on *text buffers* to copy, cut or paste their content. Selected text spans, which correspond to possibly any level of contiguous text within a document, can thus be moved in the document. This is one of the most useful function of text editors compared to handwriting, because it makes restructuring at the level of short text units simple. Lastly, a function to undo the last editing actions is usually provided to allow the user to quickly restore some previous state of a document from the history of changes.

As computers became more powerful, users gained more control over the presentation of their documents with *word processors*[5], which are now the most widely used tools for creating documents with computers. Modern word processors allow for designing and applying *style sheets* that define how specific text units such as section headings or footnotes should look like when printed, but they also allow editing the text as it would be rendered in its final form, an instantiation of the *WYSIWYG* (*What You See is What You Get*) paradigm. Style sheets are important to guarantee presentation consistency across documents, and also because they are intended to free the writer from dealing with style issues while composing. Document structure can be expressed by applying styles to text spans corresponding to section headings, which can be compiled automatically into tables of contents. *Templates* are more advanced descriptions of what documents of a specific type should contain and look like, and are typically useful for facilitating and constraining the creation of stereotypical documents.

Various other functions are provided by word processors to ease document creation and favor presentation consistency, including internal and external reference handling, index and bibliography compilation, figure and table numbering, and inclusion of specific objects such as diagrams and formulae. Modern word processors also have features of particular importance for *collaborative editing and reviewing*, including annotating documents with notes and tracking changes between successive document versions.

Another way of creating documents, using any text editor, involves adding specific commands in documents that describe the type of text spans. *Declarative formatting* is intended to have writers concentrate on the content of their texts and its description rather than on the details of their visual presentation. A small number of commands can be used to describe logical structure, define anchors and references, or to specify some characteristic of a text span. For instance, a command can be inserted to define that some consecutive words should be emphasized, and a *rendering engine* later applies the appropriate low level commands defined in the active style sheet to produce the final version in some output format suitable for printing and diffusion, such as the Portable Document Format (PDF). Popular *markup languages* include LaTeX for describing technical and scientific documents rendered with the TeX typesetting system, and the Hypertext Markup Language (HTML) for describing web documents processed by web browser rendering engines (for more information on these topics see Rahtz (2012) and Stührenberg (2012) both in this volume). The commands of these languages can be extended to define new classes or to redefine existing ones so as to allow the tailoring of documents to meet any document layout standards.[6] Such languages prove to be particularly useful for *content reuse*, as any part of a document can be reused and rendered appropriately within the context of a new document.

Other advanced functions, including language processing utilities, can be used in conjunction with text editors and word processors. We will review the most important ones in the following sections. Independently of the type of assistance that they bring to the writer and of their performance, their adoption is often directly linked to their direct integration as a *plug-in* into the writer's familiar authoring software. Note finally that the choice of an authoring system for large organizations has strong implications, in particular as regards user training and compatibility with other products (Wilkinson et al. 1998).

2.2.2. Tools for structuring documents

The structure of documents is an essential characteristic that has a direct relationship with how effectively their readers will comprehend their content. The division of complex knowledge into hierarchically organized subparts is necessary to guide readers in progressively assimilating the communicative goals of a document, but also allows to directly identify useful subparts of documents when searching for specific information. Well-structured technical documents are those that meet both user expectations and industry standards. In this section, we review existing software that can be used for organizing the content of any type of documents.

Many writers resort to pen and paper for the initial stages of drafting, where ideas are enumerated and sequences of merging, splitting and linking yield an initial document plan that is often subsequently revised during composition. Some software programs can provide assistance to help mature the structure of a document that has not yet been decided upon. *Document outliners* usually show a visual representation of a document's structure organized as a tree whose nodes are logical subparts of the document such as chapters and sections. Outliners display views of the document from any user-specified level by providing them with the ability to expand or collapse any subtree of the document structure. Editing actions can be applied on subtrees, and are in particular useful for reorganizing documents at any stage of their creation by moving entire subparts. These tools facilitate visualization of a document's logical structure and allow authors to have a global perception of a document while performing editing operations at any appropriate level.

The success of outliners certainly comes from their tight integration within word processors, which permits a writer to easily go back and forth between document structuring and text composition. Other applications primarily target the stage of knowledge structuring, often with export capabilities to use some organization of ideas as initial document plans in a word processor. *Mind maps* are graphical representations that link ideas or any relevant objects and have diverse applications related to brainstorming. Large amount of information can be structured as graphs whose visual representations support node folding for concentrating on specific areas as well as common editing operations. Mind mapping software[7] belong to the larger family of *knowledge organization tools* which also include *wikis*, whose growing popularity owes to their simplicity of use, their simple and effective model for structuring information as text, and their availability as online collaborative tools. These tools mostly differ in their degree of support of well-established document authoring systems and their support for collaborative editing. One of the main limitations of tools in this category is the absence of a satisfying solution for bidirectional synchronization between the knowledge structures on the one hand, and the evolving texts on the other.

Lastly, authors often need help for integrating and organizing existing text units coming from various sources, but this requires text interpretation capabilities that are far beyond the current state-of-the-art. Relevant research areas include *automatic text segmentation*, which is concerned with segmenting text into thematically coherent units, and *automatic text clustering*, whose object it is to group similar texts (see e.g. Manning and Schütze 1999; see also Paaß (2012) in this volume on this and related methods).

As Dale (1997) noted, "*the degree of organizational assistance that is possible depends very much on the extent to which regularity can be perceived or imposed on the text concerned*". Solutions for constrained document authoring will be reviewed in section 2.3.

2.2.3. Reference tools for authoring with a computer

As writing heavily involves referring to existing knowledge under some published form, the process of accessing information is closely intertwined with document creation. While using search engines within web browsers to search the web or internal document repositories is now very common when writers have a specific information need, *recommander systems* can be integrated within word processors to present relevant information coming from e.g. web document search during the authoring process in a proactive manner (Puerta Melguizo et al. 2009). Citations of published works can be managed with *reference management tools*, such as the BibTex system for LaTeX, which process bibliographical structured entries and produce both reference sections in documents and appropriate citations in the text using predefined style sheets.

Another important need during writing is access to reference linguistic information, which is often made possible within authoring systems. *Electronic dictionaries* are used to check the correctness of the spelling of a word or its appropriateness in some context given its definition, and *specialized lexicons* additionally contain domain-specific words such as proper names. Writers often use *synonym dictionaries* to find alternative words that may be more appropriate in the context of some text[8], or even *thesauri* which group words according to various types of similarity or relatedness (e.g. antonymy), thereby providing assistance in lexical access. The most useful reference tools for technical writers are *terminologies*, which contain expressions (*terms*) that should be used to denote concepts in a speciality field (e.g. medicine, computer science, agronomy). These notions are covered in detail in this volume (Sharoff and Hartley 2012).

Writers sometimes need to check the appropriateness of some words or expressions in specific contexts, a need which is more frequent when they are not familiar with a specific domain or do not have a native command of the language of a document.[9] *Concordancers* are software that retrieve instances of words or phrases in context from corpora and are thus useful to check if a given expression can be found in the exact context or some approximate context of the text being written. Using such tools on domain-specific or proprietary corpora can be used to check that some text span conforms to some in-house writing style, but use of literal search for expressions with web search engines has also become a common practise.

2.2.4. Language-sensitive tools

Written documents are complex linguistic forms whose primary goal is to convey information from a writer to a reader. The most efficient level of assistance that a writer could obtain would require deep understanding of the communicative goals that a document should contain to suggest the most appropriate structure and linguistic realizations to express them. The current state-of-the-art in *Natural Language Processing* (NLP) still falls short of meeting these requirements. Research, however, has produced results which over the years have been integrated with varying success into authoring aids. This section concentrates on well-established technologies, operating at shallow linguistic levels, while Section 4 will point to various research works that may find their way into future authoring aids.

The spectrum of types of language-aware assistance that authoring tools can bring to writers spans from *detection of errors* and *detection of perfectible text fragments* to *automatic error correction* and *text improvement*. An intermediate level is to provide automatic results of *critical textual analysis*, showing various statistics and measures meant to indicate linguistic quality and readability. As an alternative to automatic correction, *interactive correction and rewriting* involves the writer by suggesting alternative wordings that could be used to improve a text. With the current level of assistance that can be provided by computers, user satisfaction and adoption can vary from enthusiasm to complete rejection. This is the case, for instance, of the simple word completion feature provided by various authoring tools: in some instantiations, suggestions of word completions from a dictionary or the existing part of the document are displayed after a few letters have been typed, which can be either perceived as useful assistance in spelling or typing speed or as a distractive feature.

The types of language-sensitive tools found as optional modules in modern authoring tools still follow those of the early CRITIQUE system (Richardson and Branden-Harder 1988) developed at IBM in the 1980s: *lexical analysis*, *grammar error detection* and *stylistic analysis*. As technology improved, notably through increased consideration of syntactic and semantic features of sentences, suggestions of better candidates or even automatic corrections for frequent error patterns have been added to these tools. Another major evolution is the move from a situation where error patterns and confusion sets were manually developed, to a situation where machine learning (see e.g. Paaß (2012) in this volume) is used to derive relevant knowledge from existing text corpora.

Lexical analysis is performed by *spelling checkers* after the text has been tokenized to identify words, which in itself can be problematic as some errors can result in words being accidentally merged with surrounding words or split up. Correcting words in text relies on the following processes (Kukich 1992): the identification of errors corresponding to non-words, the correction of errors on isolated words, and the correction of errors on words in context. Determining whether a word is valid or not is generally done by looking it up in some electronic dictionary. However, dictionary coverage is by nature limited due to proper nouns, neologisms, and the impossibility of enumerating all possible word forms for morphologically-rich languages. The probability of a sequence of characters for a given language, as derived from large lexica, can also be a useful indicator that a given word is unlikely to be correct.

Studies on *spelling error patterns* reveal that there exist many different causes of errors, and that although most misspelled words differ by one character from their correct form, phonological errors, for example, are very difficult to detect and to correct. Correcting words is generally done by finding a set of candidate corrections associated with some score; *automatic correction* implies replacing some incorrect word with the candidate with the highest score, while in *interactive correction* the list of candidates is presented to the user in order of decreasing score, which is the typical scenario within authoring tools. Among the various techniques developed for correcting isolated words, finding known words at a minimum *edit distance*, measured as the minimum number of edit operations on characters (e.g. insertions and deletions) to transform one word form into another one, and exploiting probabilities of character sequences and probabilities of human-generated errors have been shown to reach good performance, in particular for proposing the correct word forms when returning reasonably small sets of candidates (Kukich 1992). However, most users would of course prefer systems proposing a single correct result and do not react well to the *false alarms* that are raised for unknown but valid words.

For some types of errors, in particular those where the misspelled words correspond to existing words (*real-word errors*), the context in which the words occur must be taken into account. This is illustrated in the sentence "*He is trying to fine out*", where the verb *fine* was used in lieu of the verb *find*, although the particle *out* imposes the latter. Various studies on English reveal that such errors can amount to more than 50 % of all errors encountered in a text, and that errors due to syntactic violations (e.g. incorrect subject-verb agreement) can amount to up to 75 % of errors (Kukich 1992). Detecting and finding these errors is typically the role of *grammar checkers*. The approach taken by many commercial systems is to use manually-built patterns of common errors. Such patterns can be based on superficial clues to describe the context of a word, which can be computed in a robust way, but syntactic information proves to be necessary in many cases. However, sentence parsing is still a difficult problem whose sources of errors are compounded in case of ill-formed input. The probabilities of sequences of words and sequences of word part-of-speeches, encoded in *language models* derived from very large text corpora, capture local dependencies and properties of word cooccurrence that can suffice to correct some types of errors. For further progress to be made, parsing technology has to become more robust as well as more accurate (Dale 1997), and large corpora of naturally-occurring errors in context (Max and Wisniewski 2010) should be used to build statistical models of errors that writers are likely to make.

Various types of notifications are now available in authoring tools to report potential style issues, including the use of colloquial or jargon expressions. Such notifications by *style checkers*, which are language-dependent, may take into account characteristics of the author and of the intended audience when relevant. The last area in which assistance may be proposed to the user encompasses other aspects of *readability support*. A typical feature consists in providing the writer with various statistics at the text level, some straightforward to calculate, such as the average number of words per sentence, others relying on more complex analyses, such as the ratio of use of passive and active voice in sentences. The word processor extension described in (Newbold and Gillam 2010) aims at providing feedback to authors on how to improve the quality of their documents, possibly at the sentence level. For instance, a description of *propositional density*, which involves the proportion of single and compound nouns in a sentence, is shown to the writer, as well as a description of *coherence*, where highly coherent sentences are defined as those with the most bonds to the themes in a document.

2.3. Constrained text authoring

2.3.1. Controlled Languages

Technical documents convey information that readers have to interpret unambiguously and often translate into concrete actions. Because of their importance, the *usability* of technical documents has received significant attention from entire industrial sectors concerned with improving the quality of their communication and reducing their support costs. Among the most important points of focus in the field of technical communication (see the present volume for a broad presentation), the *comprehensibility* of documents has long been an important concern, as shown by the existing number of textbooks and manuals for training technical writers (see e.g. Alred et al. 2008). General principles (e.g. concision, clarity, simplicity) are instantiated by numerous concrete guidelines on linguistic expression, including for example the limitation of sentence length and of the use of pronouns and relative clauses. As most technical documents and their revisions have to be multilingual, another major interest for the industrial sector was the reduction of translation costs. Producing higher quality documents simplifies the translation process[10], but for dramatic changes in translation time to occur documents had to be made more easily translated by the machine. The *translatabilty* of documents can be increased by limiting to a minimum the sources of difficulty for the machine. *Controlled languages* (see e.g. Kittredge 2005; see also Trippel 2012 and Sasaki and Lommel 2012 both in this volume) have been devised with the two above objectives in mind, initially on English and later for other European languages. Controlled languages are defined relative to a natural language and a technical domain, mainly by defining which syntactic structures and lexical items can be used, with specific senses attached. The use of domain-specific *terminologies* also participate in diminishing significantly linguistic variation (Sharoff and Hartley 2012).

The ASD Simplified Technical English[11] is one example of such a controlled language designed for aerospace maintenance documentation. Its continuous use from the early 1980s and the European Community Trademark that its specification received in 2006 are signs of the important and useful role that a controlled language can have in technical documentation. The specification comprises about 60 writing rules and about 900 words with selected meaning and approved part-of-speech. Words are described in a dictionary which provides their approved meaning together with approved and not approved examples of usage. For instance, the noun *ability* is not allowed, but its entry points to the modal verb *can* as a replacement. The sentence *One generator has the ability to supply power for all the systems* is therefore not valid, and should be rewritten as *One generator can supply power for all the systems*. Other words can be allowed but then can only appear with one consistent meaning, as it the case of the verb *follow* that can only be used with the meaning *come after*, so that the sentence *Follow the safety instructions* should be rewritten as *Obey the safety instructions*.

The latter example illustrates one of the major issues faced by *controlled language conformance checkers*, the software used to check that documents conform to the rules of a given controlled language specification. Checking whether a valid word is used in its approved meaning requires identifying its meaning in context, a very difficult task which is the object of the field of *automatic word sense disambiguation* (see e.g. Manning and Schütze (1999) and Paaß (2012)). This type of difficulty and others such as ensuring that sentences do not have more than a given number of words make it extremely difficult for conformance checkers to perform any type of automatic correction. Writing in a controlled language is therefore a skilled practice which requires significant training. However, the important reduction in complexity and ambiguity afforded by the use of controlled languages can be an important payoff for industrial sectors, notably in terms of document translation costs. Ongoing research efforts in this area include works on improving correction suggestions and using discourse-level information for error detection and correction.

2.3.2. Structured authoring

The specification of technical documents can go beyond rules constraining linguistic expression by describing valid document structures. Regulations and strong expectations from their users indeed make technical documents good candidates for content description and structuring. Logical and coherent units of text can be described as being of a given *type* (e.g. *Warning* in a plane maintenance procedure) and arranged into larger structures describing the logical structure of a document (e.g. a *Procedure* can be composed of a sequence of *Step* and *Warning* elements). In section 2.2 we already discussed the annotation of text elements for declarative formatting. Structural description can follow the same annotation principles, but targets more specifically the semantics of document elements and ensures that documents are complete and well-formed. In fact, annotations can be used as a data representation mechanism by making data types explicit from natural language documents. The actual presentation of documents can still be prepared by appropriate rendering mechanisms by applying *style sheets* that can operate on elements depending on specific attributes and their position in the document structure. A variety of *markup languages* and their support tools have been devised for annotating at various levels. The Standard Generalized Markup Language (SGML)[12], derived from IBM's Generalized Markup Language in the early 1980s and international standard since 1986, was designed to describe the structure of documents using annotations by means of *tags*. Markup languages define two syntaxes, one for including markup in text, and another one for describing the structure of valid elements. At the heart of this approach is the notion of conformance to some content model definition defined in *schemas* or *document type definitions* (DTD). The automatic process of *validation* ensures that a given document conforms to some predefined schema. The eXtensible Markup Language (XML)[13] was later developed by the World Wide Web Consortium as a simplification of SGML primarily intended for large-scale electronic publishing, but found a wider adoption and is now also commonly used as a data interchange format.

One of the main characteristics of these languages is that they were designed as *meta-languages* for defining more specific languages, such as an industry-specific document type. DocBook (Walsh 2010, Rahtz 2012), currently maintained by the Organization for the Advancement of Structured Information Standards (OASIS), is one such markup language primarily intended for general technical documentation. In DocBook, markup elements are defined at three levels: structure, block, and inline. The structure-level defines elements that correspond to document parts (e.g. chapters and parts in books) and that can in turn be described as being composed of sequences of sub-elements, some possibly optional (e.g. glossaries in books). Block-level elements typically correspond to the level of paragraphs or lists, and thus possibly annotate large fragments of text. Inline elements usually wrap shorter text fragments to indicate some properties related to a text fragment, such as internal or external references (like HTML anchors) or abstract rendering attributes (e.g. emphasis). Using the generic characteristics of these languages, SGML- and XML-aware editors can perform document validation against some specific language such as DocBook, and may even implement guided-authoring modes in which writers have to choose among valid elements given the current position in the structure of a document. This guarantee that the resulting document will be valid relative to some schema can also be accompanied with a WYSIWYG editing similar to that of word processors. Publishing software can finally produce a variety of document versions and formats suitable for diffusion with no direct intervention from document authors.

More recently, *topic-based authoring* has emerged to increase content reuse and specialization in technical publications. The Darwin Information Typing Architecture (DITA)[14], originally developed by IBM and now maintained by OASIS, defines *topics* as the units of authoring. DITA defines three types of topics, namely *Tasks*, *Concepts* and *References*. A variety of metadata are used to describe topics so that they can be efficiently found when reusing content or editing existing content. Topics can be organized into *maps* that describe the sequence and hierarchy of topics for specific document instances. For example, the description of a given procedure can be utilized both in a user manual and in a focussed web tutorial. *Conditional text* can be used to express filters or rendering preferences based on various attributes of a map, including the intended audience of a document or its type of diffusion. Again, DITA elements can be specialized to cover new topic types, and a number of industry fields organized in subcommittees have published specifications, for instance for the pharmaceutical, machine industry and semiconductor sectors. Advocates of topic-based authoring put forward the appropriateness of the granularity of topic elements, which are meant to contain information to describe elementary tasks, concepts and references and favor reuse and consistency in description and presentation. Several advantages can also be found as regards the document lifecycle, as topics can be individually reviewed by subject-matter experts, updated and translated independently of their use in some documents, thus reducing significantly publication time in some scenarios. Finally, documents assembled from consistently structured topics may be easier to read and navigate for their end users.

Structured authoring solutions have been proposed to improve the document creation and publishing processes, in particular by offering readers more consistent and structured information and by facilitating the work of technical writers through content reuse and assistance in document structuring. Nowadays, the large majority of large-scale publishing organizations have recourse to some structured authoring solution to produce high-quality and often multilingual documents. The more apparent limitation to current solutions is that there is no automatic control over the adequacy between semantic types used to describe document structure and the text fragments annotated by them. Indeed, the text entered by writers, although possibly in some controlled language, cannot be checked as being of the correct semantic type, this being inherent again to natural language interpretation capabilities. Machine learning with appropriate training data may provide some solution with *automatic text classification* (see e.g. Manning and Schütze 1999 or Paaß 2012). A drastic solution to overcome this limitation, *controlled document authoring*, will be presented in section 3.2.

3. Document Generation and Transformation by Computers

In this section, we present various applications of Natural Language Processing techniques that can produce text that can be integrated into a document creation cycle. We first briefly introduce the field of Natural Language Generation (NLG), which makes it possible to have computers generate text fully automatically. We then present Interactive Natural Language Generation, a research field in which users specify the content of documents in a controlled way and the computer generates multilingual versions of high quality documents. Finally, we describe some text transforming applications with a focus on their interactive uses.

3.1. Natural Language Generation

Some documents include the presentation of routine information, possibly on a regular basis. It is legitimate to expect that computers can play an important part in creating such documents, thereby relieving human writers from tedious activities. Delegating such a task to computers can further bring benefits of consistency, predictability, and rapidity inherent to machines. The field of *Natural Language Generation* (NLG) (see e.g. Bateman and Zock 2005) is traditionally concerned with the automatic production of textual documents from non-linguistic information. Many contexts require that some raw data be communicated and explained in written form when economic or time constraints do not allow allocating human resources to the task, possible scenarios including detailed alerts for whether events and diagnostic reports in software testing. NLG systems encode and use knowledge on a specific application domain as well as about a specific language. Difficult modeling issues need to be addressed, including the selection of what information should be presented and the characteristics defining high-quality documents. The requirements of NLG thus converge with the requirements of structured authoring and controlled languages, with the difference that specifications are intended to be used by computer programs rather than by human writers. Various approaches to generate text automatically have been proposed bearing various degrees of resemblance to activities of human writers. The most consensual approach is to divide the task in the following pipeline of processes (Reiter and Dale 2000):

1. *document planning*, which is concerned with determining what elements of information should be conveyed in a given document and with what expository structure;
2. *microplanning*, where lexical items are associated to concepts, information units are aggregated into messages, and referring expressions are chosen;
3. *surface realization*, where sentence specifications are built and surface operations such as word agreement are made.

NLG systems can occupy two different roles in the document creation process (Reiter and Dale 2000). In scenarios where documents contain both routine and more complex sections, NLG systems can be used to produce drafts which will be subsequently revised and possibly augmented by human writers. This leaves analytical work to writers while making the machine responsible for producing text which can be easily derived from raw data. The DRAFTER prototype system (Paris et al. 1995), for instance, aims at facilitating the work of technical authors writing software manuals by producing initial drafts. NLG systems can also be designed to produce final documents without human intervention, but this is mainly intended for basic document types and concerns mainly research prototype systems.

Considering the cost and complexity of building and fielding such systems, a natural concern is whether they can be viable economic alternatives to traditional document authoring. Reiter and Dale (2000) identify at least the following conditions for NLG solutions to qualify for consideration:

- the number and frequency of generated documents should be high enough to balance the costs incurred by the NLG system;
- the conformance to easily formalized standards should be a strong requirement, as machines can be programmed to consistently apply encoded rules;
- multilinguality requirements should be important, as the production of multilingual versions of documents can be produced by generation systems sharing modules and data;
- the unwillingness of writers for writing some types of stereotypical documents or the difficulty of obtaining documents at the imposed pace can justify automating the entire document creation process.

Early NLG systems were built with hand-made rules and encoded linguistic knowledge, but more recent approaches attempt to exploit language encoded in text corpora (Krahmer and Theune 2010). Hybrid systems can use statistical knowledge at various stages of the NLG process, including for example at the early stage of text structuring (Lapata 2003).

3.2. Interactive Natural Language Generation

One of the objectives of drafting documents with Natural Language Generation systems is to allow a symbiotic complementarity between machines and human writers. This can be extended by typically attributing content specification to a human expert and delegating microplanning and surface realization to a NLG system. This scenario is in fact particularly well suited to cases where content specifications do not already exist or cannot be straightforwardly derived from existing data. The process of document authoring can thus be redefined as specifying well-formed document descriptions in some formal representation which can then be used by a NLG system. Importantly, such document content representations being language-neutral by essence, they can naturally be used to produce documents in several languages. This therefore suggests a shift from multilingual document production involving monolingual authoring, followed by several independent translations, to document production through *document content authoring*, followed by multilingual automatic generation (Hartley and Paris 1997). More generally, *interactive Natural Language Generation* can involve human intervention at any step of the document generation process, as for example in fine-grained lexical choices reflecting complex analytical reasoning.

The *What You See Is What You Meant* (WYSIWYM) paradigm (Power et al. 1998) instantiates this principle by allowing a human author to interact with an evolving text reflecting the current state of a *domain model* that is iteratively populated by authoring choices. A NLG system produces a *feedback text* from the domain model, which allows the author to both check that the encoded knowledge ("What You Meant") is correct and to further refine it by choosing semantically possible values from menus associated to *anchors* in the text. For example, an initial step in a software manual might correspond to the text "**Do this action** by using *these methods*", where bold phrases in the feedback text indicate mandatory elements and italic phrases optional elements, which would typically not appear in the final form of the text if unspecified. The author might then decide to specify a value for the **Do this action** anchor, for which the system can propose several values organized in a menu (e.g. *save*, *schedule*, *start*). Selecting the *schedule* value would first update the domain model appropriately, and then trigger a regeneration of the feedback text to reflect the new state of the domain model, for example with the sentence "Schedule **this event** by using *these methods*."

This type of *controlled authoring* guarantees that documents are structurally and linguistically well-formed and makes multilingual document generation available at the reasonable cost of building NLG systems for several languages, but without the complex steps of content selection and structuring. The Multilingual Document Authoring (MDA) approach (Dymetman et al. 2000) is an evolution of WYSIWYM which allows a greater control on the specification of well-formed typed elements and of their semantic dependencies. For instance, in a drug leaflet the specification of a pharmaceutical form for a drug can impose constraints on the possibles values for modes of administration, making it impossible to define incoherent document content. Furthermore, an appropriate granularity for describing domain-specific elements can greatly limit the quantity of work by the author: simply indicating that a drug contains aspirin may automatically produce an element that will be rendered as an entire paragraph on the risks of taking aspirin in various conditions.

In spite of its apparent advantages as regards document well-formedness, document updating, and translation, interactive NLG suffers from limitations that until now have limited its use to prototype systems. The most important ones concern the difficulty of specifying, developing and updating the resources used by such systems, the impossibility of easily importing existing written documents or of entering free text, and the limited expressivity allowed to authors, who may not considered themselves as authors of the produced texts in the traditional sense. In fact, changes to work practices and redistribution of roles between man and machine may have been to date the most difficult limitations to overcome.

3.3. Text transformation by computer

Natural Language Generation was initially defined as the automatic production of textual documents from information encoded as raw non-textual data. Some scenarios however naturally call for reusing knowledge already encoded in text. In fact, human writers often use some existing documents to derive new forms of documents. The most notable applications of text reuse and transformation include the following activities:

- in *text summarization*, a text is made more concise for quick reading;
- in *text compression*, a text is shortened so that it occupies less space, for example to be used on mobile devices;
- in *information fusion*, a new text is compiled from information presented in several texts;
- in *text simplification*, a text is tailored to meet the requirements of certain readers, for example readers with reading disabilities or language learners;
- in *text normalisation*, a text is transformed to be more uniform and coherent and possibly to comply with norms for specific document types.

Automating these processes requires implementing computer programs that have the ability to generate new text from some existing text by applying a transformation function. By far the main difficulty of this process is ensuring that meaning is correctly preserved or adequately transformed from one text version to the other, while ensuring that a generated text is locally grammatically correct can be achieved to a reasonable extent (Wan et al. 2009). The family of works in this area are often described as instances of *text-to-text NLG* (see e.g. Filippova and Wan 2011)[15]. In the vast majority of cases, text-to-text applications do not resort to generic NLG systems, because deriving the information representations required by such systems through automatic text analysis is out of reach of current technologies.

Automatic summarization (see e.g. Hovy 2005) has attracted a lot of work both for the sentence selection and sentence synthesis approaches. However, early attempts at integrating such systems in word processors led to low adoption by users, which was probably the consequence of immature technology at the time and inappropriate expectations from potential users. The field of *multi-document summarization* led to promising prototypes, which for instance extract related sentences from comparable documents, such as news stories on the same topics, and apply fusion algorithms to generate new sentences (Barzilay and McKeown 2005). *Text simplification* has been initially tackled with rule-based approaches intended to simplify the complexity of sentences and improving the coherence of documents (Siddharthan 2006), and has more recently been addressed with statistical methods, for example by acquiring knowledge from the Simple English Wikipedia[16] (Zhu et al. 2010). *Text normalization* has been addressed for documents in constrained domains, such as drug leaflets, by reusing the resources of controlled authoring systems to both analyse the content of a raw document and to generate a normalized version of the document (Max 2004). Attempts at modeling common aspects of several applications have been made in the context of *statistical paraphrase generation* (Zhao et al. 2009), borrowing methods from Statistical Machine Translation. This is currently a very active research area focussing on data-driven methods (Madnani and Dorr 2010) which is likely to produce more mature technology in the years to come, in particular if significant progresses are made on the issues of text analysis and meaning preservation during transformation.

4. Writing with computers: Future directions

On December 25, 2003, the New York Times published a story about recent research results that made it possible for computers to generate paraphrases of English sentences.[17] One mentioned potential application was to help writers adapt their text to different groups of readers. Although the technology, based on statistical techniques, was at its early stages, the mention of it in the media probably revealed that such a possibility was both hoped for and unexpected. Most professional writers nowadays accept the fact that machines can handle presentation aspects of documents, and maybe at best that they can be used to spot likely spelling errors. However, the type of function just described, which could be perceived as *machine intelligence* by computer users as it poses several complex challenges requiring language sensitivity (Dale 1997) from the machine, is probably not expected to be effective in the near future. Various research areas that we have mentioned in the previous sections can in fact be used in the "augmentative scenario" predicted by Dale, in which the machine assists the human writer in performing existing tasks. The latter could concentrate on the most important matters in text authoring, drawing an analogy with what happened several decades ago with the introduction of the separation of content from presentation.

We wrap up this chapter by citing technology areas that may soon provide effective assistance to human writers, which can by no means be an exhaustive review. The first area is that of *assistance in writing and revision*. Writers and even expert writers are often looking for the most appropriate words to convey their messages. Assistance can be proposed by enhancing dictionaries with an additional index based on the notion of association between words (Ferret and Zock 2006), by proposing paraphrases in context for selected text fragments (Max 2008), or by suggesting more appropriate word choices through collocation checking (Wu et al. 2010). The hypothesis that existing text fragments are likely to be useful when writing new text can be exploited by statistical language models of likely phrase completions, as implemented in the now discontinued Google Scribe prototype system.[18] More generally, the concept of *authoring memories* could experience a similar success to what translation memories still experience in the translation industry, in particular in technical documentation as started with topic-based authoring.

The second area concerns *assistance in text structuring*. Semantic similarity between text fragments can be used to propose links across documents and passages (Knoth et al. 2010), helping writers to organize related ideas expressed as text. Requirements on the transition between sentences and document characteristics can be modeled to evaluate text coherence, in order to spot text fragments that are likely to be perceived as incoherent (Lapata and Barzilay 2005). Furthermore, useful tasks could be partly automated, such as building table-of-contents for complex documents (Branavan et al. 2007), or building back-of-the-book indexes by identifying important concepts in documents (Csomai and Mihalcea 2008). Lastly, new *collaborative authoring tools* such as wikis are likely to be enhanced with language technologies in various ways for finding, adding and structuring textual content (Hoffart et al. 2009).

Notes

1. Note that in this chapter only the textual parts of documents are considered, and not the aspects concerned with actual layout and presentation including formatting and images, which are typically managed by *publishing software*.
2. An important consideration that we will not discuss in detail in this chapter is the level of expertise and familiarity of the writer with the language in which a document has to be written: important developments are indeed under way in the domain of foreign language writing (as well as reading) assistance, where the expected level of assistance is not the same as when writing in one's mother tongue.
3. E.g. GNU EMACS: http://www.gnu.org/software/emacs.
4. http://unicode.org/standard/standard.html
5. E.g. OPEN OFFICE WRITER: http://www.openoffice.org/product/writer. html.
6. Note that some authoring software provide both the convenience of WYSIWYG editing and the use of the advanced capabilities of complex typesetting systems.
7. E.g. FREEMIND: http://freemind.sourceforge.net
8. Synonyms are frequently used to avoid redundancy in word usage, but their use is unanimously discouraged by technical writing guidelines.
9. Indeed, many technical and scientific writers who have to write in English are not native speakers.
10. It is to true, to some extent, that simpler documents can reduce the *need* for translation, as readers who are not native speakers of the language of a document may find them accessible. It is however reasonable to see the translation of all technical documents as a desirable objective.
11. http://http://www.asd-ste100.org/
12. http://www.iso.ch/cate/d16387.html; see also Stührenberg (2012) in this volume.
13. http://www.w3.org/XML/
14. http://dita.xml.org/
15. Note that, in this respect, Machine Translation is one type of text-to-text application, but involving two different languages.
16. http://simple.wikipedia.org
17. http://www.nytimes.com/2003/12/25/technology/circuits/25next.html: *WHAT'S NEXT; 'Get Me Rewrite!' 'Hold On, I'll Pass You to the Computer.'*
18. http://googleresearch.blogspot.fr/2011/05/google-scribe-now-with-automatic-text_26.html

References

Alred, Gerald J., Charles T. Brusaw and Walter E. Oliu
2008 *Handbook of Technical Writing*. St. Martin's Press.
Barrett, Edward (ed.)
1991 *Text, ConText and HyperText*. MIT Press, Cambridge, USA.
Barzilay, Regina and Kathleen McKeown
2005 Sentence fusion for multidocument news summarization. *Computational Linguistis*, 3(31): 297–328.

Bateman, John and Michael Zock.
2005 Natural language generation. In Ruslan Mitkov, editor, *The Oxford Handbook of Computational Linguistics*, chapter 15. Oxford University Press, Oxford, UK.

Branavan, S. R. K., Pawan Deshpande and Regina Barzilay
2007 Generating a table-of-contents. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics (ACL'07)*, Prague, Czech Republic.

Csomai, Andras and Rada Mihalcea
2008 Linguistically motivated features for enhanced back-of-the-book indexing. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics (ACL'08)*, Columbus, USA.

Dale, Robert
1997 Computer assistance in text creation and editing. In Ron Cole, Joseph Mariani, Hans Uszkoreit, Annie Zaenen, and Victor Zue, editors, *Survey of the State of the Art in Human Language Technology*, chapter 7.5. Cambridge University Press, Cambridge, UK.

Dymetman, Marc, Veronika Lux and Aarne Ranta
2000 XML and multilingual document authoring: Convergent trends. In *Proceedings of the 18th International Conference on Computational Linguistics (COLING'00)*, Saarbrucken, Germany.

Ferret, Olivier and Michael Zock
2006 Enhancing electronic dictionaries with an index based on associations. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics (COLING-ACL'06)*, Sydney, Australia.

Filippova, Katja and Stephen Wan (eds.)
2011 *Proceedings of the Workshop on Monolingual Text-To-Text Generation*. Association for Computational Linguistics, Portland, USA.

Flower, L. and J. R. Hayes
1980 The dynamics of composing: Making plans and juggling constraints. In L. W. Gregg and E. R. Steinberg, editors, *Cognitive processes in writing*. Erlbaum, Hillsdale, USA.

Halvorsen, Per-Kristian
1997 Document processing. In Ron Cole, Joseph Mariani, Hans Uszkoreit, Annie Zaenen, and Victor Zue, editors, *Survey of the State of the Art in Human Language Technology*, chapter 7. Cambridge University Press, Cambridge, UK.

Hartley, Anthony F. and Cécile L. Paris
1997 Multilingual document production – from support for translating to support for authoring. *Journal of Machine Translation*, 12(1): 109–128.

Hoffart, Johannes, Torsten Zesch and Iryna Gurevych
2009 An architecture to support intelligent user interfaces for wikis by means of natural language processing. In *Proceedings of the 5th International Symposium on Wikis and Open Collaboration (WikiSym'09)*, Orlando, USA.

Hovy, Eduard
2005 Text summarization. In Ruslan Mitkov, editor, *The Oxford Handbook of Computational Linguistics*, chapter 32. Oxford University Press, Oxford, UK.

Kittredge, Richard I.
2005 Sublanguages and controlled languages. In Ruslan Mitkov, editor, *The Oxford Handbook of Computational Linguistics*, chapter 23. Oxford University Press, Oxford, UK.
Knoth, Petr, Jakub Novotny and Zdenek Zdrahal
2010 Automatic generation of inter-passage links based on semantic similarity. In *Proceedings of the 23rd International Conference on Computational Linguistics (COLING'10)*, Beijing, China.
Krahmer, Emiel and Mariët Theune (eds.)
2010 *Empirical Methods in Natural Language Generation: Data-oriented Methods and Empirical Evaluation*, volume 5790 of *Lecture Notes in Computer Science*. Springer.
Kukich, Karen
1992 Techniques for automatically correcting words in text. *ACM Computing Surveys*, 24(4): 377–439.
Lapata, Mirella
2003 Probabilistic text structuring: Experiments with sentence ordering. In *Proceedings of the 41st Meeting of the Association of Computational Linguistics (ACL'03)*, Sapporo, Japan.
Lapata, Mirella and Regina Barzilay
2005 Automatic evaluation of text coherence: Models and representations. In *Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence (IJCAI'05)*, Edinburgh, Scotland.
Madnani, Nitin and Bonnie J. Dorr
2010 Generating phrasal and sentential paraphrases: A survey of data-driven methods. *Computational Linguistics*, 26: 341–387.
Manning, Chris and Hinrich Schütze
1999 *Foundations of Statistical Natural Language Processing*. MIT Press, Cambridge, USA.
Max, Aurélien
2004 From controlled document authoring to interactive document normalization. In *Proceedings of the 20th International Conference on Computational Linguistics (COLING'04)*, Geneva, Switzerland.
Max, Aurélien
2008 Local rephrasing suggestions for supporting the work of writers. In *Proceedings of the 6th International Conference on Natural Language Processing (GoTAL'08)*, Gothenburg, Sweden.
Max, Aurélien and Guillaume Wisniewski
2010 Mining naturally-occurring corrections and paraphrases from Wikipedia's revision history. In *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta.
Menke, Peter
2012 Evaluation of technical communication. In Alexander Mehler, Laurent Romary, and Dafydd Gibbon, editors, *Handbook of Technical Communication*. De Gruyter, Berlin/Boston.
Newbold, Neil and Lee Gillam
2010 The linguistics of readability: The next step for word processing. In *Pro-*

ceedings of the Workshop on Computational Linguistics and Writing at the 11th Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL'10), Los Angeles, USA.

Olive, Thierry, Ronald T. Kellogg and Annie Piolat
2002 The triple-task technique for studying the process of writing. In T. Olive and C.M. Levy, editors, *Contemporary tools and techniques for studying writing*. Kluwer Academic, Dordrecht, the Netherlands.

Olive, Thierry, Ronald T. Kellogg and Annie Piolat
2008 Verbal, visual and spatial working memory demands during text composition. *Applied Psycholinguistics*, 29: 669–687.

Paaß, Gerhard
2012 Document classification, information retrieval, text and web mining. In Alexander Mehler, Laurent Romary, and Dafydd Gibbon, editors, *Handbook of Technical Communication*. De Gruyter, Berlin/Boston.

Paris, Cécile, Keith Vander Linden, Marcus Fischer, Anthony Hartley, Lyn Pemberton, Richard Power and Donia Scott
1995 A support tool for writing multilingual instructions. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence (IJCAI'95)*, Montréal, Canada.

Pitler, Emily and Ani Nenkova
2008 Revisiting readability: A unified framework for predicting text quality. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP'08)*, Waikiki, USA.

Power, Richard, Donia Scott and Roger Evans
1998 What You See Is What You Meant: direct knowledge editing with natural language feedback. In *Proceedings of the European Conference on Artificial Intelligence (ECAI'98)*, Brighton, UK.

Puerta Melguizo, Mari Carmen, Lou Boves and Olga Muñoz Ramos
2009 A proactive recommendation system for writing: Helping without disrupting. *International Journal of Industrial Ergonomics*, 39(3): 516–523.

Rahtz, Sebastian
2012 Representation of documents in technical communication. In Alexander Mehler, Laurent Romary, and Dafydd Gibbon, editors, *Handbook of Technical Communication*. De Gruyter, Berlin/Boston.

Reiter, Ehud and Robert Dale
2000 *Building Natural Language Generation Systems*. Cambridge University Press, Cambridge, UK.

Richardson, Stephen D. and Lisa C. Branden-Harder
1988 The Experience of Developing a Large-scale Natural Language Text Processing System: CRITIQUE. In *Proceedings of the 2nd Conference on Applied Natural Language Processing (ANLP'88)*, Austin, USA.

Sasaki, Felix and Arle Lommel
2012 Multilingual computing. In Alexander Mehler, Laurent Romary, and Dafydd Gibbon, editors, *Handbook of Technical Communication*. De Gruyter, Berlin/Boston.

Sharoff, Serge and Anthony Hartley
2012 Lexicography, terminology and ontologies. In Alexander Mehler, Laurent Romary, and Dafydd Gibbon, editors, *Handbook of Technical Communication*. De Gruyter, Berlin/Boston.

Siddharthan, Advaith
2006 Syntactic simplification and text cohesion. *Research on Language and Computation*, 4(1): 77–109.

Stührenberg, Maik
2012 Foundations of markup languages. In Alexander Mehler, Laurent Romary, and Dafydd Gibbon, editors, *Handbook of Technical Communication*. De Gruyter, Berlin/Boston.

Trippel, Thorsten
2012 Controlled language structures in technical communication. In Alexander Mehler, Laurent Romary, and Dafydd Gibbon, editors, *Handbook of Technical Communication*. De Gruyter, Berlin/Boston.

Walsh, Norman
2010 *DocBook 5: The Definitive Guide*. O'Reilly Media, Sebastopol, CA.

Wan, Stephen, Mark Dras, Robert Dale, and Cécile Paris
2009 Improving grammaticality in statistical sentence generation: Introducing a dependency spanning tree algorithm with an argument satisfaction model. In *Proceedings of the 12th Conference of the European Chapter of the ACL (EACL'09)*, Athens, Greece.

Wilkinson, Ross, Timothy Arnold-Moore, Michael Fuller, Ron Sacks-Davis, James Thom and Justin Zobel (eds.)
1998 *Document Computing*. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Wu, Jian-Cheng, Yu-Chia Chang, Teruko Mitamura and Jason S. Chang. Automatic collocation suggestion in academic writing. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL'10)*, Uppsala, Sweden, 2010.

Zhao, Shiqi, Xiang Lan, Ting Liu and Sheng Li
2009 Application-driven statistical paraphrase generation. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP (ACL-IJCNLP'09)*, Singapore.

Zhu, Zhemin, Delphine Bernhard and Iryna Gurevych
2010 A monolingual tree-based translation model for sentence simplification. In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling'10)*, Beijing, China.

2. Representation of Documents in Technical Communication

Sebastian Rahtz

This chapter considers standards and models for the representation of textual content, both for the purpose of the organisation and presentation of content. It provides an overview of the characteristics and usage of most commonly used document markup formats (Office Open XML (OOXML), Open Document Format, TeX/LaTeX, DocBook, Text Encoding Initiative (TEI), XHTML, etc.) with a view on their scope and application. A central theme of the chapter is a clarification of the notion of a *document* in relation to that of *text*. The chapter will overview issues in representing the structure, content and function of documents. Further, it reviews document formats in the light of their support of these representation tasks.

It should be noted that, for convenience, this chapter deals entirely with documents and texts managed in electronic form, as is now almost universally the case. Many of the arguments would apply, however, to older analogue information as well.

1. Texts, documents and data

When does text form a document? What are the key characteristics of a structured document, as opposed to structured data or a stream of words? Do documents always consist of words, or can a set of pictures form a document?

There are at least *five* layers of conceptual activity on the spectrum from atomic building blocks to a document.

1. At the bottom level are numbers, letters, image pixels, audio fragments from which we can build all other artifacts (leaving aside for this purpose how those building blocks are stored or a computer).
2. At the next level these blocks are formed into semantically-meaningful artifacts (sentences, pictures, tables, decipherable fragments of audio).
3. The semantic blocks are combined, at some moment in time, to form a storage unit (a file, a database record, a construct in a computer program).
4. One or more storage units presented to a reader or viewer creates a ‘text’, a coherent object which can be described, has metadata, and is created by some rational process.
5. When a ‘text’ has a coherent *purpose*, we may call it a ‘document’.

It is clear that appearance alone, or storage format, or data type, does not dictate to us that we are dealing with a document. If we consider four typical streams of information that we meet on the web daily:

1. an RSS feed of news about rapidly changing current events;
2. a page of images from Flickr resulting from a search for ‘white cats’;
3. information about the current state of share prices;
4. our Inbox of unread email messages.

they are structurally similar to:

1. a set of entries from a scientist’s logbook;
2. a sequence of photos showing the growth of a cell over time;
3. statistical analysis of word usage in the works of Rudyard Kipling;
4. publication of the correspondence between two star-crossed lovers.

and may be represented using the same tools and structures on our computer. However, the first set lack the characteristics which would let us describe them as coherent documents, whereas the second set have that extra *purpose* which makes them more than a text.

It should be noted that this taxonomy, distinguishing a ‘text’ from a ‘document’, is not universal; others may distinguish, for example, between a literal transcription of a printed work (a ‘document’), and an edition of that work (a ‘text’). However, for the purpose of this chapter (and other parts of this book), we choose to distinguish early on between the storage of data in digital form (the bits and bytes and other structures), their management on computers (files, etc.), and the intellectually-complete communicating object which is the subject of much of this book.

We conclude this section by asserting that in the field of technical documentation a ‘document’ (and we will omit the ‘emphasis’ in the remainder of this chapter) has at least some of the following characteristics: it has a coherent *purpose*; it has identifiable *authorship* and associated *metadata*; it has a *beginning* and an *end*; and, finally, it has some *structure*. That there may be *several* overlapping structures visible to readers (or even to authors) does not invalidate the necessity of structure.

A document can be created from any useful types of data, combined in a variety of ways, but if it possesses none of these characteristics, it probably falls outside the scope of this chapter. For this book, of course, we may also add the characteristic that the document is *information bearing*, but that may be regarded as an optional extra by some.

1.1. The idea of a document

In the previous section, we requested that a document has some *structure*. Conventionally, this will be divided into two parts, the metadata (bibliographical information about the authorship, provenance, classification, etc. of the document) and the main information bearing component. The other chapters in this book will consider in much more detail the meaning of the inner components of text, but here it will be helpful to outline the major components, which in subsequent sections we will compare representations of:

– introductory matter (preface, acknowledgements, summary, etc.);
– main structure, often divided into several ‘chapters’;
– end matter (bibliography or references, notes, etc.).

In the field of scientific documents, we should also be aware of the important category of ‘associated data’, raw material which provides the context or justification of assertions in the text (see also Romary 2012 in this volume).

It should be clear from the description of components above that the purpose of the structure in a document is to convey *semantic* information within the document. We may recall at this point the debate which lasted through the late 1980s and the 1990s, on the question of ‘what is text’, in which various writers (largely centred around Alan Renear and Steve DeRose) theorized over describing what could be seen in an existing document, and the extent to which it would be reduced to an OHCO (Ordered Hierarchy of Content Objects, codified in SGML, and exemplified by the schemes discussed below). These influential papers (e.g. DeRose et al. 1990, DeRose et al. 1997, Hockey et al. 1999, Renear 1997, and Renear et al. 1996; the philosophers of the digital humanities have returned to the theme many times since) clarified and established for a generation of text encoders the primacy of semantic markup. Later writers refined ideas about ‘text’, establishing that it is a far more fluid affair, especially in the multiple, overlapping, views which can be asserted of it, and showed the limitations of schemes such as XML.

For the discussion in this chapter of technical communication, we are probably safe within the stable OHCO universe.

1.2. Explicit and implicit semantic information

One of the major problems in dealing with the instantiation, as opposed to the idea, of a document is the (necessary) use of visual clues to semantic structure in the work as presented to most readers or viewers. While we said in the previous section that a document often has a set of chapters, we only see this by its representation in the page. We deal with a visual set of conventions, derived from centuries of manuscripts and books.

This is well exemplified by a 19th century edition of Euclid, shown in Figure 1[1], which shows a 4-level hierarchy of content:

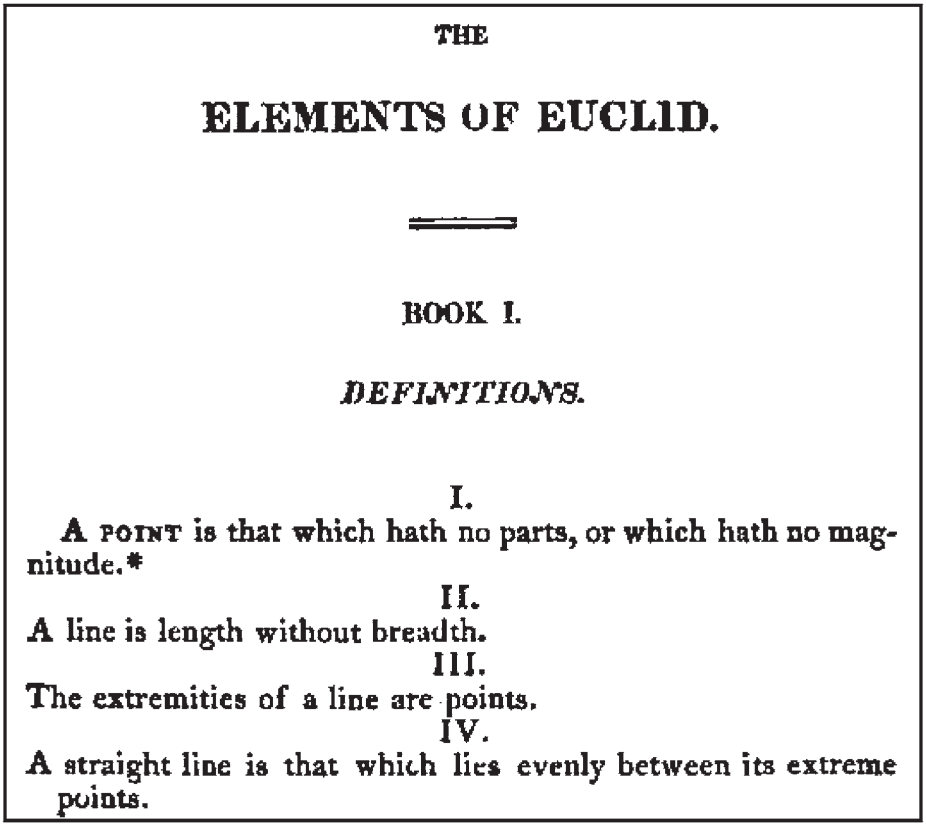
1. The work: **ELEMENTS OF EUCLID**, using centring, vertical white space, larger font;
2. Book 1: **BOOK I.**, using centring, vertical white space, separating rule, numbering, and capitals;
3. Set of definitions: ***DEFINITIONS.***, using centric, italic font and capitals;
4. Definition: **I.**, using centring and numbering.

The use here of whitespace (horizontal and vertical), font changes, colour, and numbering as part of the content, are all relying on several centuries of conventions (since Gutenberg at least); e.g. that an italic heading is usually of lesser importance than one in bold – though the 'rules' are not explicitly stated. Contrast the 19th century Euclid with the 3rd century AD Greek papyrus (Figure 2[2]) which lacks almost any visual clue (even as to word breaks), and a 10th century manuscript where the scribe has added word breaks but little else.

In the analog form represented above in the papyrus fragments and early book, we have little choice but to use visual clues, because the scribe or typesetter has to use the *same method* to convey both the structure and the content. An alternative in the digital medium is to use *hidden marks* to differentiate true content from content providing meta-information about the semantics of the content. In practical terms, this means using some technique to associate a set of words with a particular intention. Here we are concerned to establish the difference between a human reader following a stream of characters in various sizes and forms, separated by white space, and a mediating computer program which reads a stream of data and instructions and presents a view to the human eye based on the interpretation of those instructions.

It may be noted, of course, that the representation of the structure need not involve a linear stream of interspersed instructions and data (markup and text): the same effect can be achieved using data storage techniques (i.e. a database), in which words (or images, or sounds) are stored in named locations, and a second stream of data says what the locations mean. For the purposes of this chapter, however, we will stay with the simpler notion of the linear stream contained within a computer file.

We have arrived at the notion that the document is represented by a linear stream of data, of which some components are content ('words') and some are instructions, or markup; and that the markup will instruct a viewing medium on how to present the words to the reader in a manner familiar to them.



THE

ELEMENTS OF EUCLID.

BOOK I.

DEFINITIONS.

I.

A POINT is that which hath no parts, or which hath no magnitude.*

II.

A line is length without breadth.

III.

The extremities of a line are points.

IV.

A straight line is that which lies evenly between its extreme points.

Figure 1. Visual clues to structure in Euclid.

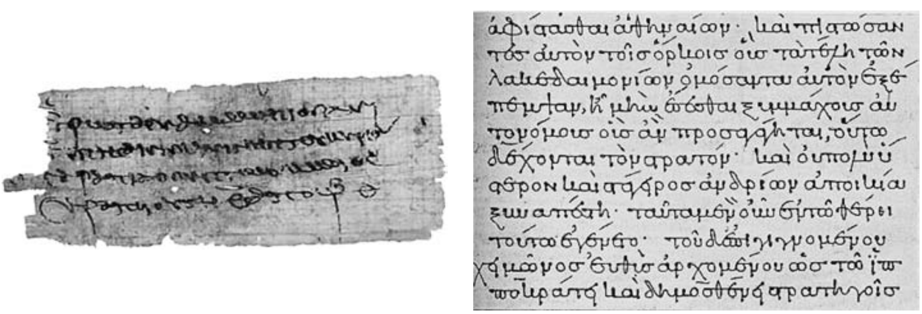


Figure 2. Greek papyrus, Oxyrhynchus Papyri vol. LXII no. 4339 P. Oxy; and 10th century manuscript copy of Thucydides.

Table 1. Commonly-used visual techniques to differentiate text components.

Technique	Example
punctuation	this is – or so we believe – the truth; yet we can never know.
white space (vertical or horizontal)	
	– on the one hand – on the other
change of font	the `fixed width font` is used to mimic the old typewriter or teletype
change of font style (italic, bold, small-caps, etc.)	the oldest *dog* in the world, however
change of font size	We may never know what the mouse thought about it
colour	Whatever you do, do not open the bottle marked POISON
page or screen breaks	

We may note that this does not mean a simple 1:1 relationship between markup and visual form. There are, for examples, minor differences like the use of *“semantics”* or *«semantics»* or *‚semantics‘* to distinguish a quoted word, high-level cultural differences like reading direction, or (more importantly) the use of a medium like sound, or Braille marks (see Kubina and Lücking 2012 in this volume), to convey distinctions in text. Many of the familiar visual techniques (see Table 1) do come from trying to record patterns of speech (most obviously the use of white space to represent pauses in speech).

What manner of instructions may we insert in the stream of data? There are two distinct models available. Firstly, we can use generalised language to describe the traditional visual effects, in more or less vague terms. Instructions like ‘use Helvetica font 8pt bold with 10pt leading’ or ‘leave 1 cm of white space’ are unambiguous, whereas ‘emphasise this word’, ‘center the title on the page’ and ‘number the item’ are more open to different interpretations. They share the characteristic, however, that they are telling a computer programme *what* to do, not *why*. Alternatively, we can use instructions which attempt to explain *why* a group of words is different from its surroundings, without saying how that should be conveyed to the reader. For example, a simple instruction saying ‘this is a paragraph’ is left open to several effects (Figure 3; text from Rahtz and Rowley 1984).

Using the latter model, the markup must be accompanied by a ‘stylesheet’. This term has been used since the start of work on SGML (ISO, 1986) to describe a second (or more) set of instructions which tell a processor how the

Although site 1 appears to be an area excavation and the remainder appear to be trenches of various dimensions the technique employed in all cuttings was the same. The whole of the site was gridded and each cutting was dug in layers.

Features were related to layers as work progressed, finds were recorded on the basis of layer and feature and small finds were additionally plotted three-dimensionally.

Although site 1 appears to be an area excavation and the remainder appear to be trenches of various dimensions the technique employed in all cuttings was the same. The whole of the site was gridded and each cutting was dug in layers.

Features were related to layers as work progressed, finds were recorded on the basis of layer and feature and small finds were additionally plotted three-dimensionally.

Although site 1 appears to be an area excavation and the remainder appear to be trenches of various dimensions the technique employed in all cuttings was the same. The whole of the site was gridded and each cutting was dug in layers.

Features were related to layers as work progressed, finds were recorded on the basis of layer and feature and small finds were additionally plotted three-dimensionally.

Figure 3. Variations in how paragraphs are displayed.

markup should be formatted; an ISO standard, DSSSL (ISO, 1994) was defined for this in the 1990s, but it was never fully implemented, and the term is now used more commonly in the context of XSL and CSS (see below). A stylesheet clearly provides a more flexible way of working. But even when using this model, there remain layers of possible semantic distinctions we can add. Italic *emphasis*, for example, is a relatively neutral technique indicating a word or phrase different in some way, and we may often explain more about *why* it is distinct (because it is a name, a title, or a Latin work, for example).

1.3. Describing or prescribing?

The combination of abstract markup and a separate interpretative stylesheet is attractive and powerful, and is widely used in modern implementations of technical documentation (although often accompanied by direct visual formatting for when an author wants to do something very specific). For it to be used with any degree of sustainability, there must be agreement on the vocabulary of the markup.

Any markup vocabulary will offer ways of representing common artifacts like text divisions, paragraphs, lists, notes, figures, bibliographies and titles. They may differ, however, on whether they regard these instructions as a framework into which text is placed, or a language to be used to describe what we see on a page. We may distinguish between

Prescriptive The text shall have a front matter with title and author, following by one or more sections, an optional set of notes, and a bibliography.
Descriptive There are such things as text sections, title statements, and notes, you may combine them as you wish.

The distinction is often made between *new writing* ('born digital') which is designed to fit a framework, and *markup of existing material* ('digitized') where a third party is trying to describe the intention of an author. It becomes important when we consider schemes to *validate* documents in the next section.

2. Models for documents

It is now time to look at the actual methods used to represent markup in technical documentation. There are three dominant models currently popular:

1. the *sui generis* TeX (Knuth, 1984) language devised by Donald Knuth, and widely used in its LaTeX (Lamport, 1986) variant – we will examine this later;
2. proprietary binary codes in the stream of words, most clearly instantiated by older versions of word-processor formats, such as Wordperfect or Microsoft Word's .doc format – we will not say any more about this, as it is entirely mimicked by the more recent XML-based versions;
3. formats based on XML, such as Text Encoding Initiative (TEI), OpenXML (ISO, 2008), used current Microsoft Office documents, OpenOffice's Open Document Format (ISO, 2006), and most text markup schemes.

The last of these, Extensible Markup Language (XML) (Bray et al., 2008), is a simple, very flexible, text format derived by the World Wide Web Consortium from SGML (ISO, 1986), (Goldfarb, 1990) in 1996; it is by some degree the most dominant and accepted format for many varieties of text exchange (even when it is delivered in a compressed package, as Microsoft Office does). For the purposes of this chapter, the language of web pages, HTML (Altheim and McCarron, 2001), can be regarded as an instantiation of XML (see also Stührenberg 2012 in this volume).

2.1. The TeX family

The TeX family of markup languages is in a unique position, for several reasons. The language and its implementation[3] are tightly bound together, so that effectively all users run the same software, with very limited interchange possibilities.[4] TeX implements a very rich and powerful typesetting language

```
\documentclass[11pt,twoside]{article}
\usepackage{times,hyperref}
\date{January 17, 1999}
\title{Simulation of Energy Loss Straggling}
\author{Maria Physicist}
\begin{document}
\maketitle
\section{Introduction}\label{intro}

\par Due to the statistical nature of ionisation energy loss,
large fluctuations can occur in the amount of energy deposited
by a particle traversing an absorber element. Continuous
processes such as multiple scattering and energy loss play a
relevant role in the longitudinal and lateral development of
electromagnetic and \emph{hadronic showers}, and in the case
of sampling calorimeters the measured resolution can be
significantly affected by such fluctuations in their
active layers.

\par An example formula
\begin{equation}
\xi / I_0
\end{equation}
\par The work at \href{http://public.web.cern.ch/public/}{CERN}
is likely to be of interest to readers of this paper.
\end{document}
```

Figure 4. A simple document in LaTeX.

(especially in the area of mathematics), which can produce high-quality printed or PDF output. However, the input code is *also* a programming language, which allows for high-level markup languages to be developed on top of it.

The best known language written using TeX is LaTeX (Lamport, 1986), which defines a notation to address the same high-level constructs as we listed in the previous section. The expressiveness of this is extended by a very well-developed community of add-on packages, allowing the author to describe everything from music to chess games to complex bibliographies all in the same system, and receive reliable output. LaTeX also has important external tools to assist with smart editing, bibliography formatting and citation management (BibTeX), and index term management (makeindex). Combined with the sophistication of the underlying math engine, this has meant LaTeX is adopted very widely for preprints, theses, and scientific book typesetting. A simple LaTeX document is listed in Figure 4.

LaTeX is a tightly-coupled system. The same language is used to describe both what is required ('centre this paragraph and set it in smaller than the surrounding paragraphs') and to program interaction with the output page ('if there is less than 3 inches left on the page, set the text slightly smaller'). This built-in algorithmic rule-based working is quite distinct from either the separate stylesheet approach of the XML languages we shall discuss in the next sections, or with the working of word-processor systems like Microsoft Word which assume human *ad hoc* interaction.

The problems with LaTeX are, however, serious. Its extensibility and macro-programming ability make it very hard to do meaningful validation of the structure, or process it with other generic tools.[5] The relative difficulty of generating high-quality web pages, or summarising semantic information, for example, as RDF, seem to place LaTeX at one end of a write-only cul-de-sac.

2.2. Concepts of XML

XML provides a general model of structured data represented as strings of text. It has a simple convention that all instructions are enclosed in < and > characters, and has two important characteristics: firstly, the set of instructions is not prescribed, allowing anyone to define a vocabulary; and secondly, there is an added concept of a schema, or set of rules, for a vocabulary, which allows a document to be checked or *validated* (for this notion see also Stührenberg 2012 in this volume).

An XML document represents a tree. It has a single root and many nodes. Each node can be a subtree, a single *element* (possibly bearing some *attributes*) or a string of *character data*. There are also *comments*, *processing instructions*, *entity references* and *marked sections*, but these are obscure and rarely encountered. The main tool is the *element* (container) and its *attributes*; the basic syntax is shown in the following:

```
<?xml version="1.0" encoding="utf -8"?>
<rootElement>
    <elementName attributeName="attributeValue">
        elementContent
    </elementName>
    <!-- this is a comment -->
</rootElement>
```

Thus an XML document is encoded as a linear string of characters. It begins with a special instruction (the XML declaration) which specifies that this *is* an XML document, and which version of the XML standard it follows; it may also specify a different character encoding for the document. Elements are then marked by start- and end-tags using < and > around a name. Comments are delimited by <!-- and --> and attribute name/value pairs are supplied on the start-tag and may be given in any order.

The elements can be in different *name spaces*, to allow for the same name being found in different vocabularies, by using a special attribute xmlns. The following example shows how math markup is differentiated by its namespace:

```
<p>This is some text, and now
    some math:
    <m:math xmlns:m="http://www.w3.org/1998/Math/MathML">
        <m:msub>
```

```
            <m:mrow>
                <m:mi>p</m:mi>
            </m:mrow>
            <m:mrow>
                <m:mi mathvariant="normal">T</m:mi>
            </m:mrow>
        </m:msub>
    </m:math>
</p>
```

Which elements are allowed (<p>, <math>, <msub>, <mi>, <mrow> in the previous example), and in what order or combination, is dictated by a separate *schema* document, which can be written in a variety of notations. Checking the document against the schema is called *validation*. The most common languages are Document Type Definition (DTD) language of XML, XML Schema (Thompson et al., 2004), and RELAX NG (ISO, 2003); they let the vocabulary designer say:

1. What the root element may be;
2. What other elements, or character data, are allowed within a given element;
3. What attributes are allowed, and whether they are optional or mandatory;
4. What sort of data is allowed inside attributes or elements (i.e. numbers, dates, URIs, choices from a fixed list).

The intention of this is very similar to that of a database schema, determining what sort of data can be put into a container, and whether it is mandatory to do so. Figure 5 shows a very simple schema, written in RELAX NG, and an XML document valid against it.

Each of the common schemas for technical documentation (OpenOffice XML, Open Document Format, DocBook, HTML, and TEI) provides a similar set of constructs and rules. They each describe

– Metadata about document
– Title and author
– Headings, indicating a hierarchy of sections
– Paragraphs
– Lists
– Tables
– Figures
– Hyperlinks
– Cross-referencing and a wide variety of inline emphasis constructs. They differ greatly, however, in the combinations they allow, the way in which arbitrary formatting or effects are supported, and the provision of specialised vocabulary for particular domains.

```
<rng:grammar
    datatypeLibrary="http://www.w3.org/2001/XMLSchema-datatypes">
    <rng:define name="people">
        <rng:element name="people">
            <rng:oneOrMore>
                <rng:ref name="person"/>
            </rng:oneOrMore>
        </rng:element>
    </rng:define>
    <rng:define name="person">
        <rng:element name="person">
            <rng:attribute name="birth">
                <rng:data type="gYear"/>
            </rng:attribute>
            <rng:optional>
                <rng:attribute name="death">
                    <rng:data type="gYear"/>
                </rng:attribute>
            </rng:optional>
            <rng:text/>
        </rng:element>
    </rng:define>
    <rng:start>
        <rng:ref name="people"/>
    </rng:start>
</rng:grammar>

<people>
    <person birth="1832" death="1888">Alcott, Louisa May</person>
    <person birth="1775" death="1817">Austen, Jane</person>
    <person birth="1861" death="1891">Balestier, Wolcott</person>
    <person birth="1757" death="1827">Blake, William</person>
    <person birth="1848" death="1895">Boyesen, Hjalmar Hjorth</person>
    <person birth="1820" death="1849">Brontë, Anne</person>
    <person birth="1818" death="1848">Brontë, Emily</person>
    <person birth="1857" death="1924">Conrad, Joseph</person>
    <person birth="1883" death="1923">Hašek, Jaroslav</person>
    <person birth="1865" death="1936">Kipling, Rudyard</person>
    <person birth="1931">Le Carré, John</person>
</people>
```

Figure 5. Minimal schema together with an XML document.

2.3. The HTML family

By far the most widely used (and most abused) markup language is HTML (Altheim and McCarron, 2001). Although influenced by SGML in its first years, it was only in the mid 1990s formalised as an SGML profile, and has seldom been fully implemented as such by processing tools (i.e. web browsers). An explicit XML-compliant version, XHTML, was completed in 2000, but does not have full acceptance 10 years later; that is now being supplanted by a second branch, HTML5 (Hickson, 2010), which extends the language at the same time, but is not expected to have total adoption for some years.

A simple HTML document is listed in Figure 6, showing the same document as LaTeX in the section above. However, HTML is normally used much more as a *container* for a structured document, wrapping it a layer of presentation, navigation and dynamic features. It has two helper technologies which make it hard to treat it on a par with other schemes:

```
<?xml version="1.0" encoding="utf-8"?>
<html xmlns="http://www.w3.org/1999/xhtml" lang="en">
  <head>
    <meta http-equiv="Content-Type" content="text/html; charset=UTF-8"/>
    <title>Simulation of Energy Loss Straggling</title>
    <meta name="author" content="Maria Physicist"/>
    <meta name="DC.Title" content="Simulation of Energy Loss Straggling"/>
  </head>
  <body>
    <div class="header">
      <h1 class="title">Simulation of Energy Loss Straggling</h1>
      <h2 class="author">Maria Physicist</h2>
    </div>
    <div class="div1" id="intro">
      <h2>
        <span class="head">Introduction</span>
      </h2>
      <p>Due to the statistical nature of ionisation energy loss,
large fluctuations can occur in the amount of energy
deposited by a particle traversing an absorber element.
Continuous processes such as multiple scattering and energy
loss play a relevant role in the longitudinal and lateral
development of electromagnetic and <span class="term">hadronic
showers</span>, and in the case of
sampling calorimeters the measured resolution can be
significantly affected by such fluctuations in their active
layers.  </p>
      <p>The work at <a href="http://public.web.cern.ch/public/">CERN</a> is
           likely to be of interest to readers of this paper.</p>
    </div>
    <div class="footer">
      <address> Maria Physicist.
      Date: January 17, 1999
      </address>
    </div>
  </body>
</html>
```

Figure 6. Simple document in HTML.

- CSS (Cascading Stylesheet language) (Çelik et al., 2009): this provides a way to attach visual properties (fonts, spacing, colours, etc.) to HTML elements, keyed by means of the **class** attribute visible on the `<div>` elements in Figure 6;
- JavaScript (ISO, 2009): almost all web browsers implement this scripting language for manipulating the page, pulling in new data, and performing actions.

The use of JavaScript in particular can make HTML a tightly-coupled language like LaTeX, where the page can change itself dynamically inside the browser. Crucially, however, HTML can also be used in other contents, and with other software, as a much more general purpose language. An example of its reuse is as the underlying formatting language of the ePub electronic book specification (IDPF, 2010).

The most powerful technique in HTML is its use of two very generic nestable container elements, `<span>` and `<div>`, which can be used with the **class** attribute to model many different structures. Additional generic attributes added by the Microdata (Hickson, 2011) or RDFa (Pemberton et al., 2008) and W3C

Table 2. Markup languages and their numbers of different elements permitted inside paragraphs.

Number	Markup language
182	TEI (all)
146	DocBook
62	TEI (lite)
35	XHTML

initiatives also make it possible to put an extra layer of semantic markup over HTML.

New elements in HTML5 allow for an even closer relationship with the implementation, in that use of conforming browsers will not just allow standardised video embedding, but also access to environmental information as the location supplied by a computer's GPS. This pushes HTML, and its companion JavaScript, closer to a functional programming language, and away from a document representation language.

2.4. The Text Encoding Initiative

The Text Encoding Initiative (TEI) (Consortium, 2010) scheme was developed during the 1990s, and then again after 2000, now at a 5th release, to create a very rich schema which can reasonably claim to have the most wide-ranging and detailed expressivity of all the markup languages. Table 2 shows the choice of elements available to a writer inside a paragraph, with the TEI offering a staggering 182 possibilities in its fullest version.

TEI was designed as *descriptive* scheme for digitisation of existing material from across the literary and linguistic spectrum, emphasising the use of markup for analysis rather than display.

A simple TEI document is listed in Figure 7. Notice the use of a container <div>, with a child <head> rather than the simple heading approach of HTML's <h1> and <h2>. In contrast, consider the short fragment of speech 'Enough enough, do you want to go buy some Polo?', marked up below with linguistic part of speech information using TEI elements:

```
<u who="#PS05C">
    <s n="1194">
        <w lemma="enough" type="AV0">Enough </w>
        <w lemma="enough" type="AV0">enough</w>
        <c type="PUN">, </c>
        <w lemma="do" type="VDB">do </w>
        <w lemma="you" type="PNP">you </w>
```

```
<TEI>
    <teiHeader>
        <fileDesc>
            <titleStmt>
                <title>Simulation of Energy Loss Straggling</title>
                <author>Maria Physicist</author>
            </titleStmt>
            <editionStmt>
                <edition>
                    <date>January 17, 1999</date>
                </edition>
            </editionStmt>
            <publicationStmt>
                <p>unpublished</p>
            </publicationStmt>
            <sourceDesc>
                <p>part of a sample document used at CERN for teaching LaTeX</p>
            </sourceDesc>
        </fileDesc>
    </teiHeader>
    <text>
        <body>
            <div xml:id="intro">
                <head>Introduction</head>
                <p>Due to the statistical nature of ionisation energy loss,
                    large fluctuations can occur in the amount of energy
                    deposited by a particle traversing an absorber element.
                    Continuous processes such as multiple scattering and energy
                    loss play a relevant role in the longitudinal and lateral
                    development of electromagnetic and <term>hadronic
                        showers</term>, and in the case of sampling calorimeters
                    the measured resolution can be significantly affected by
                    such fluctuations in their active layers. </p>
                <p>An example formula:
                    <formula notation="MathML">
                        <m:math>
                            <m:mi>x</m:mi>
                            <m:mo>/</m:mo>
                            <m:msub>
                                <m:mi>l</m:mi>
                                <m:mrow>
                                    <m:mn>0</m:mn>
                                </m:mrow>
                            </m:msub>
                        </m:math>
                    </formula>
                </p>
                <p>The work at <ref target="http://public.web.cern.ch/public/">CERN</ref>
                    is likely to be of interest to readers of this paper.</p>
            </div>
        </body>
    </text>
</TEI>
```

Figure 7. Simple document in TEI.

```
            <w lemma="want" type="VVI">want </w>
            <w lemma="to" type="TO0">to </w>
            <w lemma="go" type="VVI">go </w>
            <w lemma="buy" type="VVB">buy </w>
            <w lemma="some" type="DT0">some </w>
            <w lemma="polo" type="NN1">Polo </w>
            <c type="PUN">?</c>
        </s>
</u>
```

TEI offers so much choice for two reasons: firstly, it is a modular scheme covering a broad range of descriptive and analytic markup (Table 3 lists the currently available modules), and secondly because of its descriptive nature. Where a prescriptive scheme might dictate that lists are not permitted inside paragraphs (HTML is an example), the TEI must allow both inside and outside paragraphs, because it cannot dictate where such things may be found in existing text.

Amongst TEI's modules there are several which are more akin to database schemas than traditional text. For example, the following fragment is a description of medieval manuscript, with a bewildering variety of metadata in attributes, formal 'fields' (e.g. <rubric>) and use of inline transcriptional notes, for example, on the text of <explicit>.

```
<msItem n="c">
    <locus from="2r9" to="2r19">2r:9-2r:19</locus>
    <title type="uniform">Fra paradiso</title>
    <rubric>Fra paradiso</rubric>
    <incipit>Sva er sagt at paradis er hi<expan>n</
        expan>n oesti
        lutr pessarar veraldar.</incipit>
    <explicit>oc veria bio<supplied reason="illegible">
        r</supplied>g oc
          hitar at me<expan>n</expan>n skili eigi pangat
            komast.</explicit>
</msItem>
```

One of the TEI's key design features is its assumption that users will always choose from a subset of the elements available (the TEI Consortium provides tools to make this easy, for example, Roma (TEI, 2010)), and thus never face the choice of 182 elements. As the table earlier showed, a typical TEI profile ('Lite') only offers 62 choices inside a paragraph – though this can still be off-puttingly large.

From the point of view of technical documentation interchange, the TEI is less than ideal because it uses very general mechanisms typically qualified by unconstrained **type** attributes. For example, where HTML models a numbered list as

```
<ul>
    <li>first item</li>
    <li>second item</li>
</ul>
The TEI will use
<list type ="ordered"><item>first item</item><item>second item</item></list>
```

Table 3. Modules in the TEI scheme.

Module in TEI	**Coverage**
analysis	Simple analytic mechanisms
certainty	Certainty and responsibility
core	Elements available in all TEI documents
corpus	Language corpora
dictionaries	Dictionaries
drama	Performance texts
figures	Tables, formulae, and graphics
gaiji	Representation of non-standard characters and glyphs
header	The TEI Header
isofs	Feature structures
linking	Linking, segmentation, and alignment
msdescription	Manuscript description
namesdates	Names, dates, people, and places
nets	Graphs, networks, and trees
spoken	Transcriptions of speech
tagdocs	Documentation elements
tei	The TEI infrastructure
textcrit	Critical apparatus
textstructure	Default text structure
transcr	Representation of primary sources
verse	Verse

The freedom allowed here for other values than 'ordered' for **@type** can be confusing to the consumer of a TEI text.

In practice, users of the TEI group into *communities of practice*, adopting more strictly controlled subsets of the specification for a particular domain or workflow. For example, scholars using TEI to encode detailed descriptions of medieval manuscripts may use the subset defined by ENRICH (http://enrich.manuscriptorium.com/), while large-scale, but light touch, library digitization projects may use the *TEI in Libraries: Guidelines for Best Practices* (http://wiki.tei-c.org/index.php/TEI_in_Libraries:_Guidelines_for_Best_Practices_Working_Group). Interoperability between a manuscript description and a library digital edition of *A Christmas Carol* is rarely a high priority, so using very distinct subsets of TEI is not problematic. The two documents do, however, share a common ontology (regarding the TEI simply as a vocabulary), so future use of the two streams of data is considerably easier than if they had used entirely different schemata.

2.5. DocBook

DocBook (Walsh, 2010), developed since the 1990s, and now in its 5th release, is the best-known and best-supported of the technical writing schemas, with support from many tool vendors. In contrast to the TEI, it is highly prescriptive. It is squarely aimed at new writing, and supporting features for documenting IT systems (it descends from the requirements of the technical publisher O'Reilly. The simple example is shown in Figure 8, making no use of DocBook's specialised elements. For example (list taken from the documentation) DocBook offers five distinct ways of including examples of computer programs:

`<literallayout>` **A** `<literallayout>` does not have any semantic association beyond the preservation of whitespace and line breaks. In particular, while program listing and screen are frequently presented in a fixed-width font, a change of fonts is not ordinarily implied by `<literallayout>`.

`<programlisting>` **and** `<programlistingco>` The `<programlisting>` and `<programlistingco>` elements are verbatim environments, usually presented in Courier or some other fixed-width font, for program sources, code fragments, and similar listings.

`<screen>` **and** `<screenco>` The `<screen>` and `<screenco>` elements are verbatim or literal environments for text screen captures, other fragments of an ASCII display, and similar things.

screenshot `<screenshot>` is actually a wrapper for a `<mediaobject>` intended for screenshots of a GUI, for example.

synopsis A synopsis is a verbatim environment for command and function synopses.

In even more detail, there are 14 elements for describing aspects of a user interface (`<accel>`, `<guibutton>`, `<guiicon>`, `<guilabel>`, `<guimenu>`, `<guimenuitem>`, `<guisubmenu>`, `<keycap>`, `<keycode>`, `<keycombo>`, `<keysym>` `<menuchoice>`, `<mousebutton>`, and `<shortcut>`) and 17 more for specifying a computer language. Like the TEI, however, DocBook offers general purpose elements and attributes which can allow authors to ignore the semantic elements

DocBook gains considerably from solid support in tools for rendering into different formats (HTML, PDF, etc.) and from its well-documented and detailed syntax for computer documentation. Like the TEI, it has a rich and powerful section of elements dedicated to metadata about the document.

```
<?xml version="1.0" encoding="utf-8"?>
<article xmlns="http://docbook.org/ns/docbook" xmlns:xlink="http://www.w3.org/1999/xlink"
  version="5.0">
  <info>
    <title>Simulation of Energy Loss Straggling</title>
    <author>
      <personname>Maria Physicist</personname>
      <address>
        <city>Geneva</city>
        <postcode>CH-1211 </postcode>
        <country>Switzerland</country>
      </address>
      <email>maria.physicist@cern.ch</email>
    </author>
    <pubdate>January 17th, 1999</pubdate>
  </info>
  <sect1>
    <title>Introduction</title>
    <para>Due to the statistical nature of ionisation energy loss, large fluctuations can occur in
      the amount of energy deposited by a particle traversing an absorber element. Continuous
      processes such as multiple scattering and energy loss play a relevant role in the longitudinal
      and lateral development of electromagnetic and <firstterm>hadronic showers</firstterm>, and in
      the case of sampling calorimeters the measured resolution can be significantly affected by
      such fluctuations in their active layers.</para>
    <para>An example formula: <equation><math xmlns="http://www.w3.org/1998/Math/MathML"
          ><mi>ξ</mi><mo>/</mo><msub><mi>I</mi><mrow><mn>0</mn></mrow></msub></math></equation>
    </para>
    <para>The work at <link xlink:href="http://public.web.cern.ch/public/">CERN</link> is likely to
      be of interest to readers of this paper.</para>
  </sect1>
</article>
```

Figure 8. Simple document in DocBook.

2.6. The faithful servants

The schemas we have described above all cover the broad aspects of describing the semantics of a document, specifying metadata, and (in different areas) delve into domain-specific markup. However, there remain four specialist areas in which (wisely) none of the XML-based schemes venture. These are catered for by XML standards of their own, largely from the World Wide Web Consortium, which can be included within any document by virtue of *name-spaces*. These specialist schemes are for:

Mathematics The MathML (Poppelier et al., 2003) language provides comprehensive facilities for expressing maths in XML markup, both in a conventional presentation way, and in a less familiar but more powerful semantic fashion.

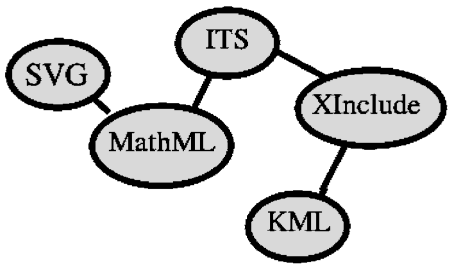
Vector graphics Scalable Vector Graphics (SVG) (Dahlström et al., 2010) defines a language for pictures, such as graphs and plots, primarily targetting the web, where it also provides for user interaction.

Internationalization The Internationalization Tag Set (ITS) (Sasaki and Lieske, 2007) provides facilities for marking up multilingual documents, for example, specifying writing direction, and for translation tools, indicating where translation should occur (see also Sasaki and Lommel 2012 in this volume).

Transclusion The small but necessary XInclude (Marsh et al., 2006) standard provides markup for one document pulling in bits of another.

Describing geographical maps The KML (Google, 2010) and GML (ISO, 2007) languages offer descriptions of mapping at a higher level than SVG; the former is in very widespread use through Google Earth and Google Maps.

We have already seen an example of MathML in our sample documents earlier, and those conveniently illustrate how two document schemas can work together. Using SVG is rather similar. Consider this picture:



It look like this in SVG XML:

```
<ellipse
    id="svg_10"
    transform="rotate(-0.516164243221283305.51266479492284,100.
                                                40528869628734)"
    filter="url(#svg_7_blur)"
    ry="31.56728"
    rx="60.44972"
    cy="100.40529"
    cx="305.51266"
    stroke-width="5"
    stroke="#000000"
    fill="#00ff00"/>
```

The existence of the servant schemas is important, because they demonstrate a possibility of document interchange and re-use beyond the relatively simple word-based systems. However, we should consider that the MathML and SVG languages are well over a decade old, but are not yet reliably supported by all the components of the modern web, and be a little cautious.

2.7. The word-processor family

The word-processor formats of the common office suites (Microsoft Office 2007 onwards, and Open Office (OpenOffice, 2010)) store their documents in XML markup (albeit as a ZIP archive containing a number of files). They each have a model which combines a very light structural model (fundamentally little more than a set of blocks on the page, with inline sequences inside the blocks) with a very rich set of visual properties, addressing combinations with named *styles*. This is potentially very powerful, when used in a disciplined way, but the set of styles can be very confusing, ranging from the typical structural functions (Figure 9:1), through purely presentational (Figure 9:2), to locally-defined domain specific (Figure 9:3). The software also allows the user to override any of the formatting *ad hoc*, which can create havoc for interchange.

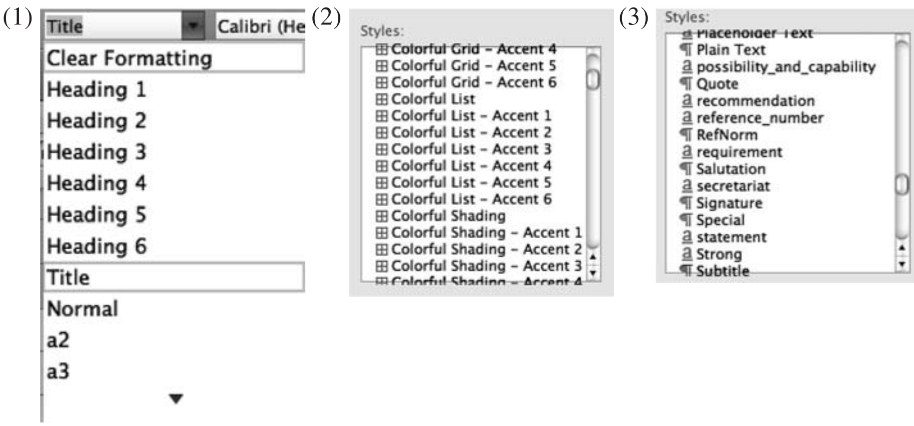


Figure 9. Style listing in Microsoft Word.

The underlying XML is relatively clean, since all of the styling information is abstracted away from the normal stream. The representation of a simple paragraph earlier in this chapter looks like this in Open XML (ISO, 2008):

```
<w:p
    xmlns:w="http://schemas.openxmlformats.org/wordprocessingml/
    2006/main">
    <w:pStyle w:val="TextPara"/>
    <w:r>
        <w:rPr/>
        <w:t xml:space="preserve">An XML document represents a (kind of)
        tree. It has a single root and many nodes Each node can be a subtree,
        a single
</w:t>
</w:r>
<w:r>
<w:rPr>
<w:b/>
</w:rPr>
<w:t xml:space="preserve">element</w:t>
</w:r>
<w:r>
<w:rPr/>
<w:t xml:space="preserve"> (possibly bearing some
</w:t>
</w:r>
<w:r>
<w:rPr>
<w:b/>
</w:rPr>
<w:t xml:space="preserve">attributes</w:t>
</w:r>
<w:r>
<w:rPr/>
<w:t xml:space="preserve">) or a string of
</w:t>
</w:r>
<w:r>
<w:rPr>
<w:b/>
</w:rPr>
<w:t xml:space="preserve">character data.</w:t>
</w:r>
</w:p>
```

and like this in Open Document Format (ISO, 2006) (ODF):

```
<text:p text:style-name="Text body"
    xmlns:text="urn:oasis:names:tc:opendocument:xmlns:
        text:1.0">An XML
    document represents a (kind of)
    tree. It has a single root and many nodes Each node
        can be a subtree,
    a single <text:span text:style-name="Emphasis">
        element</text:span>
    (possibly bearing some <text:span text:style-name="
        Emphasis">attributes</text:span>)
    or a string of
    <text:span text:style-name="Emphasis">character
        data.</text:span>
</text:p>
```

The difference between the formats is that OpenXML keeps *all* text within a spanning container, whereas Open Document Format allows ‘mixed content’ inside `<text:p>`, a mixture of character data and `<text:span>` elements.

OpenXML allows for both a style reference (`<pStyle val="Text Para"/>`) and a local override (`<rPr><b/></rPr>` says that the text should be in bold), where ODF has simply style names.

OpenXML makes especially heavy use of namespaces, defining 12 separate vocabularies for different parts of the document, which it represents using at least 20 separate files. The apparent simplicity of the markup example above masks considerably detailed specification of visual formatting, the details of tables, and hyperlinking.

The word-processor schemes are tightly coupled to presentation and the software. Separating out all the visual formatting to named style components, they operate within a single level of nesting, and only the style name provides any attempt at semantic information. There is no possibility of rules to say where styles can occur or nest, although such control can be programmed within the software itself.

2.8. Which markup scheme to choose?

It is unlikely that any of the document representation schemes described above is a universal panacea suited to everyone. It is clear that their characteristics are dictated by very different requirements, and a choice will depend on an assessment of tool support, validation needs, domain coverage, output needs, and peer-group involvement. The use of specialised vocabularies for math, graphics, etc. seems unlikely to be controversial, as the only alternative is to provide non-scaling pictures.

Whichever scheme is adopted to manage the structured documents we defined at the start of this chapter, it is inevitable that there will be *abuse* – the LaTeX users who makes section headings by setting a large bold sans-serif font, the HTML author who uses tables to lay out pages, the TEI encoder who uses <l> (designed for lines of verse) to force line breaks, the DocBook writer who uses the **style** attribute to convey all distinctions, and the Word practitioner who puts in empty paragraphs to make more vertical space.

3. Conclusion

We started by describing the characteristics of a 'document', and we have spent most of this chapter looking at a variety of document markup schemes and how they manage the relationship between that abstract document and its visual representation. There are three main reasons for making such a distinction explicit. The first is the desire to be able to *change* the visual representation for different situations or media, including non-visual representation to provide access for a wider range of readers. The second is the ability to *check* the structure of documents, and ensure that they contain what we think they do. The third is the need to *analyse* documents, and either extract data from them, or enhance them with, for example, links to other documents and resources. The first two of these claims, eloquently argued in a classic paper of 1987 by Coombs, Renear and de Rose (Coombes et al., 1987) (and in many other publications since), may be open to the criticism that the writing and publishing industry has *not*, generally speaking, adopted the idea of separating markup from content. Phrases like 'The most dramatic sign of change is the overwhelming promotion of SGML' from 1990 (DeRose et al. (1990), p.16) raise a smile 20 years later. Most writing remains tied to a particular medium, and uses *ad hoc* visual formatting. Much use is made of design features which make documents attractive, but have no semantic corollary.

It is more convincing, however, to argue that the separate representation of content, structure and layout in a document is of far more importance than just layout, because it provides the possibility of identifying, extracting and sharing *data* from documents, which brings us back to the start of the chapter. A document is more than simply a stream of data, but that does not mean it does not contain data within it. Whether that be personal names, dates, formulae, geographical locations, addresses, financial data or references, the structured document is the building block of the modern information flow.

Notes

1. See "The Elements of Euclid: viz. the first six books" on http://books.google.com/
2. See http://commons.wikimedia.org/wiki/File:Greek_manuscript_vetustissimus_Thucydides.png
3. The current TeX software, under the close control of its author Donald Knuth, was released in 1982 (Knuth, 1984) and has changed very little since that time. Despite some considerable additions of capability in the last decade, most notably pdfTeX for making PDF output, and XeTeX for processing Unicode, the adherence to the input language and backward compatibility is almost absolute. TeX is one of the oldest and cleanest examples of free software.
4. Conversions from TeX to other formats are distinctly non-trivial, because of the nature of its input language, and TeX itself supports no import from other formats, though it can be programmed to read XML. Conversions to TeX from XML formats are commonplace and easy to write.
5. It is possible, using a structured editor, to write entirely 'vanilla' LaTeX which is amenable to structured processing, but it is not easy to enforce this.

References

Altheim, Murray and Shane McCarron
2001 XHTML 1.1 – Module-based XHTML. Recommendation, W3C. URL http://www.w3.org/TR/2001/REC-xhtml11-20010531.2010-10-25.

Bray, Tim, Jean Paoli, Eve Maler, François Yergeau and C. M. Sperberg-McQueen
2008 Extensible Markup Language (XML) 1.0 (Fifth series). Recommendation, W3C. URL http://www.w3.org/TR/2008/REC-xml-20081126/. 2010-10-25.

Çelik, Tantek, Bert Bos, Ian Hickson and Håkon Wium Lie
2009 Cascading Style Sheets Level 2 Revision 1 (CSS 2.1) Specification. Candidate recommendation, W3C. URL http://www.w3.org/TR/2009/CR-CSS2-20090908.2010-10-25.

Consortium TEI
2010 TEI P5: Guidelines for Electronic Text Encoding and Interchange. Technical report, TEI Consortium. URL http://www.tei-c.org/Guidelines/P5/. 2010-10-25.

Coombes, James H. , Allen Renear and Steven J. DeRose
1987 Markup Systems and The Future of Scholarly Text Processing. *Communications of the Association for Computing Machinery*, 30.

Dahlström, Erik, Jon Ferraiolo, Fujisawa Jun, Anthony Grasso, Dean Jackson, Chris Lilley, Cameron McCormack, Doug Schepers, Jonathan Watt and Patrick Dengler
2010 Scalable Vector Graphics (SVG) 1.1 (Second Edition). Working draft, W3C. URL http://www.w3.org/TR/2010/WD-SVG11-20100622/.2010-10-25.

DeRose, Steven J. , David Durand, Elli Mylonas and Allen Renear
1990 What is Text, Really? *Journal of Computing in Higher Education*, 1:3-26.

DeRose, Steven J. , David G. Durand, Elli Mylonas and Allen H. Renear
1997 Further Context for “What is Text, Really?”. *The Journal of Computer Documentation*, 21:40-44.
Goldfarb, Charles F.
1990 *The SGML Handbook*. Oxford, Clarendon Press.
Google
2010 KML (Keyhole Markup Language). Technical report, Google. URL http://code.google.com/apis/kml/documentation/. 2010-10-25.
Hickson, Ian
2010 HTML5: A vocabulary and associated APIs for HTML and XHTML. Working draft, W3C. URL http://www.w3.org/TR/2010/WD-html5-20101019/.
Hickson, Ian
2011 HTML Microdata. Recommendation, W3C. URL http://www.w3.org/TR/microdata/. 2011-09-14.
Hockey, Susan, Allen Renear and Jerome J. McGann
1999 Panel: What is Text? A Debate on the Philosophical and Epistemological Nature of Text in the Light of Humanities Computing Research. Technical report, Association of Literary and Linguistic Computing. URL http://www.iath.virginia.edu/ach-allc.99/proceedings/hockey-renear2.html.
IDPF
2010 ePub specification. Technical report, IDPF. URL http://www.idpf.org/specs.htm. 2010-10-25.
ISO
1986 *Standard Generalized Markup Language (ISO 8879:1986 SGML)*. International Organization for Standardization, Geneva.
ISO
1994 *Information processing – Text and office systems – Document Style Semantics and Specification Language (DSSSL), ISO/IEC DIS 10179.2*. International Organization for Standardization, Geneva.
ISO
2003 *ISO/IEC 19757-2:Amd1 Document Schema Definition Language (DSDL) – Part 2: Regular-grammar-based validation – RELAX NG – Amendment 1: Compact Syntax. ISO version of the RELAX NG Compact Syntax*. International Organization for Standardization, Geneva.
ISO
2006 *ISO/IEC 26300:2006 Information technology – Open Document Format for Office Applications (OpenDocument) v1.0.* International Organization for Standardization, Geneva.
ISO
2007 *ISO 19136:2007 Geographic information – Geography Markup Language (GML)*. International Organization for Standardization, Geneva.
ISO
2008 *ISO/IEC 29500-1:2008 Information technology – Document description and processing languages – Office Open XML File Format – Part 1: Fundamentals and Markup Language Reference*. International Organization for Standardization, Geneva.

ISO
2009 *ISO/IEC 16262:2002 Information technology – ECMAScript language specification*. International Organization for Standardization, Geneva.

Knuth, Donald E.
1984 *The TeXbook*. Addison-Wesley, Reading, Massachusetts.

Kubina, Petra and Andy Lücking
2012 Barrier-free communication. In Alexander Mehler, Laurent Romary, and Dafydd Gibbon, editors, *Handbook of Technical Communication*. De Gruyter, Berlin/Boston.

Lamport, Leslie
1986 *LaTeX: A Document Preparation System*. Addison-Wesley, Reading, Massachusetts.

Marsh, Jonathan, David Orchard and Daniel Veillard
2006 XML Inclusions (XInclude) Version 1.0 (Second Edition). Recommendation, W3C. URL http://www.w3.org/TR/2006/REC-xinclude-20061115/. 2010-10-25.

OpenOffice
2010 OpenOffice.org. Technical report, Oracle. URL http://www.openoffice.org/. 2010-10-25.

Pemberton, Steven, Ben Adida, Shane McCarron and Mark Birbeck
2008 RDFa in XHTML: Syntax and Processing. A collection of attributes and processing rules for extending XHTML to support RDF. Recommendation, W3C. URL http://www.w3.org/TR/2008/REC-rdfa-syntax-20081014. 2010-10-25.

Poppelier, Nico, Robert Miner, Patrick Ion and David Carlisle
2003 Mathematical Markup Language (MathML) Version 2.0 (Second Edition). Recommendation, W3C. URL http://www.w3.org/TR/2003/REC-MathML2-20031021/. 2010-10-25.

Rahtz, Sebastian and Trevor Rowley
1984 *Middleton Stoney. Excavation and Survey in a North Oxfordshire Parish 1970–1982*. Oxford University Department for External Studies.

Renear, Allen
1997 Out of Praxis: Three (Meta)Theories of Textuality. In Kathryn Sutherland, editor, *Electronic Text. Investigations in Method and Theory*, pages 107–126. Clarendon Press, Oxford.

Renear, Allen, Elli Mylonas and David Durand
1996 Refining our Notion of What Text Really Is: The Problem of Overlapping Hierarchies. In Nancy Ide and Susan Hockey, editors, *Research in Humanities Computing*, pages 263–280. Oxford University Press, Oxford.

Romary, Laurent
2012 Scholary communication. In Alexander Mehler, Laurent Romary, and Dafydd Gibbon, editors, *Handbook of Technical Communication*. De Gruyter, Berlin/Boston.

Sasaki, Felix and Christian Lieske
2007 Internationalization Tag Set (ITS) Version 1.0. Recommendation, W3C. URL http://www.w3.org/TR/2007/REC-its-20070403/. 2010-10-25.

Sasaki, Felix and Arle Lommel
2012 Multilingual computing. In Alexander Mehler, Laurent Romary, and Dafydd Gibbon, editors, *Handbook of Technical Communication*. De Gruyter, Berlin/Boston.

Stührenberg, Maik
2012 Foundations of markup languages. In Alexander Mehler, Laurent Romary, and Dafydd Gibbon, editors, *Handbook of Technical Communication*. De Gruyter, Berlin/Boston.

TEI
2010 Roma: generating validators for the TEI. Technical report, TEI Consortium. URL http://www.tei-c.org/Roma. 2010-10-25.

Thompson, Henry S. , Murray Maloney, David Beech and Noah Mendelsohn
2004 XML Schema Part 1: Structures Second series. Recommendation, W3C. URL http://www.w3.org/TR/2004/REC-xmlschema-1-20041028/. 2010-10-25.

Walsh, Norman
2010 *DocBook 5: The Definitive Guide*. O'Reilly Media and XML Press.

3. Foundations of markup languages

Maik Stührenberg

The term *markup* depicts information (usually in textual form) that is not properly part of the original content (the primary data) but that is added to this content "to convey information about it" (Goldfarb 1991, p. 125). A *markup language* provides a notation (a syntax) to differentiate markup and the content to which the markup is added, and the respective markup units that can be used to annotate it. Historically the term derives from the classic text-publishing where an author (or a typographer) made notes (annotations) alongside the (hand-written) text to provide information about how the text should be typeset (which was done by others). This kind of annotation is called *procedural markup*, since it usually consists of concrete low level formatting commands such as "page break" or "change font size" specific to a certain text formatting language, such as SCRIPT for the IBM CP67/CMS (cf. Madnick and Moulton 1968), the mid-70's variant WATERLOO SCRIPT (cf. Department of Computing Services 1980) or TEX (cf. Knuth 1984). In contrast to that is *descriptive markup*, that is, the labeling of units of content by adding logical (semantic) information to it, thus separating content and formatting.

When we use the term *markup language* nowadays, we usually talk about a system that provides a machine-processable (and possibly human-readable) *syntax* that is used not only to annotate primary data but also to structure it. Since data and annotation are interspersed with each other this can be called as self-describing or *semistructured data* (Abiteboul, Buneman, and Suciu 2000, p. 11). Examples for such markup languages are LATEX (cf. Mittelbach et al. 2006) or HTML (Raggett 1997).

By this definition we have already specified two components of a markup language already: the notation (or linearisation) and the formal model (the data structure). The third aspect identified by Sperberg-McQueen 2002 is a mechanism, Sperberg-McQueen (2002) a mechanism for validation (a *constraint language* or grammar), coining the metaphor of a tripod (cf. Huitfeldt and Sperberg-McQueen 2004). Following Sperberg-McQueen, Huitfeldt, and Renear (2000) as well as Witt (2004) we will add a fourth leg to this metaphor by differentiating between the serialized markup and its interpretation, resulting in the metaphor of a chair. We choose this metaphor since there is a difference between the first three components and the last one, since the former three are crucial to markup languages while the latter one is optional – there are three-legged chairs as well as four-legged ones.

Nowadays, markup languages (such the above-mentioned examples) are applications of a *meta language*, that is, a language for defining markup languages. The two most established meta languages are the Standard Generalized Markup Language (SGML, ISO 8879:1986) and the Extensible Markup Language (XML, Bray, Paoli, and Sperberg-McQueen 1998). This means in return that at least one component, the syntax, is inherited from the respective meta language.[1] We will talk about the syntax of XML-based markup languages in the next section.

The data model of a markup language is dependent on its meta language on the one hand but is influenced by the constraint language that is used to define its grammar on the other hand. We will elaborate this topic in Section 2. A document grammar can be seen as a set of rules for instances of this very markup language, validating an instance by means of this grammar results in accepting or rejecting it. A document grammar in itself is constructed by the application of a constraint language, while the constraint languages available are dependent on the meta language. In Section 3, we will discuss different constraint languages that can be used to define document grammars for XML-based markup languages.

The fourth (and optional) leg of our markup chair, the differentiation between the serialized markup – Witt (2004) introduces the term "layer" – and its interpretation (the "level") will be discussed in Section 4.

1. The syntax of XML-based markup languages

As we have already mentioned in the introductory part, XML-based markup languages inherit their notation format (the syntax) from their meta language. XML celebrated its 10th birthday in 2008. During this decade it has become a widely adopted standard for the structured storage of data of many kinds (especially for semistructured data). But the roots of the specification (and therefore, elementary parts of its syntax) go back to the 1960s, when William Tunnicliffe of the Graphic Communications Association (GCA) presented the idea of separating content and formatting at a meeting of the Canadian Government Printing Office in September 1967. This resulted in the "GenCode concept" (Goldfarb 1991). In 1969 the Text Description Language (TDL) was developed at IBM by Charles F. Goldfarb, Ed Mosher, and Ray Lorie, followed by the Integrated Textual Information Management Experiment (InTIME) prototype (Goldfarb, Mosher, and Peterson 1970). In 1971 TDL was renamed as Generalized Markup Language (GML, cf. Goldfarb 1973), introducing a simple syntax, including angle brackets "<" and ">" as delimiter strings between text and markup, and the concept of start and end tags. In 1978 both GenCode and GML combined forces in the newly established ANSI Committee on Computer Languages for the Processing of Text which became the ISO committee that developed the Standard Generalized Markup Language (SGML) in the 1980's and which published the final version of the standard in 1986 (ISO 8879:1986), establishing the concept of generic markup: "The characteristics of being declarative, generic, nonproprietary, and contextual make the Standard Generalized Markup Language 'standard' and 'generalized'." (Maler and El Andaloussi 1995, p. 12).

Especially the possibility to construct new markup languages by defining new Document Type Definitions (DTD) without breaking compatibility to other applications of the SGML meta language was one of the main reasons for SGML's success. Apart from being used by IBM and the US Department of Defense, the Text Encoding Initiative (TEI) opened SGML for humanities and linguistics departments (Aguera et al. 1987; Sperberg-McQueen and Burnard 1993 and Ide 1994) and the development of the Hypertext Markup Language (Raggett 1997) as SGML application (Berners-Lee and Fischetti 1999, p. 41) established the meta language as common ground for various domains. However, HTML was suitable for structuring hypertexts, yet it lacked (and still lacks) markup for more generic texts, while SGML was considered too complex to be delivered over the web (although there were attempts made such as the ones discussed in Rubinsky and Maloney 1997). A possible solution was the development of a stripped-down version of SGML addressing the ten issues that were summarized in Bray (1996a) and in Sperberg-McQueen and Bray (1997). Amongst others, the new specification should be straightforwardly usable over the internet, should support a wide variety of applications, and should be compatible with SGML.[2] It is therefore SGML's reference concrete syntax that was mostly transferred unchanged to XML (to be more specific: to XML instances).

An XML document (that is, an XML instance) is (as well as an SGML document) physically composed of "storage units called *entities* which contain either parsed or unparsed data. Parsed data is made up of characters, some of which form character data and some of which form markup." (Graham and Quin 1999, p. 111). The *document entity* serves as the root of the entity tree (cf. Section 2) and as a starting point for software processing the document (an *XML processor*). Logically it consists of "declarations, elements, comments, character references, and processing instructions, all of which are indicated in the document by explicit markup" (Graham and Quin 1999, p. 125). Of special interest are the elements, since "[e]ach XML document contains one or more elements, the boundaries of which are either delimited by start-tags and end-tags, or, for empty elements, by an empty-element tag. Each element has a type, identified by name, sometimes called its 'generic identifier' (GI)" (Graham and Quin 1999, p. 161f). A start tag begins with the delimiter string opening angle bracket "<", followed by the GI and the delimiter string closing angle bracket ">". The end tag is constructed similar with the only difference that after the opening angle bracket the forward slash has to be written. The data between start and end tag is called the *content* of the element and can be either text, other elements or a mixture of both. Listing 1 shows an example element `a` containing the textual content “Text”.

Listing 1. An example element.

```
<a>Text</a>
```

Empty elements (that is, an element with no content) can be represented either via a start tag immediately followed by an end tag, or via an empty element tag – both variants can be seen in Listing 2.

Listing 2. The two allowed representations of an empty element.

```
<b></b>
<b/>
```

These examples show some small differences between XML and its ancestor SGML: Apart from being case-sensitive, most of the latter’s minimization features have been stripped in XML. While SGML allows for example that either the start tag or the end tag (or both) could be declared optional in the document grammar, there is only a single possible minimization left in XML which corresponds to attributes declared with default values in the document grammar (cf. Section 3). Attribute specifications can occur in the start tag (and only in the start tag) of an element and consist of the attribute name and its value (enclosed in single or double quotes, see Graham and Quin 1999, p. 166f). Each attribute name may only appear once in a start tag (although it is allowed to use the same attribute multiple times in different tags, because attributes are declared locally). The value of an attribute may not contain the opening angle bracket character. Listing 3 shows an example element `c` containing an attribute `d`.

Listing 3. An example attribute.

```
<c d=“Some Text”>More text</c>
```

Elements containing other elements (or so called mixed content, that is the before-mentioned mixture of elements and text) can be used to structure the content (cf. Listing 4).

Listing 4. Establishing structure.

```
<e>
 <f>Some text</f>
 <f>More text</f>
 <g>
  <i>Even more text</i>
 </g>
</e>
```

In Listing 3 it can be observed that it is not only possible to use the same element type (in this example the `f` element) more than once but also to nest elements into each other – as long as they nest properly. Overlapping start and end tags are not allowed, therefore the example shown in Listing 5 is not correct (not well-formed, see below).

Listing 5. Incorrect nesting of elements.

```
<j>
 <k>Some text
  <l>More text
 </k>
  </l>
</j>
```

Attributes can neither contain other attributes nor elements, however, they can be of different types: “a string type, a set of tokenized types, and enumerated types” (Graham and Quin 1999, p. 187). All of these types have to be declared in the *attribute list declaration* in the document grammar, but of special interest are two tokenized types: ID and IDREF. The value of an attribute of type ID “must uniquely identify the elements which bear them” (Graham and Quin 1999, p. 187). Values of type IDREF in contrast “must match the value of some ID attribute” (Graham and Quin 1999, p. 190). That means that a pair of elements having an ID (respective IDREF) type attribute can be connected via their attributes’ values (e. g. to link a footnote text to the corresponding footnote marker in the running text). There is an additional IDREFS tokenized type that allows to refer to one or more values of attributes of the ID type, thus establishing not only 1:1 (one-to-one) but 1:*n* (one-to-many) links.

In contrast to SGML, XML documents do not need to be *valid* according to a document grammar (although it is needed if you want to use the before-mentioned ID/IDREF linking mechanism), but may be used stand-alone, in which case they are called *well-formed* – given that they satisfy the well-formedness constraints stated in the XML specification.

The statements made so far are only a fraction of one part of the XML specification (e. g., we have not talked about the XML prolog or processing instructions), the other part deals with one grammar formalism: the XML Document Type Definition (DTD). Before we discuss XML DTDs (and alternative constraint languages) in Section 3 we will have a look at the formal model of XML instances.

2. The formal model of XML instances

Both, SGML and XML, were developed with the task of structuring textual documents in mind. As stated by Coombs, Renear, and DeRose (1987, p. 945), "[d]ocuments have a natural hierarchical structure: chapters have sections, sections have subsections, and so on, until one reaches sentences, words, and letters." This statement grounded the theory that text can be considered as an ordered hierarchy of content objects (OHCO), postulated by DeRose et al. (1990) and refined in Renear, Mylonas, and Durand (1996). Such an ordered hierarchy can be represented by a tree and therefore the formal model of an XML instance is considered to be a tree: As we have already seen in Section 1, elements nest properly, the XML specification speaks of the "entity tree" (Graham and Quin 1999, p. 251), and the XML Information Set (Cowan and Tobin 2004) uses the tree as a possible data model (although it is clearly stated that this is only an option). Other authors agree on the tree model (cf. for example Ng 2002 or Liu, Vincent, and Liu 2003) and following Klarlund, Schwentick, and Suciu (2003, p. 9) we propose the following formal definition of an *XML Tree*:[3]:

Definition 1 *An XML tree* $t = (N,E,<,\lambda,\nu)$ *consists of a directed tree* (N,E)*, where* N *is the set of nodes,* $E \subseteq N \times N$ *is the set of edges,* $<$ *is a [strict] total order on* N*,* $\lambda : N \rightarrow \Sigma$ *is a partial labeling function, and* $\nu : N \rightarrow D$ *is a partial value function; moreover, the order must be depth-first: whenever* x *is an ancestor of* y*, it holds that* $x < y$*.*

A graphical representation of the example XML instance presented in Listing 4 on p. 86 is shown in Figure 1.

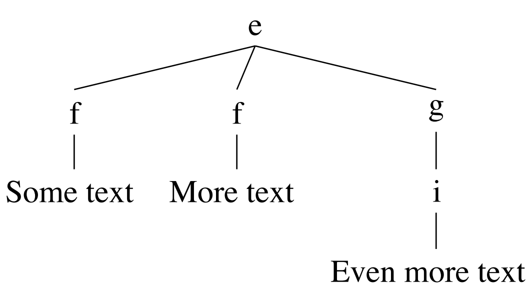


Figure 1. Tree representation of the example XML instance.

The elements are represented as nonterminal nodes in the tree with the generic identifier as label. Leafs contain text nodes.

Before we discuss the appropriateness of the tree as a data model for XML instances any further we will sketch out some limitations of it.

2.1. Limitations of the tree data model

Comparing to the general graph model a (directed) tree has some limitations: There is only a single root node, crossing arcs and cyclic paths are not allowed. In the refinement of the before-mentioned OHCO theory (Renear, Mylonas, and Durand 1996), some counterexamples against the then-called OHCO-1 theory were given, proving that text is not an ordered hierarchy (and therefore not an ordered, rooted tree). As a result, OHCO-2, "An analytical perspective on a text determines an ordered hierarchy of content objects" and OHCO-3 were established. The latter refinement first defines the term "sub-perspective" (as a proper inclusion between two perspectives *x* and *y*) to conclude "objects may overlap in a perspective, but if they do then they belong to different sub-perspectives of that perspective" (Renear, Mylonas, and Durand 1996, p. 274).

One of the counterexamples that were presented against OHCO-1 is the existence of multiple logical hierarchies. We will call a logical hierarchy a "level" (cf. Section 4) and multiple hierarchies "concurrent hierarchies". Concurrent hierarchies are defined in Dekhtyar and Iacob (2005, p. 186) as follows:

> A hierarchy is formed by a subset of the elements of the markup language used to encode the document. The elements within a hierarchy have a clear nested structure. When more than such a hierarchy is present in the markup language, the hierarchies are called concurrent.

While concurrent hierarchies can result in overlapping markup this is not compulsory – there are multi-dimensional annotations that can be easily combined in a single XML instance. On the other hand overlapping annotations can be the result of a single annotation layer (cf. Section 4 for a further discussion on this topic): Pianta and Bentivogli (2004, p. 31) use the annotation of a multiword expression (a discontinuous segment) to demonstrate overlapping markup.

(1) Coi superalcolici bisogna *andarci* veramente *piano*. Translation: People should *take it* really *easy* with liquors.

The Italian multiword expression "andarci piano" (English: "take it easy") corresponds to one lexical unit (with a single meaning). The authors try to annotate this expression on the different levels of *token* (the orthographical form), *potential word* ("an inflected word form before phonological and/or orthographic adjustment is applied to adjacent word forms"), and *lexical unit* ("one or more potential words carrying a unitary lexical meaning"), but since the token "andarci" contains two potential words ("andare" and "ci") the resulting graph would lead to multiple parentage (assuming, that the token level is considered as the basic representation level, cf. Figure 1 in Pianta and Bentivogli 2004, p. 31). As Sperberg-McQueen and Huitfeldt (2004, p. 151) stated, multiple parentage is nothing more or less than overlap.

A very simple example of overlapping hierarchies can be demonstrated by annotating the sentence "The sun shines brighter" on both the syllable and morpheme levels (cf. Listing 6).

Listing 6. Overlap example.

```
<syllables>
<morphemes>
  <s><m>The</m></s>
  <s><m>sun</m></s>
  <s><m>shine</m><m>s</m></s>
  <s><m>brigh</s>
  <s>t</m><m>er</m></s>.
</morphemes>
</syllables>
```

The overlap occurs in the word "brighter" as one can easily observe in Figure 2.

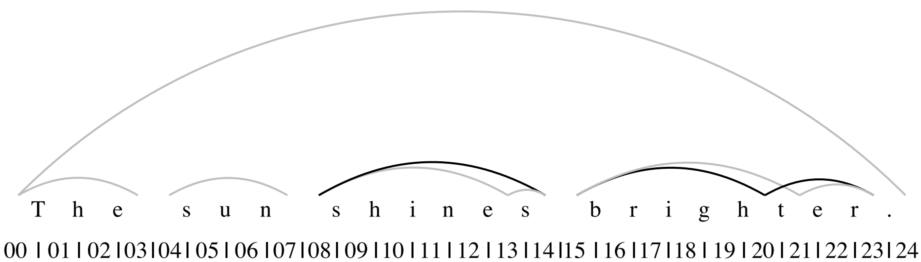


Figure 2. Graphical representation of the overlapping hierarchies.

In this case the overlap is a result of multiple annotation levels. Following Durand (1999); Durusau and O'Donnell (2002) and Witt (2004), different relations can be subsumed between two annotation levels, namely start tag identity, end tag identity, full inclusion, total identity, overlap and shared start/end point. Additionally Witt (2004) introduces meta-relations between element classes, "i. e. the set of all instances of an element". As another aspect, Sperberg-McQueen and Huitfeldt (2004) distinguish between real and *spurious* overlap. Spurious overlap exists when the instance "can be rewritten without overlap, without changing the interpretation of any character of the document". Listing 7 shows an example annotation (the second line shows the rewritten annotation). By rearranging the position of the start tags of the first occurrence of the `a` and `b` elements the overlap is eliminated.

Listing 7. Spurious overlap.

```
1 <a><b>John likes</a> Mary</b>
2 <b><a>John likes</a> Mary</b>
```

Self overlap exists when two elements of the same level overlap. Marinelli, Vitali, and Zacchiroli (2008) demonstrate this by constructing a situation where two reviewers annotate two overlapping text regions. Listing 8 shows an example XML representation.

Listing 8. Self overlap.

```
<p>Is this <comment>self <comment>overlap</comment>
</comment>?</p>
```

Since both annotators use the same comment element it is impossible to distinguish between a self overlap or proper nesting (i. e. a `comment` element containing another `comment` element).

For these reasons, several approaches to deal with overlapping markup were developed which can be classified as XML-based, XML-related or non-XML-based, and which are discussed in Stührenberg and Goecke (2008) and Marinelli, Vitali, and Zacchiroli (2008), amongst others. While the non-XML-based and XML-related options may make use of completely alternative formal models (e. g. the *Annotation Graph*, discussed in Bird and Liberman 1999 and Maeda et al. 2001, or the *Range Algebra*, which is developed by Nicol 2002b; Nicol 2002a and which forms the formal foundation for the Layered Markup and Annotation Language, LMNL, cf. Tennison 2002; Cowan, Tennison, and Piez 2006; Piez 2008), the XML-based approaches have to stick with XML's formal model. Simple XML-based solutions are the redundant encoding of information in multiple instances, the use of milestones or fragmentation which are discussed in the Guidelines of the Text Encoding Initiative (Burnard and Bauman 2011, Ch. 20, "Non-hierarchical Structures") or standoff markup (cf. Thompson and McKelvie 1997).[4] Example XML-based standoff solutions that allow concurrent hierarchies can be found in Dipper et al. (2007); Ide and Suderman (2007); Stührenberg and Goecke (2008) and Stührenberg and Jettka (2009), amongst others. Since there are XML-based solutions to the problem of overlapping hierarchies we have to challenge the formal model of a tree for XML instances.

As stated by Abiteboul, Buneman, and Suciu (2000) this formal model is only valid if we solely examine the *elements* of an XML instance. The inherent mechanism for defining unique identifiers and references by attributes of the before-mentioned ID and the respective IDREF/IDREFS token types allows for establishing links (that is, directed edges) between an element node and one or more other element node(s) elsewhere in the tree, resulting in the more general graph model. Other authors agree to this position: Polyzotis and Garofalakis (2002) discuss "general, graph-structured XML databases (where non-tree edges can arise naturally as explicit element references through id/idref attributes or XLink constructs", while Gou and Chirkova (2007) state that "ID/IDREF attributes increase the flexibility of the XML data model and extend the basic tree model to DAGs or to even more general directed graphs with cycles". Møller and Schwartzbach (2007) and Møller and Schwartzbach (2011) coin the data model as *XML Graph* (formerly known as *Summary Graph*), while Dehmer, Mehler, and Emmert-Streib (2007) define the term *Generalized Tree*, that is, a graph with a kernel hierarchical skeleton in conjunction with graph inducing peripheral edges. Following this work, Mehler (2009) classifies different *Generalized Trees* according to their non-tree edges.

There is, however, a distinctive feature that helps us draw the line between both formal models: if we want to extend the tree by adding additional edges (by using IDREF/IDREFS type attributes or other integrity feature) we have to take one step forward from well-formed XML instances to valid instances, since the inherent linking mechanism is only available by supplying a document grammar. This is why we discuss different constraint languages that can be used to define XML document grammars next.

3. XML constraint languages

In this section we will demonstrate that the constraint language (as a formalism) used to define a concrete document grammar for a given markup language has a strong influence on the formal model that can be described by an instance of this very markup language. In addition, we will classify the three most-used constraint languages according to their expressivity.

As already mentioned, XML allows for using instances without the need for a document grammar (in this case, the document cannot be assumed to be valid but only to be well-formed). If one chooses to implement a document grammar (or more general: a constraint language to define a specific document grammar) there are several available, starting with XML Document Type Definitions (DTD, cf. Bray et al. 2004 and Goldfarb and Prescod 2004), XML Schema Descriptions (XSD, often called XML Schema, cf. Walmsley 2002; Fallside 2001; Fallside and Walmsley 2004; Thompson et al. 2001; Thompson et al. 2004; Biron and Malhotra 2001 and Biron and Malhotra 2004), RELAX NG (Regular Language Description for XML New Generation, ISO/IEC 19757-2:2003; ISO/IEC 19757-2:2008) as successor of both RELAX (ISO/IEC TR 22250-1:2002) and TREX (Clark 2001), Schematron (ISO/IEC 19757-3:2006), or Document Structure Definition 2.0 (DSD and DSD2, cf. Klarlund, Møller, and Schwartzbach 1999; Klarlund, Møller, and Schwartzbach 2002 and Møller 2005). While both XML DTDs and XML Schema are standardized by the W3C, RELAX NG and Schematron are part of the ISO/IEC standard 19757, DSDL (Document Schema Definition Languages).

In principle, it is possible to differentiate XML constraint languages into two classes: grammar-based and rule-based languages. A definition for both classes can be found in van der Vlist (2003b); van der Vlist (2004) as well as in Costello and Simmons (2008):

> A grammar-based schema language specifies the structure and contents of elements and attributes in an XML instance document. For example, a grammar-based schema language can specify the presence and order of elements in an XML instance document, the number of occurrences of each element, and the contents and datatype of each element and attribute. A rule-based schema language specifies the relationships that must hold between the elements and attributes in an XML instance document. For example, a rule-based schema language can specify that the value of certain elements must conform to a rule or algorithm.

Following this definition, DTD, XSD, and RELAX NG are grammar-based constraint languages, while DSD and Schematron are rule-based constraint languages. We will only deal with grammar-based constraint languages in this section. A concrete realization will be called a "schema". Walmsley (2002, p. 4–5) defines the purpose of a schema as follows:

> A schema can be used to validate:
> – The structure of elements and attributes. For example, a product must have a number and a size, and may optionally have an effDate (effective date).
> – The order of elements. For example, number must appear before size.
> – The data values of attributes and elements, based on ranges, enumerations, and pattern matching. For example, size must be an integer between 2 and 18, effDate must be a valid date.
> – The uniqueness of values in an instance. For example, all product numbers in an instance must be unique.

While all schema languages discussed in this section support the validation of the structure of elements and attributes and the order of elements, some lack in data typing or integrity checking. Many authors have already discussed some schema languages in terms of their features: Walsh (1999) compares XML DTDs with the (to that point in time not finally released) XML Schema. Lee and Chu (2000) discuss in total six schema languages, namely DTD, XML Schema, XDR (XML-Data Reduced, cf. Frankston and Thompson 1998), SOX (Schema For Object-Oriented XML, cf. Fuchs, Maloney, and Milowski 1998; Davidson et al. 1999), Schematron and DSD and try to classify these languages according to their expressive power (with DTD as a member of the lowest class 1, and XSD, DSD, and Schematron as members of the highest class 3). Jelliffe (2001) analyses DTD, XML Schema, RELAX NG, Schematron, DSDL (not a multipart standard at that time), xlinkit (Nentwich, Emmerich, and Finkelstein 2001; Nentwich et al. 2002), Examplotron (van der Vlist 2003c) and HOOK (Jelliffe 2002) regarding the features encoding, linking, infoset, structures, static data typing, local reference integrity, web reference integrity, co-occurrence constraints and value defaulting (we will discuss some of these features later on in this section). Van der Vlist (2001) compares DTD, XML Schema, RELAX NG, Schematron, and Examplotron and discusses differences according the construction of a PSVI (Post Schema Validation Infoset), integrity, and vendor support. Møller and Schwartzbach (2006) analyze DTD, XML Schema, DSD2 and RELAX NG and precisely sum up the disadvantages of DTD and XSD (for example the use of unqualified locals in XSD which can be very confusing). In addition, the authors try to give best-practice-rules for using XML Schema which result in reducing the expressivity of this constraint language. We will compare the expressivity of DTD, XSD, and RELAX NG in the next section. Ansari, Zahid, and Doh (2009) discuss the very same four constraint languages, however, the comparative analysis remains a little shallow. In comparison, the analysis accomplished by Manguinhas (2009) discusses several aspects, includes expressive power of the four constraint languages DTD, XSD, RELAX NG, and Schematron, following the taxonomy established by Murata et al. (2005). This taxonomy is enhanced in Stührenberg and Wurm (2010) by introducing a new grammar type, namely the unambiguous restrained-competition grammar (URCG).

3.1. Expressivity of constraint languages

There is a close relationship between XML constraint languages and formal logic, therefore, it should be no surprise that a large number of authors have already discussed this topic, such as Brüggemann-Klein and Wood (1992); Brüggemann-Klein (1993); Brüggemann-Klein and Wood (1992, 1997a, 2002, 2004); Lee, Mani, and Murata (2000); Hopcroft, Motwani, and Ullman (2000); Rizzi (2001); Mani (2001); Murata, Lee, and Mani (2001); Berstel and Boasson (2002); Sperberg-McQueen (2003); Klarlund, Schwentick, and Suciu (2003); Murata et al. (2005); Martens, Neven, and Schwentick (2005); Lu et al. (2005); Kilpeläinen and Tuhkanen (2007); Comon et al. (2007); Martens, Neven, and Schwentick (2009); Gelade, Martens, and Neven (2009); Gelade et al. (2010). Kracht (2011) uses modal logic to implement a semantic for XML instances and includes the XML PATH LANGUAGE (XPath, Clark and DeRose 1999) to the discussion (similar to Benedikt and Koch 2009).

Usually, DTDs can be seen as extended context-free grammars (ECFG): “A DTD is essentially a context-free grammar, with its own notation for describing the variables and productions. […] Rather, the language for describing DTDs is essentially a CFG notation […]” (Hopcroft, Motwani, and Ullman 2000, p. 199).

An ECFG is a context-free grammar that allows regular expressions on the right side of the production rules. This is accepted by Brüggemann-Klein and Wood (1997b) or Klarlund, Schwentick, and Suciu (2003):

> Document types in SGML are defined, essentially, by bracketed, extended context-free grammars […]. The right-hand sides of productions, called model groups, are essentially regular expressions with two major differences. First, model groups allow three new operators ?, &, and +. Second, the model groups must be unambiguous in the sense of Clause 11.2.4.3 of the standard. (Brüggemann-Klein and Wood 1997b, p. 2).

Newer analysis, such as Murata, Lee, and Mani (2001); Murata et al. (2005); Møller and Schwartzbach (2006) use tree grammars to compare the expressive power of different XML constraint languages.[5] The main difference between (E)CFGs and (local) tree grammars is the fact that the former produce strings while the latter produce trees (cf. Gladky and Melčuk 1969, p. 1; Murata et al. 2005, p. 662 and p. 686–687). In addition, tree grammars are allowed to have more than one start symbol (cf. Bruggemann-Klein and Wood 1998, p. 7).

Following Murata et al. (2005) and Stührenberg and Wurm (2010), we first define a regular tree grammar before we introduce more restrictive types.

Definition 2 *A regular tree grammar (RTG) is a 4-tuple $G = (N, T, S, P)$, where:*

- *N is a finite set of nonterminals,*
- *T is a finite set of terminals,*
- *S is a set of start symbols, where S is a subset of N,*
- *P is the set of production rules of the form $A \rightarrow a(r)$, where $A \in N$, $a \in T$, and r is a regular expression over N; A is the left-hand side, a(r) is the right-hand side, a is the terminal (or label) which is introduced by the rule, and r is called the content model of this production rule.*

In addition we need the definition of an interpretation to make a statement whether a tree is valid against a grammar or not (a tree *t* is valid against an RTG *G* if there is an interpretation) introduced in Murata et al. (2005, p. 663f).

Definition 3 *An interpretation I of a tree t against a regular tree grammar G is a mapping from each node e in t to a non-terminal, denoted I(e), such that:*

- *I(e) is a start symbol when e is the root of t, and,*
- *for each node e and its child nodes $e_0, e_1, \ldots, e_n$ there exists a production rule $A \rightarrow a(r)$ in G such that*
  - *I(e) is A,*
  - *the terminal symbol (label) of e is a, and*
  - *$I(e_0), I(e_1), \ldots, I(e_n)$ matches r.*

Furthermore, we have to introduce the concept of competing rules (Stührenberg and Wurm 2010):

Definition 4 *We call two rules of a RTG competing, if they introduce the same terminal in the right hand side, but have different left hand sides (non-terminals). Thus,* $A \rightarrow a(r)$ *und* $B \rightarrow a(r^0)$ *are competing rules.*

Following Murata et al. (2005), DTDs can be classified as local tree grammars. A local tree grammar is an RTG without competing rules (without competing nonterminals). This implies that in a DTD elements are declared global (and attributes are declared local).

The next class of grammars introduced by Murata et al. (2005) is the one of single-type tree grammars which roughly corresponds to XML Schema:

Definition 5 *A single-type tree grammar is a regular tree grammar where competing nonterminals must not occur in the same content model.*

This prohibition of competing nonterminals in the same content model can be found in the XML Schema specification (cf. Thompson et al. 2004, Section 3.8.6) and is called "Element declarations consistent (EDC)" constraint:

> If the particles contains, either directly, indirectly (that is, within the particles of a contained model group, recursively) or implicitly two or more element declaration particles with the same name and target namespace, then all their type definitions must be the same top-level definition, that is, all of the following must be true:
> 1. all their type definitions must have a non-absent name.
> 2. all their type definitions must have the same name.
> 3. all their type definitions must have the same target namespace.

However, there are XML schema document grammars that have greater expressive power than single-type tree grammars. This is the case, when wildcards are used. Wildcards allow for elements or attributes without constraining the names. In this case we have to deal with a restrained-competition tree grammar (cf. Murata et al. 2005, p. 675 and the following definition given in Stührenberg and Wurm 2010).

Definition 6 *A restrained-competition grammar (RCG) is an RTG, where competing nonterminals must not occur in the same content model and with the same prefix of nonterminals; we thus disallow rules with identical left-hand side, terminals, and content models of the form* $(\Gamma A \Delta)$ *and* $(\Gamma B \Delta')$*, where A and B are competing nonterminals, and where uppercase Greek letters refer to possibly empty sequences of nonterminals.*

Again, following Murata et al. (2005), RELAX NG grammars can be classified as regular tree grammars. However, as Stührenberg and Wurm (2010) state, there is another class of tree grammars that may fit better to RELAX NG, the one of unambiguous restrained-competition grammars (URCGs).

We will not discuss this grammar class here due to space restrictions but will have a further look at a crucial point that is essential when discussing XML constraint languages under formal aspects, that is the concept of determinism.

3.2. Determinism

Determinism is an important property for XML instances, schema languages, and interpretation (and as a result: validation). If a grammar is deterministic, parsing can be much more efficient.[6]

This artificial constraint is rooted in the times of SGML when deterministic finite automata where used for context-checking of content models: "A content model cannot be ambiguous; that is, an element or character string that occurs in the document instance must be able to satisfy only one primitive content token without looking ahead in the document instance." (ISO 8879:1986, Section 11.2.4.3).

And, furthermore:

> Checking for conformance to a content model is essentially equivalent to the problem of recognizing (accepting or rejecting) regular expressions, and regular expressions therefore provide a useful theoretical basis for some aspects of content checking in SGML. It can be shown (by Kleene's theorem) that a regular expression is equivalent to a deterministic finite automaton (DFA). A parser could in theory, therefore, handle any *model group* by reducing it to a regular expression and constructing a DFA with state transition paths corresponding to the tokens of the *model group*. Practice, however, presents some difficulties. […] A[nother] problem lies with the construction of the DFA. One method is first to construct a nondeterministic finite automaton (NFA) directly from the regular expression, then to derive a deterministic equivalent. Although this and other algorithms for DFA construction are non-polynomial and hardly intractable for the human-readable models envisaged by SGML, they may be too resource-intensive for some implementations. This International Standard avoids these problems by eliminating the need to construct a DFA. This it does by prohibiting models that are ambiguous or require "look-ahead"; […] As a result, context checking can be done by simplified algorithms that use only NFAs. (Goldfarb 1991, p. 558–559).

Since there is a prohibition of a look-ahead, this form of non-ambiguity is called "1-unambiguity" by Brüggemann-Klein and Wood (1997a, p. 183) in contrast to the concept of ambiguity described by Book et al. (1971). It is quite interesting that the handling of determinism is different in SGML DTDs, XML DTDs, and XSD: While SGML forbids "ambiguous content models", the XML specification speaks of "non-deterministic content models" (Bray, Paoli, and Sperberg-McQueen 1998, Appendix E), and XML Schema introduces the Unique Particle Attribution rule (UPA):

> A content model must be formed such that during validation of an element information item sequence, the particle contained directly, indirectly or implicitly therein with which to attempt to validate each item in the sequence in turn can be uniquely determined without examining the content or attributes of that item, and without any information about the items in the remainder of the sequence. (Thompson et al. 2001, Section "3.8.6 Constraints on Model Group Schema Components", Subsection "Schema Component Constraint: Unique Particle Attribution").

Another interesting fact is, that only XML Schema has a normative prohibition of non-deterministic content models (the before-mentioned UPA), while both SGML and XML cover this issue in the non-normative section, only. In SGML and XML DTDs, the existence of an non-deterministic content model is a "non-fatal error", that is, the parser is allowed to continue the parsing process (vgl. Graham and Quin 1999, p. 279) nevertheless. There are parsers that accept the content model shown in Listing 9 which is a simple test case.

Listing 9. Example of a non-deterministic content model (DTD) which is accepted by some XML parsers.

```
<!ELEMENT a (b, b, b, b?, b?)>
```

Since XSD has introduced the normative UPA, a similar content model (cf. Listing 10) is rejected by every validating parser.

Listing 10. Example of a non-deterministic content model (XSD) which is not accepted.

```
<xs:element name="a">
 <xs:complexType>
  <xs:sequence>
   <xs:element ref="b"/>
   <xs:element ref="b"/>
   <xs:element ref="b"/>
   <xs:element ref="b" minOccurs="0" maxOccurs="1"/>
   <xs:element ref="b" minOccurs="0" maxOccurs="1"/>
  </xs:sequence>
 </xs:complexType>
</xs:element>
```

Of course, the correct way to specify a content model that allows 3 to 5 occurrences of the `b` element is either the one shown in Listing 11 or the one in Listing 12 that tries to stay as close to the DTD realization as possible.

Listing 11. Example of a deterministic content model (XSD).

```
<xs:element name="a">
 <xs:complexType>
  <xs:sequence>
   <xs:element ref="b" minOccurs="3" maxOccurs="5" />
  </xs:sequence>
 </xs:complexType>
</xs:element>
```

Listing 12. Variant of a deterministic content model (XSD).

```
<xs:element name="a">
 <xs:complexType>
  <xs:sequence>
   <xs:element ref="b"/>
   <xs:element ref="b"/>
   <xs:element ref="b"/>
   <xs:sequence minOccurs="0">
    <xs:element ref="b"/>
    <xs:element ref="b" minOccurs="0"/>
   </xs:sequence>
  </xs:sequence>
 </xs:complexType>
</xs:element>
```

Fuchs and Brown (2003) provide a detailed discussion of the UPA together with an algorithm for UPA testing.

In case of mixed content models XML DTD explicitly prohibit multiple occurrences of the same element types in the same content model (cf. Bray, Paoli, and Sperberg-McQueen 1998 and Bray et al. 2008, Section “3.2.2 Mixed Content” and Listing 13).

Listing 13. Example of a mixed content model that is not allowed in XML DTDs.

```
<!ELEMENT a (#PCDATA | b, b)*>
```

However, a similar content model is allowed in XML Schema under the UPA, since it is still deterministic (cf. Listings 14 and 15):

Listing 14. Example of a mixed content model that is allowed in XML Schema.

```
<xs:element name=“a”>
 <xs:complexType mixed=“true”>
  <xs:sequence>
   <xs:element ref=“b”/>
   <xs:element ref=“b”/>
  </xs:sequence>
 </xs:complexType>
</xs:element>
```

Listing 15. Example of a mixed content model that is not allowed in XML Schema.

```
<xs:element name="a">
 <xs:complexType mixed="true">
  <xs:sequence>
   <xs:element ref="b" minOccurs="0"/>
   <xs:element ref="b" minOccurs="0"/>
  </xs:sequence>
 </xs:complexType>
</xs:element>
```

RELAX NG in contrast has some advantages over the other two constraint languages: it allows for ordering of the element child nodes in mixed content (van der Vlist 2003b, p. 57–58), and co-occurrence constraints can be used to specify the content model of an item according to the value of another item, allowing non-deterministic content models which cannot be realized neither in DTD nor in XSD (van der Vlist 2003b, p. 62–63). Listing 16 shows such an attribute-element constraint. The content model of the section element is constrained by the value of the type attribute: if the value is set to “global”, either section or para child elements are allowed, if the value is set to “sub”, it must contain only para children.

Listing 16. Example of an attribute-element constraint in RELAX NG (compact syntax).

```
start = element text { (element. section | element.para)+ }
element. section =
  element section {
      attribute type { "global" }?,
      (element. section | element.para)+
 }+
 | element section {
       attribute type { "sub" }?,
      element.para
   }+
element.para = element para { text }
```

Note, that Listing 16 shows RELAX NG’s compact syntax which is similar to the one used by DTD. There is an alternative XML syntax available as well. In April 2012 the World Wide Web Consortium published XML Schema 1.1, which supports co-occurrence constraints known as *conditional type assignment* – either by using *type alternatives* or *assertions* (cf. Peterson et al. 2012, for XSD simple types; Gao, Sperberg-McQueen, and Thompson 2012, for XSD complex types).[7]

At least in theory, RELAX NG is the most expressive constraint language amongst the three discussed in this section. However, there may be non-formal aspects that could lead to choose another grammar formalism, such as data typing and integrity.

3.3. Data typing and integrity

Regarding data typing there are again differences between the three most-used XML constraint languages. We will discuss this issue together with the integrity feature since the ID and IDREF data types play an important role.

XML DTD has no strong datatyping. It allows the occurrence of “parsed character data”, #PCDATA, strings that do not contain the characters ‘<’, ‘&’, or the character sequence ‘]]>’ (Graham and Quin 1999, p. 23) in elements, and the use of CDATA (strings that do not contain the characters ‘<’, ‘&’ or the quotation character used to surround the string, cf. Graham and Quin 1999, p. 25) and NMTOKEN or NMTOKENS (further constrained strings that, e. g., must not contain the blank whitespace character or that use it for separating a list of NMTOKEN) in attributes. XML Schema comes with a wide range of pre-defined data types (cf. Biron and Malhotra 2001, 2004), containing integer, dates, or duration, amongst others. In addition, user-defined data types are possible as well, allowing for either constraining or extending the built-in simple types (that is strings) or defining new complex types (that is content models that contain either attributes or other elements).

RELAX NG has no data types of its own, in fact, as part of a multi-part standard it relies on externally defined data type libraries, including both generic and application-specific type libraries.[8] Usually, most users will choose to use the W3C XML Schema type library. This includes the before-mentioned pre-defined data types but not the user-defined types. In contrast to XML DTD and XSD, RELAX NG has no built-in integrity feature. In fact, even when loading the W3C XML Schema data type library defined in Biron and Malhotra (2001) and in Biron and Malhotra (2004) the linking mechanism is not supported:

> The semantics defined by [W3C XML Schema Datatypes] for the ID, IDREF and IDREFS datatypes are purely lexical and do not include the cross-reference semantics of the corresponding [XML 1.0] datatypes. The cross-reference semantics of these datatypes in XML Schema comes from XML Schema Part 1. Furthermore, the [XML 1.0] cross-reference semantics of these datatypes do not fit into the RELAX NG model of what a datatype is. Therefore, RELAX NG validation will only validate the lexical aspects of these datatypes as defined in [W3C XML Schema Datatypes]. (Clark and Kawaguchi 2001a)

However, there is the possibility of using DTD Compatibility which is a library for checking both the lexical spaces of ID and IDREF (or IDREFS) data types and supporting default values in attributes (which is otherwise not supported in RELAX NG, but for attributes in DTD and for both attributes and elements in XSD, cf. Clark and Kawaguchi 2001b and van der Vlist 2003a, p. 98–103), though the expressivity of RELAX NG content models is constrained to that of a local tree grammar:

> For compatibility with XML DTDs, RELAX NG DTD Compatibility defines an annotation that can be used to specify default attributes values. However, this can only be used for content models that do not go beyond XML DTDs in their use of attributes. (Clark 2001, Section 8: "Attributes").

It is possible to implement both datatype libraries together, although a RELAX NG processor is not required to support this behavior. In general that means that using integrity features in RELAX NG comes with a price to pay.

We will conclude this section with a short overview of the distribution of the afore-mentioned constraint languages.

3.4. Distribution of constraint languages

The remarks made by van der Vlist (2001) are still valid: the constraint language that is mostly used is XML DTD. DTDs are easy to develop – not least because of their limited expressive power – and every validating XML parser supports this grammar formalism since it is part of the XML specification. XML Schema and RELAX NG are supported by an increasing number of software products. Although XSD had a head start the results should be balanced by now. Following is a short (and by no means comprehensive) list of software that has proven useful. The editors *Emacs*[9] (in conjunction with the nXML mode[10]) and *xmloperator*[11] both support validating an XML instance against a RELAX NG schema. Other products, such as *Liquid XML Studio*[12], *Stylus Studio*[13] or *Altova XML Spy*[14] support validating instances against XSD (and DTD of course); *oXygen*[15] supports all three constraint languages. In addition, there are some stand-alone validating parsers that support DTD, XSD, and RELAX NG, such as the *Oracle Multi-Schema XML Validator (MSV)*[16].

It is interesting that more and more specifications use RELAX NG as primary constraint language (although others are supported as secondary targets as well), such as Docbook[17] (see Walsh and Muellner 1999, and Max 2012 and Rahtz 2012, both in this volume) or the schema files for the TEXT ENCODING INITIATIVE. This can be traced back to the increasing software support on the one hand, and the ease of learning of this constraint language on the other hand, since most of these specifications do not make use of RELAX NG's more expressive formal power.

As a conclusion, Table 1 compares the three discussed constraint languages in short form.

Table 1. Comparison of XML constraint languages.

	XML DTD	XSD	RELAX NG
Author	W3C		OASIS/ISO
Type	Grammar[1]		
Grammar class	LTG	STG/RCG[2]	RTG/URCG
Status	W3C Recommendation[3]		ISO standard
Syntax[†]	Proprietary	XML	XML/Proprietary
PSVI	✓ (– –)	✓	×
Structures	✓	✓	×
Data types	✓ (– –)	✓	✓ (– –)[4]
Integrity	✓[5]	✓[6]	×[7]
Content model	Deterministic[8]		Non-deterministic
XML Namespace	×	✓	✓
Use	++	+	+
Support	++	++	+ (Increasing)
Comment	Official W3C standard		Strong mathematical foundation

✓/X – Feature existent/not existent
Distribution: ++ – very strong/+ – strong/o – satisfying/– – weak/– – – very weak
† "XML" means that the serialization of the grammar uses the notation of XML instances, "Proprietary" means that another notation is used.

1 XSD could be classified as both grammar-based and object-oriented (cf. van der Vlist 2003b, p. 372).
2 RTG if wildcards are used.
3 DTD is a part of the XML specification, XSD is defined in specifications of its own.
4 Implementation of externally defined data type libraries possible.
5 Via ID/IDREF/IDREFS token type attributes. See Fan and Libkin (2002) for a formal analysis of integrity constraints in XML DTDs.
6 Via xs:ID/xs:IDREF/xs:IDREFS data types and xs:key/xs:keyref, xs:unique. See Arenas, Fan, and Libkin (2002) for a formal analysis of integrity constraints in XML Schema.
7 Via ID/IDREF/IDREFS token type attributes by using the DTD compatibility mode (which constraints the expressive power).
8 In terms of 1-unambiguity; normative in XSD.

4. Levels and layers

As the last "leg" of our markup chair we will discuss in short the differentiation between levels and layers. As already stated in the introductory part of this chapter this distinction was introduced by Witt (2004) when dealing with concurrent hierarchies:

> To avoid confusion when talking about multiply structured text and text ideally organized by multiple hierarchies, the terms "level" or "level of description" is used when referring to a logical unit, e. g. visual document structure or logical text structure. When referring to a structure organizing the text technically in a hierarchically ordered way the terms "layer" or "tier" are used. A level can be expressed by means of one or more layers and a layer can include markup information on one or more levels.

Differentiating between both the concept and its serialization is usually not necessary when handling single annotations where the concept (the *level*) and the notation (the *layer*) correlate in 1:1 (one-to-one) relation. However, other relations are possible as well, such as 1:*n*, *n*:1, or *n*:*m*. An example for a 1:*n* (one-to-many) relation would be a POS (part-of-speech) analysis in which every word class annotation is stored in a separate file. Another example would be the use of multiple POS tagger resulting in several annotation instances. In case of a *n*:1 relation (many-to-one), multiple annotation levels are comprised in a single notation, for example when combining different non-overlapping hierarchies such as syntax and morphology. A *n*:*m* relation (where both *n,m* > 1, many-to-many) involves both the splitting and mixture of levels and can be rarely observed (cf. Goecke et al. 2011).

Especially the 1:*n* relation where multiple annotations over the same primary data can be easily compared can be a useful tool for analyzing the quality of the annotation process (e. g. when comparing either human annotators or software tools). Stührenberg and Jettka (2009) describe both a format and an accompanying toolkit that allows such analysis.

5. XML – now and in the future

Since the meta language XML was introduced by Sperberg-McQueen and Bray (1997) at the ACH-ALLC 97 conference, XML has experienced a number of maturation stages. Starting as an emerging technology (although backed up by SGML's heritage) as a meta language for designing sustainable markup languages for structuring annotated data of different disciplines it is now the foundation of many today's content creation, technical protocols, and workflows in a wide range of disciplines (both scientific and economic). Content production and delivery tools such as the already mentioned DocBook, DITA (Day 2010) or

EPUB (Gylling et al. 2011) make use of XML as well as protocols such as SOAP (Mitra and Lafon 2007 and Gudgin et al. 2007) or multimedia metadata specifications such as the Extensible Metadata Platform (XMP, Adobe 2010a; Adobe 2010b; Adobe 2010c). We have now reached the point that structured markup languages – and especially XML-based markup systems – are here to stay and that their use and inclusion in today's products is self-evident. Take database systems for example. While a decade ago, relational databases have represented the dominant concept, native XML databases and hybrid approaches have been introduced to the market. The next logical step will be that XML support will be taken for granted and together with other NoSQL databases these products will take a substantial part of the database market. In general, XML-based technologies have started to loose their special nimbus they once had when the Extensible Markup Language was introduced. This is a normal stage in a maturation process and it means that it is only natural that there are already other notation formats (especially for internet-based services) such as the JavaScript Object Notation (JSON, Crockford 2006b)[18] that starts to replace XML in various but not all situations (see Crockford 2006a for a comparison between JSON and XML). In addition, efforts to reach convergence between XML as the already established technology and JSON have already been undertaken, such as Hunter (2011) or JXML[19], amongst others. However, it is quite interesting to see that one of the main expected fields of work for XML (and one of the main reasons for developing XML as a stripped-down version of SGML, see Section 1) did not materialize: the web was not conquered by XML as a meta language for structuring content but as a foundation for protocols. This conclusion is confirmed by the fact that the upcoming HTML version, HTML5 (Hickson 2012) is not an XML application, that is, it does not use XML's syntax (although there is an alternative syntax that is XML-based, the proposed syntax is not).

To sum it up, we believe that XML will still be an actively used technology ten years from now on. But the question how much it will be used and in which fields (apart from content creation and delivery) depends on whether there will be a revised XML specification (the 1.1 version, described in Bray et al. 2006 has only minor changes compared to the initial version). Discussion about the features of such a "2.0" release of XML have already started, although it is not clear if a subset of XML such as MicroXML[20] or an extended version of XML is the right answer for tomorrow's needs.

6. Conclusion – or what is missing

In this chapter we tried to give a short overview of XML-based markup languages. It was not our goal to give the uninformed reader an introduction to XML as such, but to introduce her to the different aspects of XML as a meta language that determine the resulting markup language. What we did not address due to space restrictions are other XML accompanying specifications such as XPath (Berglund et al. 2007), XSLT (Kay 2007), or XQuery (Boag et al. 2007). Especially XSLT's potential has grown between the past two versions and promises to even grow further with the upcoming version 3.0 (Kay 2010b, the version number will change after adoption) which will include streaming processing (allowing for very large input files, cf. Kay 2010a) or iterations over items in a sequence. XProc (Walsh, Milowski, and Thompson 2010) provides an easy to use approach for specifying a sequence of operations to be performed on a collection of input documents (either XML instances or raw text – since XSLT 2.0 it is possible to process non-XML input, cf. Haupt and Stührenberg 2010, for an example application).

Notes

1. Note, that SGML had both an abstract and a reference concrete syntax. While the former could be used to define SGML markup "constructs in terms of delimiter roles and character classes, but not (for the most part) specific characters" (Goldfarb 1991, p. 198), the latter assigned "specific characters to the delimiter roles and character classes" (Goldfarb 1991, p. 198). A delimiter role is a syntactic token representing a delimiter string, that is the concrete sign that is used in the textual representation to separate the (begin and end of the) annotation from the primary data.
2. We will not go into details here why it became necessary to develop XML as a proper subset of SGML (by adding ISO/IEC JTC1/WG4 1997 to the standard) after the success of HTML. Milestones on the route to XML are enlisted in DeRose (1997, p. 182–183), including BASIC SGML and MINIMAL SGML (both are parts of the SGML specification itself, cf. Sections 15.1.1 and 15.1,2, and Goldfarb 1991, S. 478 f.), PSGML (Sperberg-McQueen 1992), the TEI Interchange Subset (Sperberg-McQueen and Burnard 1993, Ch. 42, "Formal Grammar for the TEI-Interchange-Format Subset of SGML"), SGML Lite (Bos 1995) or the MINIMIZED GENERALIZED MARKUP LANGUAGE (MGML, Bray, Tim 1996b). The design goals for the specification are not only stated in the standard itself but in Bray (1996a); Sperberg-McQueen and Bray (1997) and Goossens and Rahtz (1999) as well.
3. Our single change is that we add the restriction that the order of the nodes should be strict.
4. Standoff annotation was introduced in Ide, Priest-Dorman, and Véronis (1996) as "Remote Annotation". Further use of the concept can be found in the TIPSTER pro-

ject (Architecture Committee for the TIPSTER Text Phase II Program 1996). While McEnery, Xiao, and Tono (2006, p. 44) call it "Standalone Annotation", this form of markup is nowadays widely adopted under the current term and is discussed in the current version of the TEI Guidelines as well.

5. Due to space restrictions we will not give an introduction to tree grammars here. See Gladky and Melčuk (1969) or Gécseg and Steinby (1997) for a detailed introduction.
6. Although there are already algorithms discussed that allow for the efficient parsing of non-deterministic content models, cf. work done by Kilpeläinen (1999); Kawaguchi (2001); Murata et al. (2005); Hosoya (2010) and Kilpeläinen (2011).
7. Another option is the use of embedded Schematron business rules (cf. Robertson 2002) or the emulation of a similar behavior via xsi:type attributes or xs:key elements (cf. van der Vlist 2003a, p. 65 and van der Vlist 2007, p. 45–51).
8. RELAX NG includes a primitive data type system, but it contains only the two data types token and string which differ only in whitespace processing (van der Vlist 2003b, p. 70).
9. Cf. http://directory.fsf.org/project/emacs/.
10. Cf. http://www.thaiopensource.com/nxml-mode/.
11. Cf. http://www.xmloperator.net/.
12. Cf. http://www.liquid-technologies.com/xml-studio.aspx.
13. Cf. http://www.stylusstudio.com/xml_product_index.html.
14. Cf. http://www.altova.com/xmlspy.html.
15. Cf. http://www.oxygenxml.com.
16. Cf. http://msv.java.net/.
17. Cf. http://www.docbook.org/schemas/.
18. More Details can be found at http://www.json.org/.
19. More details can be found at http://xml.calldei.com/JsonXML.
20. An Editor's Draft can be found at http://home.ccil.org/~cowan/MicroXML.html.

References

Abiteboul, Serge, Peter Buneman and Dan Suciu
2000 *Data on the Web: From Relations to Semistructured Data and XML*. San Francisco, California: Morgan Kaufmann Publishers.

Adobe
2010a *Extensible Metadata Platform (XMP) Specification Part 1: Data Model, Serialization, and Core Properties*. Tech. rep. Adobe Systems Incorporated.
2010b *Extensible Metadata Platform (XMP) Specification Part 2: Additonal Properties*. Tech. rep. Adobe Systems Incorporated.
2010c *Extensible Metadata Platform (XMP) Specification Part 3: Storage in Files*. Tech. rep. Adobe Systems Incorporated.

Aguera, Helen, Robert A. Amsler, David T. Barnard, Lou Burnard, Roy Byrd, Nicoletta Calzolari, David Chesnutt, Yaacov Choueka, Jacques Dendien, Paul A. Fortier, Thomas Hickey, Susan Hockey, Nancy M. Ide, Stig Johansson, Randall

Jones, Robert Kraft, Ian Lancashire, D. Terence Langendoen, Charles (Jack) Meyers, Junichi Nakamura, Wilhelm Ott, Eugenio Picchi, Carol Risher, Jane Rosenberg, Jean Schumacher, J. Penny Small, C. M. Sperberg-McQueen, Paul Tombeur, Frank Tompa, Donald E. Walker, and Antonio Zampolli.

1987 *TEI PCP1. The Preparation of Text Encoding Guidelines. Closing Statement of the Vassar Planning Conference*. URL: http://www.tei-c.org/Vault/SC/teipcp1.txt. 2012-02-04.

Ansari, Muhammad Shahid, Norman Zahid and Kyung-Goo Doh

2009 "A Comparative Analysis of XML Schema Languages". In: *Database Theory and Application. International Conference, DTA 2009, Held as Part of the Future Generation Information Technology Conference, FGIT 2009, Jeju Island, Korea, December 10–12, 2009. Proceedings*. Ed. by Dominik Slezak et al. Vol. 64. Berlin, Heidelberg: Springer, pp. 41–48.

Architecture Committee for the TIPSTER Text Phase II Program

1996 *TIPSTER Text Phase II Architecture Concept*. Tech. rep. National Institute of Standards and Technology. URL: http://www-nlpir.nist.gov/related_projects/tipster/docs/con112.doc. 2012-02-04

Arenas, Marcelo, Wenfei Fan and Leonid Libkin

2002 "What's Hard about XML Schema Constraints". In: *Proceedings of the 13th International Conference on Database and Expert Systems Applications*. Vol. 2453. Lecture Notes In Computer Science. Berlin, Heidelberg: Springer, pp. 269–278.

Benedikt, Michael and Christoph Koch

2009 "XPath Leashed". In: *ACM Computing Surveys* 41, 3:1–3:54.

Berglund, Anders, Scott Boag, Don Chamberlin, Mary F. Fernández, Michael Kay, Jonathan Robie, and Jérôme Siméon.

2007 *XML Path Language (XPath). Version 2.0*. W3C Recommendation 23 January 2007. World Wide Web Consortium. (W3C). URL: http://www.w3.org/TR/2007/REC-xpath20-20070123/. 2012-02-04.

Berners-Lee, Tim and Mark Fischetti

1999 *Weaving the Web*. New York, NY, USA: HarperCollins.

Berstel, Jean and Luc Boasson

2002 "Formal properties of XML grammars and languages". In: *Acta Informatica* 38, pp. 649–671.

Bird, Steven and Mark Liberman

1999 "Annotation graphs as a framework for multidimensional linguistic data analysis". In: *Proceedings of the Workshop "Towards Standards and Tools for Discourse Tagging"*. Association for Computational Linguistics, pp. 1–10.

Biron, Paul V. and Ashok Malhotra

2001 *XML Schema Part 2: Datatypes*. W3C Recommendation 02 May 2001. World Wide Web Consortium. URL: http://www.w3.org/TR/2001/REC-xmlschema-2-20010502/. 2012-02-04.

2004 *XML Schema Part 2: Datatypes Second Edition*. W3C Recommendation 28 October 2004. World Wide Web Consortium (W3C). URL: http://www.w3.org/TR/2004/REC-xmlschema-2-20041028/. 2012-02-04.

Boag, Scott, Don Chamberlin, Mary F. Fernández, Daniela Florescu, Jonathan Robie, and Jérôme Siméon
2007 *XQuery 1.0: An XML Query Language*. W3C Recommendation 23 January 2007. World Wide Web Consortium (W3C). URL: http://www.w3.org/TR/2007/REC-xquery-20070123/. 2012-02-04.

Book, Ronald et al.
1971 "Ambiguity in Graphs and Expressions". In: *IEEE Transactions on Computers* 20, pp. 149–153.

Bos, Bert
1995 *SGML-Lite – an easy to parse subset of SGML*. 4 July1995. URL: http://www.w3.org/People/Bos/Stylesheets/SGML-Lite.html. 2012-02-04.

Bray, Tim
1996a *Draft DD-1996-0001 – Design Principles for XML*. URL: http://www.textuality.com/sgml-erb/dd-1996-0001.html. 2012-02-04.
1996b *MGML – an SGML Application for Describing Document Markup Languages*. Unpublished draft paper for SGML '96. URL: http://www.textuality.com/mgml/index.html. 2012-02-04.

Bray, Tim, Jean Paoli, and C. M. Sperberg-McQueen
1998 *Extensible Markup Language (XML) 1.0*. W3C Recommendation 10 February 1998. World Wide Web Consortium (W3C). URL: http://www.w3.org/TR/1998/REC-xml-19980210.html. 2012-02-04.

Bray, Tim, Jean Paoli, C. M. Sperberg-McQueen, Eve Maler, and François Yergeau
2004 *Extensible Markup Language (XML) 1.0 (Third Edition)*. W3C Recommendation 04 February 2004. World Wide Web Consortium (W3C). URL: http://www.w3.org/TR/2004/REC-xml-20040204/. 2012-02-04.

Bray, Tim, Jean Paoli, C. M. Sperberg-McQueen, Eve Maler, François Yergeau, and John Cowan
2006 *Extensible Markup Language (XML) 1.1 (Second Edition)*. W3C Recommendation 16 August 2006, edited in place 29 September 2006. World Wide Web Consortium (W3C). URL: http://www.w3.org/TR/2006/REC-xml11-20060816. 2012-02-04.
2008 *Extensible Markup Language (XML) 1.0 (Fifth Edition)*. W3C Recommendation 26 November 2008. World Wide Web Consortium (W3C). URL: http://www.w3.org/TR/2008/REC-xml-20081126/. 2012-02-04.

Brüggemann-Klein, Anne
1993 "Formal Models in Document Processing". Habilitation. Albert-Ludwig-Universität zu Freiburg i.Br.

Brüggemann-Klein, Anne and Derick Wood
1992 "Deterministic Regular Languages". In: *STACS 92. 9th Annual Symposium on Theoretical Aspects of Computer Science Cachan, France, February 13–15, 1992 Proceedings*. Ed. by Alain Finkel and Matthias Jantzen. Vol. 577. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, pp. 173–184.
1997a "One-Unambiguous Regular Languages". In: *Information and computation* 142, pp. 182–206.

1997b “The Validation of SGML Content Models”. In: *Mathematical and Computer Modelling* 25.4, pp. 73–84.
1998 *Regular Tree Languages Over Non-Ranked Alphabets.* Unpublished manuscript. URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.50.5397&rep=rep1&type=pdf. 2012-02-04.
2002 *The parsing of extended context-free grammars*. HKUST Theoretical Computer Science Center Research Report HKUST-TCSC-2002-08. The Hong Kong University of Science and Technology Library.
2004 “Balanced Context-Free Grammars, Hedge Grammars and Push-down Caterpillar Automata”. In: *Proceedings of Extreme Markup Languages*. Montréal, Québec.

Burnard, Lou and Syd Bauman (eds.)
2011 *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. Oxford et al.: published for the TEI Consortium by Humanities Computing Unit, University of Oxford.

Clark, James
2001 *The Design of RELAX NG*. URL: http://www.thaiopensource.com/relaxng/design.html. 2012-02-04.

Clark, James and Steven J. DeRose
1999 *XML Path Language (XPath). Version 1.0.* W3C Recommendation 16 November 1999. World Wide Web Consortium (W3C). URL: http://www.w3.org/TR/1999/REC-xpath-19991116/. 2012-02-04.

Clark, James and Kohsuke Kawaguchi
2001a *Guidelines for using W3C XML Schema Datatypes with RELAX NG.* Committee Specification 7 September 2001. OASIS – Organization for the Advancement of Structured Information Standards. URL: http://www.relaxng.org/xsd-20010907.html.2012-02-04.
2001b *RELAX NG DTD Compatibility.* Committee Specification 3 December 2001. OASIS – Organization for the Advancement of Structured Information Standards. URL: http://www.relaxng.org/compatibility-20011203.html. 2012-02-04.

Comon, Hubert, Max Dauchet, Rémi Gilleron, Christof Löding, Florent Jacquemard, Denis Lugiez, Sophie Tison, and Marc Tommasi
2007 *Tree Automata Techniques and Applications. Release 12 October 2007.* URL: http://www.grappa.univ-lille3.fr/tata. 2012-02-04.

Coombs, J. H., Allen H. Renear, and Steven J. DeRose
1987 “Markup Systems and the Future of Scholarly Text Processing”. In: *Communications of the ACM* 30.11, pp. 933–947.

Costello, Roger L. and Robin A. Simmons
2008 *Tutorials on Schematron.* URL: http://www.xfront.com/schematron/index.html. 2012-02-04.

Cowan, John, Jeni Tennison, and Wendell Piez
2006 “LMNL Update”. In: *Proceedings of Extreme Markup Languages*. Montréal, Québec.

Cowan, John and Richard Tobin
2004 *XML Information Set (Second Edition).* W3C Recommendation 4 February 2004. World Wide Web Consortium (W3C). URL: http://www.w3.org/TR/2004/REC-xml-infoset-20040204/. 2012-02-04.

Crockford, Douglas
2006a “JSON: The Fat-Free Alternative to XML”. In: *Proceedings of XML 2006 conference*. Boston.
2006b *The application/json Media Type for JavaScript Object Notation (JSON)*. Request for Comment (RFC) 4627. Internet Engineering Task Force. URL: http://www.ietf.org/rfc/rfc4627. 2012-02-04.
Davidson, Andrew, Matthew Fuchs, Mette Hedin, Mudita Jain, Jari Koistinen, Chris Lloyd, Murray Maloney, and Kelly Schwarzhof
1999 *Schema for Object-Oriented XML 2.0.* W3C Note. World Wide Web Consortium (W3C). URL: http://www.w3.org/TR/NOTE-SOX2012-02-04.
Day, Don
2010 *Darwin Information Typing Architecture (DITA) Version 1.2*. OASIS Standard. OASIS – Organization for the Advancement of Structured Information Standards. URL: http://docs.oasis-open.org/dita/v1.2/os/spec/DITA1.2-spec.html. 2012-02-04.
Dehmer, Matthias, Alexander Mehler and Frank Emmert-Streib
2007 “Graph-theoretical Characterizations of Generalized Trees”. In: *Proceedings of the 2007 International Conference on Machine Learning: Models, Technologies & Applications (MLMTA’07), June 25–28, 2007, Las Vegas*, pp. 113–117.
Dekhtyar, Alex and Ionut Emil Iacob
2005 “A framework for management of concurrent XML markup”. In: *Data & Knowledge Engineering* 52.2, pp. 185–208.
Department of Computing Services
1980 *Waterloo SCRIPT Reference Manual*. SLAC-SCS-USER-NOTE-119. Waterloo, Ontario, Canada: University of Waterloo.
DeRose, Steven J.
1997 *The SGML FAQ Book. Understanding the Foundation of HTML and XML*. Electronic Publishing Series. Boston, Dordrecht, London: Kluwer Academic Publishers.
DeRose, Steven J., David G. Durand, Elli Mylonas, and Allen H. Renear
1990 “What is text, really?” In: *Journal of Computing in Higher Education* 1.2, pp. 3–26.
Dipper, Stefanie, Michael Götze, Uwe Küssner, and Manfred Stede
2007 “Representing and Querying Standoff XML”. In: *Datenstrukturen für linguistische Ressourcen und ihre Anwendungen. Data Structures for Linguistic Resources and Applications. Proceedings of the Biennial GLDV Conference 2007*. Ed. by Georg Rehm, Andreas Witt, and Lothar Lemnitzer. Tübingen: Gunter Narr Verlag, pp. 337–346.
Durand, David G.
1999 “Palimpsest: Change-Oriented Concurrency Control for the Support of Collaborative Applications”. Dissertation. Boston University Graduate School of Arts and Sciences.
Durusau, Patrick and Matthew Brook O’Donnell
2002 “Concurrent Markup for XML Documents”. In: *Proceedings of the XML Europe 2002*. Barcelona, Spain.
Fallside, David C.
2001 *XML Schema Part 0: Primer*. World Wide Web Consortium.

Fallside, David C. and Priscilla Walmsley
2004 *XML Schema Part 0: Primer Second Edition.* W3C Recommendation 28 October 2004. World Wide Web Consortium (W3C). URL: http://www.w3.org/TR/2004/REC-xmlschema-0-20041028/. 2012-02-04.
Fan, Wenfei and Leonid Libkin
2002 "On XML integrity constraints in the presence of DTDs". In: *Journal of the ACM* 49.3, pp. 368–406.
Frankston, Charles and Henry S. Thompson
1998 *XML-Data reduced*. Draft. Microsoft & University of Edinburgh.
Fuchs, Matthew and Allen Brown
2003 "Supporting UPA and restriction on an extension of XML Schema". In: *Proceedings of Extreme Markup Languages*. Montréal, Québec.
Fuchs, Matthew, Murray Maloney and Alex Milowski
1998 *Schema for Object-Oriented XML.* Submission to the World Wide Web Consortium. World Wide Web Consortium (W3C). URL: http://www.w3.org/TR/1998/NOTE-SOX-19980930/. 2012-02-04.
Gao, Shudi (Sandy), C. M. Sperberg-McQueen, and Henry S. Thompson
2012 *W3C XML Schema Definition Language (XSD)* 1.1 Part 1: Structures. W3C Recommendation 5 April 2012. World Wide Web Consortium (W3C). URL: http://www.w3.org/TR/2012/REC-xmlschema11-1-20120405/. 2012-02-04.
Gécseg, Ferenc and Magnus Steinby
1997 "Tree Languages". In: *Handbook of Formal Languages*. Ed. by Grzegorz Rozenberg and Arto Salomaa. Vol. 3: Beyond Words. Berlin Heidelberg New York Tokyo: Springer-Verlag, pp. 1–68.
Gelade, Wouter, Wim Martens, and Frank Neven
2009 "Optimizing Schema Languages for XML: Numerical Constraints and Interleaving". In: *SIAM Journal on Computing* 38.5, pp. 2021–2043.
Gelade, Wouter, Wim Martens, and Frank Neven
2010 "Simplifying XML schema: single-type approximations of regular tree languages". In: *Proceedings of the twenty-ninth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems of data*. PODS '10. New York, NY, USA: ACM, pp. 251–260.
Gladky, A. V. and I. A. Melčuk
1969 "Tree grammars (Δ-grammars)". In: *Proceedings of the 1969 conference on Computational linguistics*. COLING '69. Stroudsburg, PA, USA: Association for Computational Linguistics, pp. 1–7.
Goecke, Daniela, Dieter Metzing, Harald Lüngen, Maik Stührenberg, and Andreas Witt
2011 "Different Views on Markup. Distinguishing levels and layers". In: *Linguistic Modeling of Information and Markup Languages. Contribution to Language Technology.* Ed. by Andreas Witt and Dieter Metzing. Vol. 41. Text, Speech and Language Technology. Springer.
Goldfarb, Charles F.
1973 *Design Considerations for Integrated Text Processing Systems*. Technical Report 320-2094. IBM Cambridge Scientific Center. URL: http://www.sgmlsource.com/history/G320- 2094/G320- 2094.htm. 2012-02-04.
1991 *The SGML Handbook*. Ed. by Yuri Rubinsky. Oxford: Oxford University Press.

Goldfarb, Charles F., Edward J. Mosher and Theodore I. Peterson
1970 “An Online System for Integrated Text Processing”. In: *Proceedings of the 33rd Annual Meeting of the American Society for Information Science (ASIS Proceedings)*. Vol. 7.
Goldfarb, Charles F. and Paul Prescod
2004 *The XML Handbook*. 5th ed. The Charles F. Goldfarb Definite XML Series. Prentice Hall.
Goossens, Michel and Sebastian Rahtz
1999 *The LaTeX Web Companion*. Addison-Wesley Series on Tools and Techniques for Computer Typesetting. Addison-Wesley Professional.
Gou, Gang and Rada Chirkova
2007 “Efficiently Querying Large XML Data Repositories: A Survey”. In: *IEEE Transactions on Knowledge and Data Engineering* 19.10, pp. 1381–1403.
Graham, Ian S. and Liam R. E. Quin
1999 *XML Specification Guide*. Ed. by Cary Sullivan. Wiley Computer Publishing.
Gudgin, Martin, Marc Hadley, Noah Mendelsohn, Jean-Jacques Moreau, Henrik Frystyk Nielsen, Anish Karmakar, and Yves Lafon
2007 SOAP Version 1.2 Part 1: *Messaging Framework (Second Edition).* W3C Recommendation 27 April 2007. World Wide Web Consortium (W3C). URL: http://www.w3.org/TR/2007/REC-soap12-part1-20070427/. 2012-02-04.
Gylling, Markus, William McCoy, Elika Etemad, and Matt Garrish
2011 *EPUB Content Documents 3.0.* Recommended Specification 11 October 2011. International Digital Publishing Forum. URL: http://idpf.org/epub/30/spec/epub30-contentdocs-20111011.html. 2012-02-04.
Haupt, Stefanie and Maik Stührenberg
2010 “Automatic upconversion using XSLT 2.0 and XProc”. In: *Proceedings of Balisage: The Markup Conference*. Vol. 5. Balisage Series on Markup Technologies. Montréal, Québec.
Hickson, Ian
2012 *HTML5. A vocabulary and associated APIs for HTML and XHTML.* W3C Working Draft 29 March 2012. World Wide Web Consortium (W3C). URL: http://www.w3.org/TR/2012/WD-html5-20120329/. 2012-02-04.
Hopcroft, John E., Rajeev Motwani and Jeffrey D. Ullman
2000 *Introduction to Automata Theory, Languages, and Computation*. 2nd ed. Amsterdam: Addison Wesley Longman.
Hosoya, Haruo
2010 *Foundations of XML Processing. The Tree-Automata Approach*. Cambridge University Press.
Huitfeldt, Claus and C. M. Sperberg-McQueen
2004 “Markup Languages and Complex Documents (MLDC)”. Presentation given at the colloquium of the HKI (Historisch Kulturwissenschaftliche Informationsverarbeitung) of the University of Cologne, 10 Decembre 2004. URL: http://old.hki.uni-koeln.de/material/huitfeldt.ppt. 2012-02-04.
Hunter, Jason
2011 “A JSON Facade on MarkLogic Server”. In: *XML Prague 2011 Conference Proceedings*. Ed. by Jirka Kosek. 2011-519. Institute for Theoretical Computer Science. Prag.

Ide, Nancy M.
1994 "Encoding standards for large text resources: the Text Encoding Initiative". In: *Proceedings of the 15th conference on Computational linguistics*. Morristown, NJ, USA: Association for Computational Linguistics, pp. 574–578.
Ide, Nancy M., Greg Priest-Dorman and Jean Véronis
1996 *Corpus Encoding Standard (CES)*. Tech. rep. Expert Advisory Group on Language Engineering Standards (EAGLES).
Ide, Nancy M. and Keith Suderman
2007 "GrAF: A Graph-based Format for Linguistic Annotations". In: *Proceedings of the Linguistic Annotation Workshop*. Prague, Czech Republic: Association for Computational Linguistics, pp. 1–8.
ISO 8879:1986
Information Processing – Text and Office Information Systems – Standard Generalized Markup Language. International Standard. Geneva: International Organization for Standardization.
ISO/IEC 19757-2:2003
Information technology – Document Schema Definition Language (DSDL) – Part 2: Regular-grammar-based validation – RELAX NG (ISO/IEC 19757-2). International Standard. Geneva: International Organization for Standardization.
ISO/IEC 19757-2:2008
Information technology – Document Schema Definition Language (DSDL) – Part 2: Regular-grammar-based validation – RELAX NG (ISO/IEC 19757-2). International Standard. Geneva: International Organization for Standardization.
ISO/IEC 19757-3:2006
Information technology – Document Schema Definition Language (DSDL) – Part 3: Rule-based validation – Schematron. International Standard. Geneva: International Organization for Standardization.
ISO/IEC JTC1/WG4
1997 *ISO 8879 TC2*. Technical Corrigendum. ISO/IEC. URL: http://www1.y12.doe.gov/capabilities/sgml/wg8/document/1955.htm. 2012-02-04.
ISO/IEC TR 22250-1:2002
Information technology – Document description and processing languages – Regular Language Description for XML – Part 1: RELAX Core. International Standard. Geneva: International Organization for Standardization.
Jelliffe, Rick
2001 *The Current State of the Art of Schema Languages for XML*. Presentation given at the XML Asia Pacific 2001 Conference. URL: http://www.estebel.com/articles/RickJelliffe.pdf. 2012-02-04.
2002 *Resource Directory (RDDL) for Hook 0.2. A One-Element Language for Validation of XML Documents based on Partial Order.* URL: http://xml.ascc.net/hook/. 2012-02-04.
Kawaguchi, Kohsuke
2001 *Ambiguity Detection Of RELAX Grammars.* URL: http://www.kohsuke.org/relaxng/ambiguity/AmbiguousGrammarDetection.pdf. 2012-02-04.

Kay, Michael
2007 *XSL Transformations (XSLT) Version 2.0*. W3C Recommendation 23 January 2007. World Wide Web Consortium (W3C). URL: http://www.w3.org/TR/2007/REC-xslt20-20070123/. 2012-02-04.

Kay, Michael
2010a "A Streaming XSLT Processor". In: *Proceedings of Balisage: The Markup Conference*. Vol. 5. Balisage Series on Markup Technologies. Montréal, Québec.
2010b *XSL Transformations (XSLT) Version 2.1*. W3C Working Draft. World Wide Web Consortium.

Kilpeläinen, Pekka
1999 "SGML and XML Content Models". In: *Markup Languages – Theory & Practice* 1.2, pp. 53–76.
2011 "Checking determinism of XML Schema content models in optimal time". In: *Information Systems* 36.3, pp. 596–617.

Kilpeläinen, Pekka and Rauno Tuhkanen
2007 "One-unambiguity of regular expressions with numeric occurrence indicators". In: *Information and Computation* 205.6, pp. 890–916.

Klarlund, Nils, Anders Møller, and Michael I. Schwartzbach
1999 *Document Structure Description 1.0*. Technical Report. BRICS (Basic Research in Computer Science, Aarhus University). URL: http://www.brics.dk/DSD/dsddoc.html. 2012-02-04.
2002 "The DSD Schema Language". In: *Automated Software Engineering* 9.3, pp. 285–319.

Klarlund, Nils, Thomas Schwentick and Dan Suciu
2003 "XML: Model, Schemas, Types, Logics, and Queries". In: *Logics for Emerging Applications of Databases*. Ed. by Jan Chomicki, Ron van der Meyden, and Gunter Saake. Berlin, Heidelberg: Springer, pp. 1–41.

Knuth, Donald E.
1984 *The $T_EX$book*. Addison-Wesley.

Kracht, Marcus
2011 "Modal Logic Foundations of Markup Structures in Annotation Systems". In: *Modeling, Learning and Processing of Text Technological Data Structures.* Ed. by Alexander Mehler, Kai-Uwe Kühnberger, Henning Lobin, Harald Lüngen, Angelika Storrer, and Andreas Witt. Studies in Computational Intelligence. Dordrecht: Springer.

Lee, Dongwon and Wesley W. Chu
2000 "Comparative Analysis of Six XML Schema Languages". In: *ACM SIGMOD Record* 29.3, pp. 76–87.

Lee, Dongwon, Murali Mani, and Makoto Murata
2000 *Reasoning about XML Schema Languages using Formal Language Theory.* Technical Report RJ# 10197, Log# 95071. IBM Almaden Research Center. URL: http://pike.psu.edu/publications/ibm-tr-2000.pdf. 2012-02-04.

Liu, Jixue, Millist Vincent and Chengfei Liu
2003 "Local XML functional dependencies". In: *Proceedings of the 5th ACM international workshop on Web information and data management*. WIDM '03. New York, NY, USA: ACM, pp. 23–28.

Lu, Shiyong, Yezhou Sun, Mustafa Atay, and Farshad Fotouhi
2005 "On the consistency of XML DTDs". In: *Data & Knowledge Engineering* 52.2, pp. 231–247.

Madnick, Stuart E. and Allen Moulton
1968 "Script, An On-Line Manuscript Processing System". In: *IEEE Transactions on Engineering Writing and Speech* 11.2, pp. 92–100.

Maeda, Kazuaki, Steven Bird, Xiaoyi Ma, and Haejoong Lee
2001 "The Annotation Graph Toolkit: Software Components for Building Linguistic Annotation Tools". In: *HLT '01: Proceedings of the first international conference on Human language technology research*. Morristown, NJ, USA: Association for Computational Linguistics, pp. 1–6.

Maler, Eve and Jeanne El Andaloussi
1995 *Developing SGML DTDs: From Text to Model to Markup*. Upper Saddle River, New Jersey: Prentice Hall PTR.

Manguinhas, Hugo
2009 "Schema Languages for XML". In: *Actas do INForum – Simpósio de Informática 2009.* Ed. by Luis Rodrigues and Rui Lopes. Lisbon, Portugal: Faculdade de Ciências da Universidade de Lisboa, pp. 299–310. URL: http://inforum.org.pt/INForum2009/docs/full/paper_35.pdf. 2012-02-04.

Mani, Murali
2001 "Keeping chess alive: Do we need 1-unambiguous content models?" In: *Proceedings of Extreme Markup Languages*. Montréal, Québec.

Marinelli, Paolo, Fabio Vitali, and Stefano Zacchiroli
2008 "Towards the unification of formats for overlapping markup". In: *New Review of Hyper-media and Multimedia* 14.1, pp. 57–94.

Martens, Wim, Frank Neven and Thomas Schwentick
2005 "Which XML Schemas Admit 1-Pass Preorder Typing?" In: *Database Theory – ICDT 2005*. Ed. by Thomas Eiter and Leonid Libkin. Vol. 3363. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, pp. 68–82.
2009 "Complexity of Decision Problems for XML Schemas and Chain Regular Expressions". In: *SIAM Journal on Computing* 39.4, pp. 1486–1530.

Max, Aurélien
2012 "Document authoring: from word processing to text generation". In: *Handbook of Technical Communication.* Ed. by Alexander Mehler, Laurent Romary, and Dafydd Gibbon. Berlin/Boston: De Gruyter.

McEnery, Anthony, Richard Xiao and Yukio Tono
2006 *Corpus-Based Language Studies. An Advanced Resource Book*. Routledge Applied Linguistics. Oxon: Routledge.

Mehler, Alexander
2009 "Generalised Shortest Paths Trees: A Novel Graph Class Applied to Semiotic Networks". In: *Analysis of Complex Networks: From Biology to Linguistics*. Ed. by Matthias Dehmer and Frank Emmert-Streib. Weinheim: Wiley-VCH, pp. 175–220.

Mitra, Nilo and Yves Lafon
2007 *SOAP Version 1.2 Part 0: Primer (Second Edition).* W3C Recommendation 7 April 2007. World Wide Web Consortium (W3C). URL: http:// www.w3.org/TR/2007/REC-soap12-part0-20070427/. 2012-02-04.

Mittelbach, Frank, Michel Goossens, Johannes Braams, David Carlisle, and Chris Rowley
2006 *The LaTeX Companion*. 2nd ed. Addison-Wesley Series on Tools and Techniques for Computer Typesetting. Addison-Wesley Professional.

Møller, Anders
2005 *Document Structure Description 2.0*. Technical Report. BRICS (Basic Research in Computer Science, Aarhus University). URL: http://www.brics.dk/DSD/dsd2.html. 2012-02-04.

Møller, Anders and Michael I. Schwartzbach
2006 "Schema Languages". In: *An Introduction to XML and Web Technologies*. Harlow, England: Addison-Wesley.
2007 "XML Graphs in Program Analysis". In: *PEPM '07: Proceedings of the 2007 ACM SIGPLAN symposium on Partial evaluation and semantics-based program manipulation*. New York, NY, USA: ACM.
2011 "XML Graphs in Program Analysis". In: *Science of Computer Programming* 76.6, pp. 492–515.

Murata, Makoto, Dongwon Lee and Murali Mani
2001 "Taxonomy of XML Schema Languages using Formal Language Theory". In: *Proceedings of Extreme Markup Languages*. Montréal, Québec.

Murata, Makoto, Dongwon Lee, Murali Mani, and Kohsuke Kawaguchi
2005 "Taxonomy of XML Schema Languages Using Formal Language Theory". In: *ACM Transactions on Internet Technology* 5.4, pp. 660–704.

Nentwich, Christian, Wolfgang Emmerich and Anthony Finkelstein
2001 *xlinkit: links that make sense*. Tech. rep. Department of Computer Science, University College London. URL: http://www.cs.ucl.ac.uk/staff/a.finkelstein/papers/htxlinkit.pdf. 2012-02-04.

Nentwich, Christian, Licia Capra, Wolfgang Emmerich, and Anthony Finkelstein
2002 "xlinkit: A Consistency Checking and Smart Link Generation Service". In: *Transactions on Internet Technology (TOIT)* 2.2, pp. 151–185. URL: http://www.cs.ucl.ac.uk/staff/a.finkelstein/papers/acmxlinkit.pdf. 2012-02-04.

Ng, Wilfred
2002 "Maintaining Consistency of Integrated XML Trees". In: *Proceedings of the Third International Conference on Advances in Web-Age Information Management*. WAIM '02. London, UK: Springer-Verlag, pp. 145–157.

Nicol, Gavin Thomas
2002a *Attributed Range Algebra. Extending Core Range Algebra to Arbitrary Structures*.
2002b "Core Range Algebra: Toward a Formal Model of Markup". In: *Proceedings of Extreme Markup Languages*.

Peterson, David, Shudi (Sandy) Gao, Ashok Malhotra, C. M. Sperberg-McQueen, and Henry S. Thompson
2012 *W3C XML Schema Definition Language (XSD) 1.1 Part 2: Datatypes*. W3C Recommendation 5 April 2012. World Wide Web Consortium (W3C). URL: http://www.w3.org/TR/2012/REC-xmlschema11-2-20120405/. 2012-07-04.

Pianta, Emanuele and Luisa Bentivogli

2004 "Annotating Discontinuous Structures in XML: the Multiword Case". In: *Proceedings of LREC 2004 Workshop on "XML-based richly annotated corpora"*. Lisbon, Portugal, pp. 30–37.

Piez, Wendell

2008 "LMNL in Miniature". Presentation given at the Goddag workshop, Amsterdam, 1–5 December 2008. URL: http://www.piez.org/wendell/LMNL/Amsterdam2008/presentation-slides.html. 2012-02-04.

Polyzotis, Neoklis and Minos Garofalakis

2002 "Statistical Synopses for Graph-Structured XML Databases". In: *Proceedings of the 2002 ACM SIGMOD International Conference on Management of Data, Madison, Wisconsin, June 3–6, 2002*. Ed. by Michael J. Franklin, Bongki Moon, and Anastassia Ailamaki. ACM, pp. 358–369.

Raggett, Dave

1997 *HTML 3.2 Reference Specification*. W3C Recommendation 14 January 1997. World Wide Web Consortium. URL: http://www.w3.org/TR/REC-html32-19970114. 2012-02-04.

Rahtz, Sebastian

2012 "Representation of documents in technical communication". In: *Handbook of Technical Communication*. Ed. by Alexander Mehler, Laurent Romary, and Dafydd Gibbon. Berlin/Boston: De Gruyter.

Renear, Allen H., Elli Mylonas and David G. Durand

1996 "Refining our notion of what text really is: The problem of overlapping hierarchies". In: *Research in Humanities Computing. Selected Papers from the ALLC/ACH Conference, Christ Church, Oxford, April 1992* 4, pp. 263–280.

Rizzi, Romeo

2001 "Complexity of Context-free Grammars with Exceptions and the inadequacy of grammars as models for XML and SGML". In: *Markup Languages – Theory & Practice* 3.1, pp. 107–116.

Robertson, Eddie

2002 *Combining Schematron with other XML Schema languages*. URL: http://www.topologi.com/public/Schtrn_XSD/Paper.html. 2012-02-04.

Rubinsky, Yuri and Murray Maloney

1997 *SGML on the Web: Small Steps Beyond HTML*. Charles F. Goldfarb Series On Open Information Management. Upper Saddle River, New Jersey: Prentice Hall.

Sperberg-McQueen, C. M.

1992 *Poor-Folks SGML (PSGML). A Rational Subset of SGML*. URL: http://www.tei-c.org/Vault/ED/edw36.gml. 2012-02-04.

2002 "What matters?" In: *Proceedings of Extreme Markup Languages*. Montréal, Québec.

2003 "Logic grammars and XML Schema". In: *Proceedings of Extreme Markup Languages*. Montréal, Québec.

Sperberg-McQueen, C. M. and Tim Bray

1997 "Extensible Markup Language (XML)". In: *Proceedings of ACH-ALLC '97. Joint International Conference of the Association for Computers and the Humanities (ACH) and the Association for Literary & Linguistic Computing (ALLC)*. Queen's University, Kingston, Ontario.

Sperberg-McQueen, C. M. and Lou Burnard (eds.)
1993 *TEI P2: Guidelines for the Encoding and Interchange of Machine Readable Texts*. Oxford et al.: published for the TEI Consortium by Humanities Computing Unit, University of Oxford.
2003 "Logic grammars and XML Schema". In: *Proceedings of Extreme Markup Languages*. Montréal, Québec.
Sperberg-McQueen, C. M. and Claus Huitfeldt
2004 "GODDAG: A Data Structure for Overlapping Hierarchies". In: *Proceedings of the 5th International Workshop on the Principles of Digital Document Processing (PODDP 2000)*. Ed. by Peter King and Ethan V. Munson. Lecture Notes in Computer Science 2023. Springer, pp. 139–160.
Sperberg-McQueen, C. M., Claus Huitfeldt, and Allen H. Renear
2000 "Meaning and interpretation of markup". In: *Markup Languages – Theory & Practice* 2.3, pp. 215–234.
Stührenberg, Maik and Daniela Goecke
2008 "SGF – An integrated model for multiple annotations and its application in a linguistic domain". In: *Proceedings of Balisage: The Markup Conference*. Vol. 1. Balisage Series on Markup Technologies. Montréal, Québec.
Stührenberg, Maik and Daniel Jettka
2009 "A toolkit for multi-dimensional markup: The development of SGF to XStandoff". In: *Proceedings of Balisage: The Markup Conference*. Vol. 3. Balisage Series on Markup Technologies. Montréal, Québec.
Stührenberg, Maik and Christian Wurm
2010 "Refining the Taxonomy of XML Schema Languages. A new Approach for Categorizing XML Schema Languages in Terms of Processing Complexity". In: *Proceedings of Balisage: The Markup Conference*. Vol. 5. Balisage Series on Markup Technologies. Montréal, Québec.
Tennison, Jeni
2002 "Layered Markup and Annotation Language (LMNL)". In: *Proceedings of Extreme Markup Languages*. Montréal, Québec.
Thompson, Henry S. and David McKelvie
1997 "Hyperlink semantics for standoff markup of read-only documents". In: *Proceedings of SGML Europe '97: The next decade – Pushing the Envelope*. Barcelona, pp. 227–229.
Thompson, Henry S., David Beech, Murray Maloney, and Noah Mendelsohn
2001 *XML Schema Part 1: Structures.* W3C Recommendation 2 May 2001. World Wide Web Consortium (W3C). URL: http://www.w3.org/TR/2001/REC-xmlschema-1-20010502/. 2012-02-04.
2004 *XML Schema Part 1: Structures Second Edition.* W3C Recommendation 28 October 2004. World Wide Web Consortium (W3C). URL: http://www.w3.org/TR/2004/REC-xmlschema-1-20041028/. 2012-02-04.
Van der Vlist, Eric
2001 *Comparing XML Schema Languages*. URL: http : / / www . xml . com / pub / a / 2001 / 12 / 12 / schemacompare . html. 2012-02-04.
2003a *Examplotron*. URL: http://examplotron.org/0/7/. 2012-02-04.
2003b *XML Schema*. Köln: O'Reilly.

2004 "ISO DSDL Overview". In: *Proceedings of XML Europe*. Amsterdam.
2007 "XML Schema Languages". In: *Advanced XML Applications*. Boston, Massachusetts: Thomson Course Technology, pp. 37–90.

Walmsley, Priscilla
2002 *Definitive XML Schema*. The Charles F. Goldfarb Definite XML Series. Prentice Hall PTR.

Walsh, Norman
1999 *Understanding XML Schemas.* URL: http://www.xml.com/pub/a/1999/07/schemas/index.html. 2012-02-04.

Walsh, Norman, Alex Milowski and Henry S. Thompson
2010 *XProc: An XML Pipeline Language.* W3C Recommendation 11 May 2010. World Wide Web Consortium (W3C). URL: http://www.w3.org/TR/2010/REC-xproc-20100511/. 2012-02-04.

Walsh, Norman and Leonard Muellner
1999 *DocBook: The Definitive Guide*. O'Reilly.

Witt, Andreas
2004 "Multiple Hierarchies: New Aspects of an Old Solution". In: *Proceedings of Extreme Markup Languages*. Montréal, Québec.

4. Controlled Language Structures in Technical Communication

Thorsten Trippel

1. Introduction

In many environments language is used in situations where reoccurring communicative events have the consequence that – to avoid miscommunication – some standard language is being used. For example the communication of air traffic controllers and aircraft pilots are such occasions, where miscommunication is fatal and is avoided by formulaic, standardized language (see also Ladkin, 2012, in this volume). However, in expert communication the use of formulaic, standardized language is not an exception and not restricted to cases in which miscommunication is live endangering. Sometimes it is a matter of efficiency, avoiding inaccuracies and saving time.

When talking about controlled language, we refer to language and language structures that may be part of the natural language – that is the language used in informal conversations – but it is required that there is some kind of authority licensing the use of this language. Figure 1 illustrates the degrees of freedom vs. the amount of control of different kinds of languages in terms of syntactic (grammatical constructs) and lexical (words with unique meanings) features. Computer or programming languages are the most restricted languages in terms of lexicon and permissible syntax, air traffic control language is an example for human language on the controlled language side, while technical writing is on the natural language side, though partly controlled, while conversation is seen as the most variable in terms of syntax and lexicon. Folksonomies in this illustration stand for hierarchical lexical resources which are rather fixed but where the words do not have major syntactic restrictions. Metadata is included, being usually more restricted lexically (see section 2 for more details on metadata) and following stricter syntactic patterns.

In the case of air safety communication, the restrictions are a standard set of communication patterns issued internationally by appropriate authorities, here this is the *International Civic Aviation Organization* (ICAO, see for example the ICAO DOC 9432, 2007) or national mirroring organizations (for example the UK Civil Aviation Authority in the United Kingdom, see CAP 413 (2009)) with standard phraseology for radio communication of aircrafts and air traffic controllers.

In the industry, there might be other guidelines decided on by ‘the boss’ or a company’s communication department. The most usual case of control is the

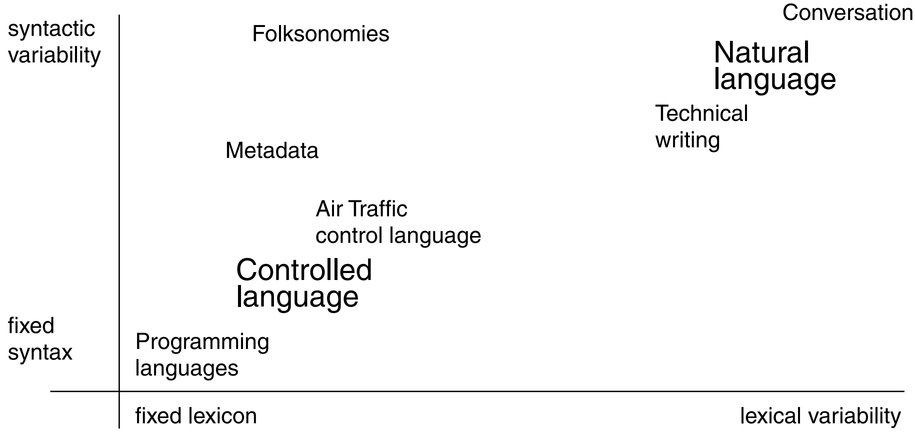


Figure 1. Illustration of the degrees of variability of between Natural languages and Controlled languages.

use of a controlled vocabulary (see also Chu, 2003, chapter 4) – often highly technical terms – and the request to restrict oneself to 'simple' structures, for example, to avoid passives, subordinate clauses, subjunctives, negation, double negation, etc.

In technical environments where computers and computer services communicate with each other there is usually a fixed grammar for the communication between the systems and a restricted vocabulary. When saying grammar, this does not refer to school grammars, but formal grammars, that is, a list of rules saying which 'words' or classes of words or which structures can occur in sequence of each other (see also Stührenberg, 2012, in this volume). In technical communication of computer systems the complete expressivity of language is consciously avoided, but the allowed structures are restricted to the concrete communicative purposes of the system. The same is true for the vocabulary: for large parts of such communicative instances the vocabulary is completely fixed, no creativity allowed and synonyms and ambiguities are disallowed.

The precision and unambiguous use of expressions is typical for technical communication and distinguishes technical communication form natural communication, which allows for a creative use of words, and which has different ways for expressing the same thing. The creativity is restricted by grammars which scholars try to describe in syntactic theories, with words lexicographers try to record in dictionaries with possibly all of their *meanings*[1].

Technical vocabularies are in fact studied and recorded by terminologists, who do not refer to *words* and *meanings*, but *terms* and *concepts* (for an introduction to terminology see for example Sager, 1990; Wüster, 1991) (see also Sharoff and Hartley, 2012, in this volume). The concept refers to the one-and-only abstract meaning, all signs – such as written words or painted symbols –

represent this concept. In technical communication by experts, there may be more than one term per concept, in electronic applications there is usually only one [2].

When talking about *technical communication* in this sense, it refers to the communication with a specific purpose in a technical context, for example, by experts or according to specific protocols. In this article we exclude protocols for networking operations such as HTTP, TCP/IP, etc. (for protocols see also Diewald, 2012, in this volume) but restrict ourselves to formally restricted patterns that are intended for human perception.

1.1. Simplified languages for internationalization purposes

A major part for the simplification of languages is the reduction of the grammatical complexity. Though simplification is well known from the development of trade languages and pidgins, in internationalization contexts this is also used. Basically there are two simplification strategies, reducing the number of words and reducing the allowed structures. While the earlier is usually achieved by defining a core vocabulary, which could at least in principle be checked automatically, the reduction of structures is harder, because it requires a syntactic awareness.

In learner dictionaries of a language, for example, in Crowther (1995), definitions use a restricted vocabulary to allow learners of a language to understand the definitions with a limited number of words. The *Aerospace and Defence Industries Association of Europe* (ASD[3]) has published, for example, a specification of a restricted language called *Simplified Technical English* for the aircraft industries, see ASD-STE100. This specification allows for the reduction of the vocabulary for technical maintenance of aircrafts to assist in the international contexts.

By leaving out function words in restricted vocabularies the step towards structural reduction is also small; avoiding constructions is easier if the required function words are not included. One step further into controlled structures is of course the phraseology defined for air traffic, where a limited number of utterances is defined with slots for local and situational adjustment, for example, to communicate which way an aircraft has to taxi and where it has to stop and wait.

1.2. Fixed terminology in technical text

Terminology in technical texts is extensively studied for technical writers and for industrial translation (see for example Sager, 1990, for an introduction) (see also Max 2012 in this volume). Besides a communicative function, terminology is also used for branding purposes in the industry, that is to distinguish texts of one company from that of others. As technical texts, such as user manuals frequently have to be translated into different languages, involving various translators, a standardization within a project or for a company is seen both, as a branding activity and a strive for consistency (see also Sasaki and Lommel, 2012, in this volume).

A simple example of the need for consistency is a technical manual consisting of an illustration with the named parts (for example, the user manual of a household device such as a microwave oven) and a text describing the use of the different parts and switches. If the naming is inconsistent, references from text to illustration are bound to fail. If, as is common practice, the translation of such a manual is done by various translators, the likelihood for inconsistent translation increases. To avoid inconsistencies, technical applications – terminology management systems such as the industry's SDL-Trados Multiterm[4], Star's WebTerm[5] or ProTerm[6] – are used to support translators in the translation process. For example, when a translation memory system such as SDL-Trados Translator's workbench, Wordfast[7], etc. is being used, terms available in a specialized lexicon called *termbase* are highlighted based on tokenization of the source and the contents of the terminology management system.

If a term is identified – which also may include some morphological processing to avoid the use of all grammatical forms of the term and include only one form such as a nominative singular for nouns – the translator is presented with a highlighted hint pointing to the concept in the termbase, possibly with the negotiated term in the target language. If various terms in the target language exist, in industrial projects there are preferred terms manually selected by a project leader to be used and deprecated terms that are not to be used at all, for example, the competitors' brand names.

Within industrial writing the terms are specified in styleguides or processed in *terminology management systems* (TMS). Contrary to dictionaries, the TMS structures the entries not by a lexical form – or a lemma – but by concept. Hence each article in the TMS represents an abstract concept, usually with a definition, and references to related concepts, the subject field and maybe to a hierarchy of concepts, an *ontology* (see also Sharoff and Hartley, 2012, in this volume) or *concept system*. Within the article of the concept all instantiations or terms for the included languages are embedded, possibly with separate definitions per language. This does not imply that there are different definitions or that the concept is expressed by terms in different languages which can have different scopes and restrictions. In fact this means that a concept can have more than one definition, because the definition is only an approximation to the concept and not the concept itself, which can be more abstract. For this reason definitions can exist for the same concept that seem to be deviating from each other.

The industrial and academic use of this kind of vocabulary resulted in early standardization efforts, for example with the ISO committee for Terminology (ISO TC 37) being established as early as in year 1947[8]. For the translation industry it became necessary to define exchange formats for terminological databases, resulting in a family of ISO standards, for example, ISO 12200:1999 (1999) (the *Machine Readable Terminology Interchange Format*, MARTIF), later the *Terminology Markup Framework* (TMF, ISO 16642:2003) with the *TermBase eXchange* format (TBX, ISO 30042:2008). MARTIF defined an SGML syntax (see Stührenberg, 2012, in this volume) for negotiated interchange, that is, the interchange partners had to exchange knowledge about the data categories used in the applications used for the interchange. Later developments with TMF and TBX basically use a two level system: First the data categories and their hierarchy is defined with reference to standard data categories. Based on the data category definition, the interchange between different systems is defined. This interchange then does not dependent on the names used for the data categories in a specific application because the semantics of the names is predefined.

2. Controlled language for semantic applications

Controlled vocabularies for semantic applications behave similarly to terminology (comp. Chu, 2003, chapter 4.2): They are organized in thematic fields or domains, often also organized hierarchically in thesauri. Controlled languages also show further restrictions in linguistic structures, such as syntax and morphology (see also Wyner et al., 2010). By semantic applications here we understand a set of computer programmes that have routines interpreting the content of an input and changing the behavior based on the content. This means that a semantic application analyses the input structures and content, and if the structure and content contain a key, the application adapts to the key. This corresponds to applications using the *Semantic Web* as explained by Berners-Lee et al. (2001). Very simple forms are for example, the recognition of links in text, for example, some text processors use character sequences beginning with http:// and render and interpret them as links. This can either be accomplished by identifying a key expression or by analyzing the structure of an expression, recognizing the internal structures and substructures of URLs. Both correspond to a controlled vocabulary and a restricted syntax. More sophisticated is the use of highly structured information, often referred to as *metadata*.

Metadata is structured data for describing and finding something[9]. Typical examples are highly structured formulaic texts such as descriptions of experiments in the sciences or bibliographical references. In bibliographical references, for instance, books have authors and titles, while journal articles show other structures, because they also need to refer to the journal title and possibly the editor. With the structure often also some restrictions of length or other formal properties are imposed.

Electronically available texts and material to be used by semantic applications can also impose structures, that are otherwise only latently included, such as the data categories mentioned in the citations. Though it may not be required for human users to mark a name as the name of the author, for semantic applications it may be relevant.

Based on this distinction different approaches for semantic applications can be used. A controlled vocabulary for content description without structures is suitable for statistical language processing such as text classification and indexing using machine learning techniques. These require a sufficiently large size of data and the structures are only statistically relevant.

Controlled structures on the other side are more independent of statistics. Of course the development of structures is an overhead that is appropriate only if the number of instances is large enough but the classification and description is readily available from the start. Classification in terms of hierarchies and structured description by metadata are used in semantic applications (see for example Breitman et al., 2007, pp. 7–11, for a brief introduction). For the description of metadata various schemas exist. In semantic applications these schemas are usually interpreted by appropriate programmes, not by humans. Hence, the syntax is often unrelated to human language. Instead it is designed in a way that it can be easily checked for consistency. Checking the consistency includes a test if the controlled vocabulary was used.

2.1. Descriptions of documents of resources by metadata

One specific type of controlled language can be seen in the descriptions of documents and other classes of resource. For this purpose, *metadata* is being used. Metadata consists of structured descriptive information that is primarily not relevant for the described resource. For example, the bibliographical information in a library catalogue for a book is not essential for the book, but for locating the book in a large library it is essential to have the location or stock number in combination with author and title information.

For libraries and archives, the *Dublin Core Metdata initiative* (DCMI[10])

introduced 13, later 15 descriptive categories for written material (in the meantime a new approach is using a more modular concept, compare Hillmann, 2005; Coyle and Baker, 2009), among them information on *title*, *author*, *intellectual property rights*, *year of publication*, etc. These are usually referred to as the *Dublin Core* (DC). As these are partly unspecific, there were options introduced to further specify these categories. In the meantime, with the introduction of an abstract model, the fixed number of elements has been abolished.

For language resource one of the first extensions of DC for language resources was introduced by the *Open Language Archive Community* (OLAC[11], see Simons and Bird, 2008), which especially referred to linguistic fields, participants in a dialogue recorded, various languages etc. However, this standard was not sufficient for many resources to extensively describe them. For spoken language data the *ISLE MetaData Initiative* (IMDI; see IMDI, 2001) introduced a hierarchy of data categories to describe everything relevant for recorded dialogues, including information on the speakers with their relevant background, recording environments, used technology, etc. The use of IMDI for spoken language recordings points to a problematic issue: different kinds of resources require different data categories, if the metadata is not only used for filing resources in an archive, but also to support content based searches in a catalogue.

Metadata can hence be used from various perspectives:

– locating a known resource in an archive or repository: this was the intention of the DC metadata;
– understanding a resource, without analysing or reviewing it: this can partly be achieved by structured metadata as IMDI, where, for example, audio recordings can be reviewed based on properties of the speakers without listening to the recording;
– referencing to a resource: by using identificatory properties, it is possible to refer to a resource. An example are references in a book pointing to a list of works cited which themselves should contain enough identifying properties to locate the resource in an archive or repository.

A technology allowing for each perspective simultaneously was introduced with the *Component MetaData Infrastructure* (CMDI, Broeder et al., 2010), originally developed for electronic repositories for linguistic resources.

– The reference to resources in a repository is given by using persistent and unique identifiers. Hence citations can use these as references to the resource.
– By the resource dependent flexible creation of hierarchical metadata schemas, all relevant information describing a resource can be included in the metadata.
– By a mapping of data categories to data categories used in the repository catalogues, it is possible to use this type of metadata also for catalogues.

The idea of CMDI is comparatively simple: From the set of all possible data categories the relevant ones are selected and hierarchically organized or grouped. Groups are called components, which can be reused later and combined with other data categories and components. Components that are not part of other components and are used to extensively describe one particular type of resource are called *profile* in this infrastructure. At the end, for each class of resources there is a defined profile. The component metadata infrastructure provides a component registry, that is a web based editor for creating components and profiles, storing the profiles and components in a central infrastructure for easy

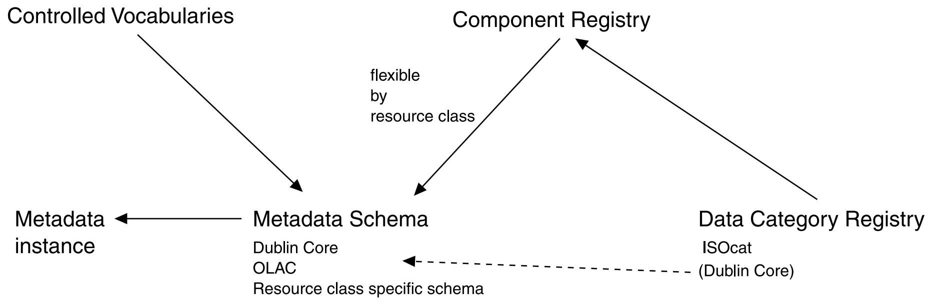


Figure 2. Integrated system for metadata: data categories from a registries are used to create a metadata schema implicitly or by the component registry, which may also include a controlled vocabulary for the metadata instance.

reuse also by others. For this purpose each component receives a persistent identifier, that points solely to this profile.[12]

Figure 2 shows the connection between the *metadata instance*, for example a library catalogue record, the metadata schema and a data category registry (see section 2.3). Dublin Core is included not only as a metadata schema, but also as a registry, because the data categories for that schema are explicitly defined by the providers.

For the distribution of CMDI-Metadata the *Open Archive Protocol for Metadata Harvesting* (OAI-PMH; see OAI-PMH, 2008) is being used. This protocol, building on the standard protocol for websites, that is, HTTP, allows for providing metadata over a central infrastructure. It is frequently used by libraries and archives to exchange their metadata[13].

2.2. Classification systems for documents and resources

Classification systems are systems to organize items into classes, that is grouping them hierarchically. In contrast to ontologies that are – at least originally – targeted at providing a system for all the worlds knowledge (See for example Breitman et al., 2007, chapter 2), classification systems are more restricted, for example, to the products in a shop, goods in a warehouse, etc. (see for example ISO 22274:2011 for a discussion). For example goods can be classified by size, color, material, etc. The goal of such a system is not to be sound, i.e., that the degree of similarity within each class is the same and each item is placed in the best possible class and that there is only one correct classification system, but to create classes that increase the efficiency in sorting and organizing.

Documents and resources, such as data in a given language, pose additional challenges to a classification system. One problem is that a classification system may be dependent on culture and language (see also Sasaki and Lommel, 2012, in this volume). To approach the problem of classification systems in internationalization contexts, ISO 22274:2011 provides the principles required. It bases classification systems on the grounds of concept systems as used in terminology. In this way, it is not the document, but a concept that is placed into the classification system. However, as expressed by the standard, the purpose of a classification system is to allow for sorting of material or immaterial items, not necessarily to have a simplified model about concepts and their interrelation in a hierarchical form.

For a classification it is required that the classes are clearly separated from each other and that it is unambiguous (ISO 22274:2011 provides the background). That means that when an item is classified, there is exactly one class. Though new developments may require the creation of new classes, i.e. the classification system is extensible, it is essential that each item is part of a class.

In consequence, the metadata categories allow for grouping and classification of documents and other resources. For the systematic classification by metadata, the categories as such need to be defined and listed.

2.3. Registries for data categories

A central issue for framework based approaches and classification is a data category registry or a list of data categories with some form of definition. In a framework based approach as for TMF and TBX there are usually two components: the selection of data categories from a set of data-categories with the organisation of these data categories into hierarchies and a specific syntax[14].

In the case of TBX, the specification is using the *eXtensible Constraint Specification* (XCS), which is a special XML syntax (see Stührenberg, 2012, in this volume), providing additional constraints defined by the schema for the TBX instance. A data category registry provides the set of data categories.

Similarly to termbases, a data category registry has a structure in which the data category is an abstract concept rather than a name or a definition that has a more or less arbitrary identifier (For a more detailed discussion, see Kemps-Snijders et al., 2008). The concept of a data category may have various definitions and names but the reference to the data category as expressed by the identifier uniquely refers to the data category.

The data category repository for TBX, which is also used for data categories for language resources, is the ISO Data Category Repository *ISOcat*, see ISO 12620:2009 and www.isocat.org. As the successor of earlier lists of data categories, ISOcat is an open repository, which means that new categories can easily be added. For standardization purposes, categories that are supposed to be reused by other parties are reviewed regularly by an expert group, if they are accepted, they are made part of a standardized set of data categories. In the case of redundancy or unclear definitions, the data categories remain in the realm of the creator, though they may be public. Each data category has a unique identifier, which can be used for addressing this data category.

Metadata resources point to data category registries and to data category profiles. At the same time metadata also refers to primary data. As these resources may be distributed, identification is crucial. For the identification, an international standard was established with ISO 24619:2011, which provides the principles for the identification of language resources.

In the case of data categories, identifiers are used as references to the data category repository. Appropriate applications resolve the identifier automatically, for example, using the *handle system* (see http://www.handle.net). Handles allow as other *Persistent Identifiers* (PID, see Schroeder, 2010) such as *Uniform Resource Names* (URN, see for example Schöning-Walter, 2010) and *Digital Object Identifier* (DOI, see for example Brase, 2010) to uniquely reference objects persistently. In this case, data categories are defined in a data category repository and each data category receives such a persistent identifier, which is used to address its definition independent of physical location or concrete web address. This is part of the semantic application in the sense that the meaning of the data category is not defined by the document or the underlying grammar, but stored externally. The definitions of ISOcat, though created by experts, are created for human readability, not for a specific semantic framework or formalism.

2.4. Controlled vocabularies

Terminology is supposed to control the vocabulary of translators and technical authors, a standard phraseology as for air traffic communication also restricts the vocabulary used significantly. But these are only policies and guidelines that are used for training and assessing. Technically, the correct use of terminology is not enforceable in oral communication or in translation contexts, not counting possible liabilities or sanctioning by a supervisor. In formulaic contexts, it is possible to use other mechanisms to enforce the use of specific wording or patterns. For example, it is possible to restrict the input of a form to a predefined set of words, the controlled vocabulary. This controlled vocabularies is not only defined externally for example in a data category repository but also formally within a document grammar. In XML, for example, the schema for defining the syntax allows for providing a fixed set of values for elements or a specific type of contents, see Thompson et al. (2004) and Biron and Malhotra (2004). Types, for example, can be dates, which have to follow a specific pattern to be allowed within a specific location of a text.

It is possible to control the vocabulary if in a given communication instance only – and without exception – a limited number of vocabulary or specific patterns can be used, but nothing else. On webpages, the concept of controlled vocabularies for filling in forms is often realized by pick-lists, without an option of inserting free text. Though this may be possible for addresses, picking from a list of countries, for example, this may be a bit more problematic in cases where the list is not completely closed or if it is too large. For example, using the address example again, the list of city names in a country could be seen as fixed, but seen worldwide, new cities appear or merge, and providing an extensive list might be problematic because of the size of the list. In this case a controlled vocabulary does not help.

3. Searching with controlled languages

Searching for documents and resources is ubiquitous with network availability virtually everywhere and the use of search engines for locating resources. Google, Bing and similar services provide extensive searches using a full text index, basically creating a wordlist with references to the pages on which a word appears (see also Paaß, 2012, in this volume). Additionally, they rank pages according to some company secrets, and they leave out high frequency words from indexing, such as function words in a natural language. One component frequently used for ranking is the position of the word on a page, for example, words in titles and headings are interpreted as being more important than words towards the end of a page. This results in the whole industry of search engine optimization with the goal of receiving the highest possible page rank[15].

Full text search with a combined ranking mechanism show a trade-off between search precision and recall. Precision and recall are statistical values measuring the amount of correctly found documents (see also Menke, 2012; Paaß, 2012, both in this volume). By recall the number of correctly found documents is compared to the number of documents that would have been correct, while precision evaluates the number of correctly found documents in relation to the whole number of retrieved documents. Problematic cases are all cases of ambiguity, i.e., polysemous expressions or homonyms, and morphologically complex expressions. To avoid that, controlled vocabularies and languages provide a way out, especially when using a fixed structure.

3.1. Metadata in search engines

There are three classes of search engines distinguished here:

- full text index based search engines such as Google, Bing and Yahoo,
- full text search engines based on metadata providing semantic indexing, and
- structured semantic search engines, for example by faceted searches.

The HTML-standard already contains a way of including metadata such as keywords and author information in *meta*-tags in the header of websites. This metadata is indexed by search engines such as *Google*, *Bing*, and *Yahoo*, and can be used for the ranking of search results. More recently, the named search engines were part of an initiative to include ontological information in websites. For semantic information, webpages may include references to an extensible concept system, the details can be found at http://www.schema.org. It can be expected that these search engine providers for webpages will support more specific searches based on these semantic tags in the future.

More widely used currently using fixed structures for searches can be found in library catalogues, often called expanded or advanced search [16].

For this kind of search, often the user sees a table with fields to fill in, the fields are labeled with *title*, *author*, etc., which of course are the metadata categories. By inserting all bits of information into the fields, the search results are more precise by not lowering the recall. For example, somebody searching for all works by William Shakespeare will easily find them when entering the name in the author-field, while a full text search would also find all texts about Shakespeare. This search method usually does not involve any normalization, including search for orthographic variants; hence this search method is not only relying on the structure of the bibliographical data, but also the quality. Full text search engines tend to have heuristics to handle such inconsistencies, because they expect a lower input data quality.

Other uses of controlled vocabularies in search are becoming more and more popular in business applications, such as webshops. Webshops such as Amazon.com classify their products and use keywords for describing products. Hence for customers it becomes possible to search either by the product class, using the hierarchical classification system, or by the keywords. Keywords can either follow a closed vocabulary, or be assigned by non-experts, in which case they are creating a folksonomy, i.e., a collection of associated words or tags which create some form of a hierarchy[17]. A study on classification of keywords can be found in Heckner et al. (2008). If the keywords are used with a classification system, the result is often a search interface called *faceted browser* (Hearst, 2006; Sacco and Tzitzikas, 2009, provide the theoretical background to highly structured search interfaces).

3.2. Faceted browsing

In a faceted browser, instead of inserting a text, the search process operates on lists of values presented to the user. One example is the Flamenco system (see for example Hearst, 2006) using the classes of metadata categories. In a faceted system, documents are classified according to their properties in a hierarchy and a user navigates through this hierarchy – usually by selecting one option upon which other, still existing options can be selected. That also means that options, that do no longer exist are not available any more for further selection. For example if a classification system distinguishes animate and inanimate, after selecting the inanimate-option there is no way of selecting human. Usually a *numerical volume indicator* (NVI) provides feedback to the user on the size of the search space after selecting the options.

Barkey et al. (2011) present a faceted browser for linguistic resources that uses a similar model as Flamenco. However, they use an adaptive system, that is: depending on the previously selected data categories not only the NVI for the remaining options is adjusted, but partly new data categories appear that are specialized to classify the remaining options. For example, after selecting the resource type *tool* for software tool, input and output formats appear as a pick-list, while for *corpus* a newly occurring option is *tag set*. These newly occurring facets only make sense for these subclasses; in general, they would be more confusing then helpful in the search process.

The options in a faceted browser can be both, free text or a controlled vocabulary. However, the classification works best if the number of classes is sufficiently large, but general enough not to become too complex. Hence it is practical to try to stick to a controlled vocabulary as often as possible. For language resources, this is sometimes a bit tricky, as various users might use different classification systems. An example for this is the class of *genre*, in which the classification is controversial and not unambiguous. In the component metadata infrastructure, as in XML Schema languages in general, the definition of a closed vocabulary hence solves the problem slightly, and as for classification systems a unique classification receives precedence before theoretical soundness.

4. Summary and Outlook

This article discussed aspects of controlled language structures in technical communication. Technical communication is seen here both in technology mediated communication (applications) and in specific technical situations where two or more persons communicate directly with each other, such as in conversations for air traffic control. This technical communication is characterized by restrictions in lexical use, i.e. specific words to be used with strictly defined semantics, and a restricted syntax. More flexibility can be found in technical writing and in the translation of technical texts, such as manuals, prephrased texts and handbooks. This communication shows a high degree of fixed terminology, though otherwise it is as unrestricted as other natural language. For these kinds of texts authors and translators are assisted by specialized applications for terminology management.

Controlled language structures are also important for semantic applications, where terminology and metadata are used, especially for applications that can be used in metadata descriptions for items in a catalogue and in search applications. Metadata depends on the type of resource it is used for, resulting in a classification of resources as well. The component metadata infrastructure latently builds profiles for metadata descriptions on top of a resource classification in the field of language resources. Such a system can be adjusted to other types of resources as well.

Controlled language structures will gain more importance in the future, for web applications, such as search engines based on metadata and incorporating concept systems. This is due to the fact that the amount of searchable data increased dramatically in the past and will continue to do so, resulting in unusable search results from full text index based search engines. Google, Bing and Yahoo point in that direction by working on a common concept system, which will have a great impact for professional information providers whose business model is based on being found through searches.

Additionally controlled language structures will gain even more prominence in technical writing and translation. It is to be expected that with the needs for further internationalization of documents, machine translation will become more prominent. And as some problems of machine translation can be avoided by restricting the lexicon and syntax of the source language text, it can be expected that especially multi-national companies will provide policies to avoid structures and wording that cannot be handled by their preferred MT system. This will have an impact on technical writing and translation.

The field of controlled language for specific purposes on the other hand will remain a niche for specific, critical communication. For example general simplified language to be used for the European Union seems not to be on the (political) agenda, though specialized simplified languages are used and not to be abolished for security critical fields such as air traffic.

Controlled language structures will be part of most technical communication in one way or the other and concept-systems, metadata and searches will play an integral part of that. This article gave an introduction and some background for all of these areas.

Notes

1. *Meaning* is one of these expressions we continuously refer to but where we have only a limited way of defining it. We use it in a natural way here and try to avoid going into formal semantics or the philosophy of meaning. It is accepted that it is not entirely clear that a word – another intuitively yet problematic expression – having more than one definition in a dictionary really refers to different meanings(see also Sharoff and Hartley, 2012, in this volume).
2. There is no principle reason for this, but the whole disambiguation problem is avoided if the meaning of a term is unique.
3. http://www.asd-europe.org/site/
4. Available via the company website http://www.trados.com/.
5. A previous terminology management system by Star was TermStar, more details on WebTerm is available at the companies website http://www.star-group.net.
6. See http://www.documatrix.com
7. See http://www.wordfast.net for information on the Wordfast translation memory system.
8. The current title of this ISO technical committee is *Terminology and other language and content resources*, with an original scope to standardize terminology for technical writers and translators, see http://www.iso.org/iso/standards_development/technical_committees/other_bodies/iso_technical_committee.htm?commid=48104
9. We adopt here a definition used by librarians, see for example http://www2.sub.uni-goettingen.de/intrometa.html, rather than the omnipresent recursive definition, see for example Breitman et al. (2007), p. 7.
10. http://dublincore.org/
11. http://www.language-archives.org/
12. For more information on the component registry see http://catalog.clarin.eu/ds/ComponentRegistry/, visited in October 2011).
13. As the registration as a data provider is not required, no full list exists. A possibly incomplete list of OAI-PMH-data providers can be found at http://www.openarchives.org/Register/BrowseSites.
14. Another example from standardisation here is the Lexical Markup Framework (ISO 24613:2008), others are still under development by ISO TC 37.
15. There is plenty of literature on Search Engine Optimization, for example, Alby and Karzauninkat (2007) provides a good motivation in German. For a motivation in English, see Enge et al. (2009), also providing some statistics on search engine use.
16. Morville and Rosenfeld (2007, p. 185,) already point out that advanced search, though expected, is less frequently used. If this is due to inappropriate interfaces or missing applications, is not discussed.
17. In his paper on data mining on folksonomies, Hotho (2010) also provides a formal definition of a folksonomy and explores hierarchies involved.

References

Alby, T. and S. Karzauninkat
2007 *Suchmaschinenoptimierung: Professionelles Website-Marketing fur besseres Ranking*. Carl Hanser Verlag, Munchen.

ASD-STE100
2010 Simplified technical english: International specification for the preparation of maintenance documentation in a controlled language. Specification Issue 5, AeroSpace and Defence Industries Association of Europe, ASD-STAN – Standardization, Avenue de Tervuren 270, B-1150 Brussels, Belgium. Specification ASD-STE100.

Barkey, R., E. Hinrichs, C. Hoppermann, T. Trippel and C. Zinn
2011 Trailblaizing through forests of resources in linguistics. In *Proceedings of Digital Humanties (DH), June 19–22, Stanford: Stanford University*, Stanford. Stanford University.

Berners-Lee, T., J. Hendler and O. Lassila
2001 The Semantic Web a new form of web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American*.

Biron, P. and A. Malhotra
2004 XML Schema Part 2: Datatypes. Second edition. http://www.w3.org/TR/xmlschema-2/ [Last accessed 03/29/2012].

Brase, J.
2010 9.4.2 der digital objekt identifier (doi). In H. Neuroth, A. Oßwald, R. Scheffel, S. Strathmann, and K. Huth (eds), *nestor Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung*. Kompetenznetzwerk Langzeitarchivierung, urn:nbn:de:0008-20100305186.

Breitman, K., M. Casanova and W. Truszkowski
2007 *Semantic Web: Concepts, Technologies and Applications*. NASA Monographs in Systems and Software Engineering. Springer, London.

Broeder, D., M. Kemps-Snijders, D. Van Uytvanck, M. Windhouwer, P. Withers, P. Wittenburg and C. Zinn
2010 A data category registry- and component-based metadata framework. In *Proceedings of the 7th Conference on International Language Resources and Evaluation (LREC 2010)*, Malta.

CAP 413
2009 Radiotelephony manual. Technical Report 413, UK Civil Aviation Authority, London.

Chu, H.
2003 *Information Representation and Retrieval in the digital age*. ASIST monograph series. Information Today, Inc., Medford, New Jersey, USA.

Coyle, K. and T. Baker
2009 Guidelines for Dublin Core Application Profiles. Dublin Core Metadata Initiative. Technical report, Dublin Core Metatdata Initiative. http://dublincore.org/documents/2009/05/18/profile-guidelines/.

Crowther, J. (ed.)
1995 *Oxford Advanced Learner's Dictionary of Current English*. Oxford University Press, Oxford, fifth edition edition.

Diewald, Nils
2012 Decentralized social network sites. In Alexander Mehler, Laurent Romary, and Dafydd Gibbon, editors, *Handbook of Technical Communication*. De Gruyter, Berlin/Boston.

Enge, E., S. Spencer, R. Fishkin and J. Stricchiola
2009 *The Art of SEO*. O'Reilly Media, Inc., Sebastopol, CA.

Hearst, M.
2006 Design recommendations for hierarchical faceted search interfaces. In *Proceedings of the SIGIR 2006 Workshop on Faceted Search*, pages 26–30.

Heckner, M., S. Mühlbacher and C. Wolff
2008 Tagging tagging. Analysing user keywords in scientific bibliography management systems. *Journal of Digital Information*, 9(2):1-17.

Hillmann, D.
2005 Using Dublin Core – the elements. Technical report, Dublin Core Metadata Initiative. http://dublincore.org/documents/2005/11/07/usageguide/elements.shtml. [Last accessed 03/29/2012]

Hotho, A.
2010 Data mining on folksonomies. In Giuliano Armano, Marco de Gemmis, Giovanni Semeraro, and Eloisa Vargiu, editors, *Intelligent Information Access*, volume 301 of *Studies in Computational Intelligence*, pages 57–82. Springer, Berlin / Heidelberg.

ICAO DOC 9432
2007 Manual of radiotelephony. Technical Report 9432, International Civil Aviation Organization (ICAO). This document is regularly adjusted to professional requirements.

IMDI
2001 Metadata elements for session descriptions, draft proposal version 2.5. http://www.mpi.nl/ISLE/documents/draft/ISLE_MetaData_2.5.pdf. [Last accessed 03/29/2012]

ISO 12200:1999
1999 Computer applications in terminology – machine-readable terminology interchange format (MARTIF) – negotiated interchange. Technical report, ISO.

ISO 12620:2009
2009 Terminology and other language and content resources – specification of data categories and management of a data category registry for language resource. Technical report, ISO.

ISO 16642:2003 and ISO 16642:2003
2003 Computer applications in terminology – terminological markup framework. Technical report, ISO. International Standard.

ISO 22274:2011
2011 Systems to manage terminology, knowledge and content – internationalization and concept-related aspects of classification systems. Technical report, ISO. Draft international standard.

ISO 24613:2008
2008 Language resource management – Lexical Markup Framework (LMF). Technical report, ISO. International Standard.

ISO 24619:2011
2011 Language resource management – persistent identification and sustainable access (PISA). Technical report, ISO.

ISO 30042:2008
2008 Term-Base eXchange (TBX) format specification. Technical report, ISO. International Standard.

Kemps-Snijders, M., M. Windhouwer, P. Wittenburg and S. Wright
2008 ISOcat: Corralling data categories in the wild. In N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odjik, S. Piperidis, and D. Tapias, editors, *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco. European Language Resources Association (ELRA).

Ladkin, Peter
2012 Verbal communication protocols in safety-critical system operations. In Alexander Mehler, Laurent Romary, and Dafydd Gibbon, editors, *Handbook of Technical Communication*. De Gruyter, Berlin/Boston.

Max, Aurélien
2012 Document authoring: from word processing to text generation. In Alexander Mehler, Laurent Romary, and Dafydd Gibbon, editors, *Handbook of Technical Communication*. De Gruyter, Berlin/Boston.

Menke, Peter
2012 Evaluation of technical communication. In Alexander Mehler, Laurent Romary, and Dafydd Gibbon, editors, *Handbook of Technical Communication*. De Gruyter, Berlin/Boston.

Morville, P. and L. Rosenfeld
2007 *Information Architectrue for the World Wide Web*. O'Reilly Media, Inc., Sebastopol, CA.

OAI-PMH
2008 The open archives initiative protocol for metadata harvesting. Technical report, OAI. Protocol Version 2.0 of 2002-06-14, http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm. [Last accessed 03/29/2012]

Paaß, Gerhard
2012 Document classification, information retrieval, text and web mining. In Alexander Mehler, Laurent Romary, and Dafydd Gibbon, editors, *Handbook of Technical Communication*. De Gruyter, Berlin/Boston.

Sacco, G. and Y. Tzitzikas (eds.)
2009 *Dynamic Taxononmies and Faceted Search*. Springer, Berlin, Heidelberg.

Sager, J.
1990 *A practical course in terminology processing*. Benjamins, Amsterdam/Philadelphia.

Sasaki, Felix and Arle Lommel
2012 Multilingual computing. In Alexander Mehler, Laurent Romary, and Dafydd Gibbon, editors, *Handbook of Technical Communication*. De Gruyter, Berlin/Boston.

Schöning-Walter C.
2010 9.4.1 Der Uniform Resource Name (URN). In H. Neuroth, A. Oßwald, R. Scheffel, S. Strathmann, and K. Huth (eds.), *nestor Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung*. Kompetenznetzwerk Langzeitarchivierung, urn:nbn:de:0008-20100305176.

Schroeder, K.
2010 9.4 Persistent Identifier (PI): ein Überblick. In H. Neuroth, A. Oßwald, R. Scheffel, S. Strathmann, and K. Huth (eds.), *nestor Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung*. Kompetenznetzwerk Langzeitarchivierung, urn:nbn:de:0008-20100305164.

Sharoff, Serge and Anthony Hartley
2012 Lexicography, terminology and ontologies. In Alexander Mehler, Laurent Romary, and Dafydd Gibbon, editors, *Handbook of Technical Communication*. De Gruyter, Berlin/Boston.

Simons, G. and S. Bird
2008 OLAC metadata. Technical report, Open Language Archive Community. http://www.language-archives.org/OLAC/metadata-20080531.html. [Last accessed 03/29/2012]

Stührenberg, Maik
2012 Foundations of markup languages. In Alexander Mehler, Laurent Romary, and Dafydd Gibbon, editors, *Handbook of Technical Communication*. De Gruyter, Berlin/Boston.

Thompson, H., D. Beech, M. Maloney and N. Mendelsohn
2004 XML Schema Part 1: Structures. Second edition. http://www.w3.org/TR/xmlschema-1/. [Last accessed 03/29/2012]

Wüster, E.
1991 *Einführung in die allgemeine Terminologielehre und terminologische Lexikographie*. Springer, Wien, 3rd edition.

Wyner, A., K. Angelov, G. Barzdins, D. Damljanovic, B. Davis, N. Fuchs, S. Hoefler, K. Jones, K. Kaljurand, T. Kuhn, M. Luts, J. Pool, M. Rosner, R. Schwitter and J. Sowa
2010 On controlled natural languages: Properties and prospects. In N. Fuchs, editor, *Controlled Natural Language: Workshop on Controlled Natural Language, CNL 2009, Marettimo Island, Italy, June 8–10, 2009, Revised Papers*, volume 5972 of *LNCS: LNAI*. Springer, Berlin, Heidelberg, New York.

5. Document Classification, Information Retrieval, Text and Web Mining

Gerhard Paaß

This chapter overviews state of the art technologies of modern information retrieval and data mining. It focuses on quantitative document models and their utilization for various tasks such as topic detection & tracking (TDT), text categorization, topic clustering and topic labeling. Further, it includes web content and structure mining as well as the rising area of ontology mining. The chapter puts emphasis on standard technologies in the present area and their application to technical communication. This includes but is not limited to digital libraries, the semantic web and technical documentation. A central part of the chapter will relate to a linguistic analysis of different classification dimensions. This does not only relate to specifying content, function, structure, and layout-oriented classification dimensions, but also to their interrelation and interdependency. Thus, special emphasis will be put on structure-oriented retrieval models which explore annotated data as an additional retrieval resource. Further, the chapter also presents results on the various roles of linguistic markers in automatic text classification and information retrieval.

1. Introduction

In developed countries, information and communication technologies (ICT) now contribute more to value creation than the classic technologies of automotive and mechanical engineering (Federal Ministry of Economics and Technology, 2009). Methods of acquiring, seeking and processing knowledge are a strategically vital issue in the context of globalized competition. Current search engines like Google or Yahoo are able to retrieve documents on the web containing specific words or phrases within fractions of seconds. They are routinely used by millions of users and allow the access to information from all over the world with unprecedented ease. In spite of their success the major search engines rely on matching sequences of characters and are unable to capture the semantics of words. Currently they cannot distinguish between different meanings of the same word, e.g. the car make jaguar and the animal jaguar. Therefore they usually fail in the focused retrieval of relational knowledge, e.g. finding all mentions of politicians who are scientists.

The ultimate goal as defined for the *Semantic Web* by Berners-Lee et al. (2001) is to extract a network of facts and relations between facts from a document which represents the semantic content of that document. In view of the enormous number of documents on the Internet, in libraries and archives, it is evident that manual annotation is infeasible. In the recent decade, methods for the automatic extraction of information from unstructured text in a collection (*corpus*) have opened new perspectives for organizing and analyzing documents and retrieving facts. These *text mining* approaches are based on data mining and statistics and employ concepts from linguistics and natural language processing. They treat the words, phrases and documents as data items having an extremely large number of different features. Two basic types of methods are used:

- *Classification methods* assign predefined classes to words, phrases and documents. The methods are *supervised* as they are trained using a representative set of text examples labeled with the correct classes. An example is the assignment of documents to thematic categories.
- *Clustering methods* group words, phrases or documents according to context similarity. These methods are *unsupervised,* processing a set of training documents which need no prior labeling. An example is grouping of words according to their thematic field using topic modeling.

As this process extracts structured information from plain documents, it is often called *information extraction*. Some methods aim at assigning words and phrases to the concepts of an *ontology*, a set of concepts and their relationships describing a domain (Gruber, 1993). The category system of *Wikipedia*, for instance, can be considered as an ontology (Mika, 2007).

Usually the methods are applied in a sequence forming a *semantic analysis pipeline,* where later steps use previously extracted features as inputs: First methods identify simple syntactic features like part-of-speech tags of words. Then *terms* are identified, which are the words and phrases carrying meaning, e.g. person names. Subsequently the terms are clustered according to semantic topics or may be assigned to different *concepts* of an ontology. Finally *relations* between concepts are extracted or even more complex *events* characterized by a number of relations.

Ultimately the original texts as well as the annotations are indexed in a *search engine*. By an appropriate query interface users can retrieve documents containing specific words as well as semantic concepts or relations between facts.

A brief introduction into text mining methods is given by Hotho et al. (2005). Feldman and Sanger (2007) present a Text Mining Handbook geared towards practical issues for large scale applications. The comprehensive book of Manning et al. (2008) introduces information retrieval and the extraction of semantic annotations. Sarawagi (2009) gives a short introduction to entity and relation extraction. These sources provide pointers to most aspects addressed in this chapter.

This chapter gives an overview over the approaches currently used. In the following section we discuss automatic procedures to structure a collection of documents, e.g. by assigning topic indicators to them in an automatic way. The third section considers methods for more finegrained semantic annotation of individual words and phrases within documents. The aim is to identify semantic concepts independent from the wording. The fourth section presents approaches to find relations between concepts and to capture the statements expressed in the text in more detail. The final section discusses how document collections with extracted indicators, concepts and relations may be searched to retrieve relevant content for the user.

2. Annotation of documents and passages

2.1. Document preprocessing

The digital documents of a corpus are usually stored as sequences of characters. Analysis algorithms require a number of preprocessing steps, before they can analyze the content. *Preprocessing* may include the following (p.57 Feldman and Sanger, 2007):

– *Layout Analysis* interprets the layout of a document to infer structural properties like title, abstract, sections, or the authors of a document. An overview is given by Konya et al. (2008).
– *Tokenization* aims at finding the *tokens* or *terms*, which are the words of a text as well as numbers, punctuation, etc. Usually a set of rules is used for this task. Part of this task is splitting the text into sentences (Manning et al., 2008).
– *Stop word removal* often is applied to omit frequent function words (like determiners, etc.) from the text as they usually carry no meaning and cause unnecessary processing load. Now and then very rare words (e.g. appearing less than 3 times in a corpus) are deleted as it is difficult to estimate their semantics.
– *Lemmatization* methods try to map verb forms to the infinite tense and nouns to the singular form. However, in order to achieve this, the word form has to be known, i.e. the part of speech of every word in the text document has to be assigned. Since this tagging process is usually quite time consuming and still error-prone, in practice frequently stemming methods are applied (Manning et al., 2008, p.30).
– *Stemming* methods try to build the basic forms of words, i.e. strip the plural ‘s’ from nouns, the ‘ing’ from verbs, or other affixes. A stem is a natural group of words with equal (or very similar) meaning. After the stemming process, every word is represented by its stem. A well-known rule based

stemming algorithm has been originally proposed by Porter (1980). He defined a set of production rules to iteratively transform (English) words into their stems.

- *Part-of-speech (POS)* tagging assigns syntactic labels (like noun, verb, adjective, etc.) to the tokens. In addition gender and temporal information may be added. POS-tagging can be done using rules or may employ more complex sequential models (Brants, 2000).
- *Weight determination* tries to estimate the information content of a word. Large weights are assigned to terms that are used frequently in relevant documents but rarely in the whole document collection. Let us denote tf(d;t) as the frequency of term (e.g. word) t in document d and df(t) the fraction of documents containing term t. Then the tf-idf weight is defined as tfidf = tf(d;t)*log(1/df(t)) taking into account the specificity of a term as well as its frequency in a document (Salton and Buckley, 1988). There are many variants of these weighting schemes (Manning et al., 2008, p.107).

Usually these tasks are solved with a set of rules or regular expressions providing a flexible means of matching patterns in text (Goyvaerts and Levithan, 2009). Note also that not all preprocessing steps have always to be performed.

2.2. Text classification

Well known access structures for document collections are library catalogs or book indexes, which allow fast access to relevant information. *Text classification* (Sebastiani, 2002, p.64) (Manning et al., 2008, p.234–320) aims at assigning predefined *categories* or classes to text documents. An example would be to automatically label each incoming technical document with a category like mechanics, electronics, or automotive. Classification starts with a *training set* $D = \{d_1, \ldots, d_n\}$ of documents that are already labeled with a category $y \in Y$ (e.g. mechanics, electronics, …). The document d is usually represented by a vector of features, e.g. the counts of words appearing in the document (*vector space representation* (Salton, 1989)). The task is then to determine a *classification model*, a function

$$f_\theta : d \rightarrow y \qquad (1.1)$$

which is able to assign the correct class y to a new document d_{new} not contained in the training set. The model function f_θ has a vector θ of free *parameters*. The optimal parameter value θ^* is selected by optimization methods in such a way that f_{θ^*} can reconstruct the classes in the training set in an optimal way. This process is referred to as *training*. Subsequently the classifier can be applied to a new document yielding the classification $f_{\theta^*}(d_{new}) \in Y$.

Text classification can be used to get more focused search results. If, for instance, we retrieve documents with category automotive containing the term jaguar we will get only documents describing aspects of jaguar cars.

2.2.1. Performance measures

To measure the performance of a trained classification model a random fraction of the labeled documents is set aside and not used for training. We may classify the documents of this *test set* with the classification model and compare the predicted labels with the true labels. The fraction of correctly classified documents in relation to the total number of documents is called *accuracy* and is a first performance measure. Often, however, the target class covers only a small percentage of the documents. Then we get a high accuracy if we assign each document to the alternative class. To avoid this effect different measures of classification success are often used (van Rijsbergen, 1975; Baeza-Yates and Ribeiro-Neto, 1999). *Precision* quantifies the fraction of retrieved documents that are in fact relevant, i.e. belong to the target class. *Recall* indicates which fraction of the relevant documents is retrieved (see also Menke in this volume).

$$\text{precision} = \frac{\#\{\text{relevant} \cap \text{retrieved}\}}{\#\text{retrieved}} \qquad \text{recall} = \frac{\#\{\text{relevant} \cap \text{retrieved}\}}{\#\text{relevant}} \tag{1.2}$$

If the training set and the test set are independent random samples of the underlying document collection then performance estimates derived from the test set are unbiased estimates of the true performance (Bengio and Grandvalet, 2004).

Obviously there is a trade off between precision and recall. Most classifiers internally determine for a document d some degree of membership in the target class. If only documents of high membership degree are assigned to the target class, the precision is high. However, many relevant documents might have been overlooked, which corresponds to a low recall. When on the other hand the search is more exhaustive, recall increases and precision goes down. The *F-score* is a compromise of both for measuring the overall performance of classifiers.

$$F = \frac{2}{1/\text{recall} + 1/\text{precision}} \tag{1.3}$$

Often the costs of classification errors depend on the class. In email classification, for instance, it is much more costly to categorize a legitimate email as spam (and delete it) than to classify a spam email as legitimate (and keep it). These costs can be incorporated into specific performance measures Manning et al. (2008) and can even be taken into account by most training algorithms for classifiers (Sebastiani, 2002).

There are a number of benchmark data sets, e.g. the Reuters-21587 collection, which have been used to assess the relative performance of classifiers (Manning et al., 2008, p.258). In the following section we discuss some classifiers relevant for text mining and illustrate their properties.

2.2.2. *Naive Bayes Classifiers*

Probabilistic classifiers start with the assumption that the words of a document d_i have been generated by a probabilistic mechanism. It is supposed that the class y_{d_i} of a document d_i has some relation to the words w_k which appear in the document. This may be described by the conditional distribution $p(w_1, \ldots, w_K \mid y_{d_i})$ of the K words given the class of the document. Then the *Bayesian formula* yields the probability of a class given the words of a document (Mitchell, 1997)

$$p(y \mid w_1, \ldots, w_{n_i}) = \frac{p(w_1, \ldots, w_{n_i} \mid y)p(y)}{\sum_{j=1}^{m} p(w_1, \ldots, w_{n_i} \mid y_j)p(y_j)} \tag{1.4}$$

Note that each document is assumed to belong to exactly one of the m categories in Y. The *prior probability* $p(y)$ denotes the probability that an arbitrary document belongs to class y before its words are known. Often the prior probabilities of all classes may be taken to be equal. The conditional probability on the left is the desired *posterior probability* that the document with words $w_1, \ldots, w_{n_i}$ belongs to category y. We may assign the class with highest posterior probability to our document.

For document classification it turned out that the specific order of the words in a document often is not very important. Even more we may assume that for documents of a given class a word appears in the document irrespective of the presence of other words. This leads to a simple formula for the conditional probability of words given a class y: $p(w_1, \ldots, w_{n_i} \mid y) = p(w_1 \mid y) *\ldots* p(w_{n_i} \mid y)$. Combining this naive independence assumption with the Bayesian formula defines the Naive Bayesian classifier Manning et al. (2008, p.237). Simplifications of this sort are required as many thousand different words occur in a corpus.

The naive Bayes classifier involves a training step which requires the estimation of the probabilities of words $p(w_j \mid y)$ for each class y by the relative frequencies in the documents of a training set which are labeled with y. In the classification step the estimated probabilities are used to classify a new document d_{new} according to the Bayes rule. Although this model is unrealistic due to its restrictive independence assumption it yields surprisingly good classifications (Joachims, 1998) although it is usually inferior to more complex approaches (Caruana and Niculescu-Mizil, 2006). Currently it is used in the Thunderbird email client to identify spam emails. It may be extended into several directions (Sebastiani, 2002).

As the effort for manually labeling the documents of the training set is high, some authors in addition use unlabeled documents for training. Assume that from a small training set it has been established that word w_i is highly correlated with class y. If from unlabeled documents it may be determined that word w_j is highly correlated with w_i, then also w_j is a good predictor for class y. *Semi-supervised classifiers* exploit unlabeled documents to improve classification performance. Nigam et al. (2000) used labeled as well as unlabeled documents for a nave Bayes classifier and were able to reduce the classification error by up to 30 %.

2.2.3. Generalization and overfitting

Data mining provides a large number of classification models which may all be applied to text classification. In contrast to the nave Bayes classifier they often make far less restrictive assumptions and are able to represent much broader forms of functional dependencies. For *decision tree classifiers* (Breiman et al., 1984; Schmid, 2010) the final class depends on combinations of specific input features (e.g. the presence of specific words in a document) such that all considered combinations form a tree-like structure. *Nearest neighbor classifiers* (Cover and Hart, 1967) define a distance measure for the documents d_i (e.g. based on the number of common words) and selects the class for a new document d according to the classes of its nearest neighbors in the training set. These and further text classifiers are discussed by Feldman and Sanger (2007, p.70–76).

Usually classifiers employ a *bag-of-words* representation (Salton, 1989), where a document is represented by a vector indicating the count of words it contains. It is clear that this representation ignores essential aspects of a text, especially the order of words (see (Halliday and Hasan, 1976) for a classical book on text structure). Hence often other features are added, e.g. *bigram* or *n-gram* features representing sequences of 2 or n words. However, this drastically increases the number of parameters. If, for instance, a document collection contains 100000 different words then there are about 10^{10} possible bigrams and the nave Bayes model requires a probability estimate for each bigram. For limited training sets this results in very unreliable parameter estimates and sharply reduces model performance.

To avoid this *overfitting* phenomenon there are two different strategies available. The first strategy actively selects input features which are likely to have an effect on the class variable and ignores the remaining features. *Feature selection methods* take into account the strength of the relation between a specific feature and the class y, as measured by the correlation, the information gain or other indices (Sebastiani (2002) gives an overview of such indexes). Experiments show that this strategy may reduce the number of features while increasing the performance (Feldman and Sanger, 2007, p.69).

2.2.4. Support vector machines

The other approach to avoid overfitting is the utilization of models with an inherent *regularization* which reduces the complexity of model structures. A very successful classifier with this property is the *Support Vector Machine* (SVM), a supervised classification algorithm (Vapnik, 1995) that has been successfully applied to text classification tasks (Joachims, 1998). As usual a document d is represented by a – possibly weighted – vector $x = (w_1, \ldots, w_k)$ of the counts of its words or bigrams. A single SVM can only separate two classes: a positive class (indicated by $y = 1$) and a negative class (indicated by $y = -1$). In the space of input vectors x a hyperplane may be defined by setting $y = 0$ in the following linear equation

$$y = f_\theta(x) = b_0 + \sum_{j=1}^{k} b_j w_j \tag{1.5}$$

The parameter vector is given by $\theta = (b_0, b_1, \ldots, b_k)$. The SVM algorithm determines a hyperplane which is located between the positive and negative examples of the training set. The parameters b_j are adapted in such a way that the distance ξ – called *margin* – between the hyperplane and the closest positive and negative example documents is maximized, as shown in figure 1. This amounts to a constrained quadratic optimization problem which can be solved efficiently for a large number of input vectors.

The documents having distance ξ from the hyperplane are called *support vectors* and determine the actual location of the hyperplane. Usually only a small fraction of documents are support vectors. A new document x_{new} is clas-

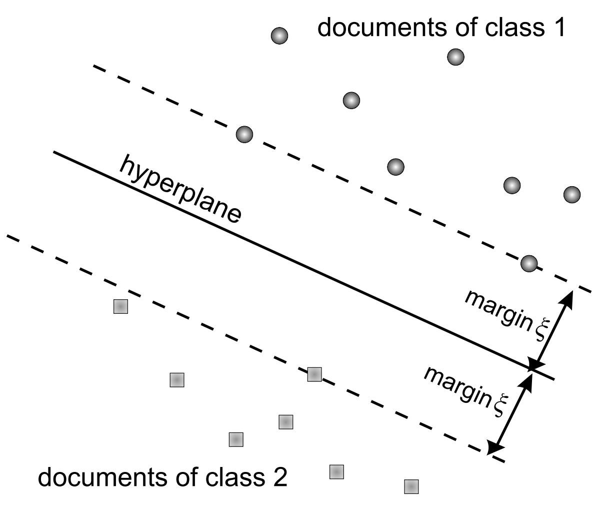


Figure 1. The hyperplane estimated by the support vector machine separates the two classes in the n-dimensional space.

sified as $y = 1$ if the value $f_{\theta*}(x_{new}) > 0$ and as $y = -1$ otherwise. In case that the document vectors of the two classes are not linearly separable a hyperplane is selected such that as few as possible document vectors are located on the wrong side.

SVMs can be extended to a non-linear predictor by transforming the usual input features in a non-linear way using a *feature map*, e.g. the quadratic feature map containing products $w_i w_j$

$$\varphi(w_1, \ldots, w_k) = (w_1, \ldots, w_k, w_1 w_1, \ldots, w_1 w_k, \ldots, w_k w_1, \ldots, w_k w_k) \qquad (1.6)$$

Subsequently a hyperplane may be defined in the expanded input space. Such non-linear transformations define extensions of scalar products between input vectors, which are called *kernels* (Shawe-Taylor and Cristianini, 2004).

If the kernels fulfill some symmetry conditions the feature map φ need not be known explicitly. In this way kernels even may be applied to non-vector elements like trees (Reichartz et al., 2010) and graphs (Gärtner, 2003).

The most important property of SVMs is that according to the theoretical bounds (Vapnik, 1995) learning is nearly independent of the dimensionality of the feature space. It it can cope with an extremely large number of features as it inherently selects a number of data points (the support vectors) defining the hyperplane, ignoring the remaining data points. This allows good generalization and makes SVMs especially suitable for the classification of texts(Joachims, 2002, p.114). To limit the random effects of unnecessary features and to reduce computation time there exist nevertheless feature selection approaches for the SVM as surveyed by Nguyen and De la Torre (2010).

In a broad empirical test Sebastiani (2002) evaluated different classifiers on benchmark data. It turns out that the level of performance of course depends on the document collection. Table 1 gives some representative results achieved for the Reuters 20 newsgroups collection (Sebastiani, 2002, p.38). Boosted trees, SVMs, and k-nearest neighbors usually deliver top-notch classifier performance, while naive Bayes and decision trees are less reliable.

Table 1. Performance of Different Classifiers for the Reuters collection.

Classifier method	F-score in %
naive Bayes	79.5
decision tree C4.5	79.4
k-nearest neighbor	85.6
SVM	87.0
boosted tree	87.8

Cesa-Bianchi et al. (2006) study hierarchical classification with high-level categories containing lower-level categories when an instance could belong to more than one class node in the underlying taxonomy. They use a simple hierarchy of Support Vector Machines (SVM) with a top-down evaluation scheme and yield good performance on this kind of task. Liu et al. (2005) apply SVMs for the classification of the complete *Yahoo!* directory and note a small gain in considering the hierarchy as a whole. A more general approach that can be applied to modeling relations between classes is presented by Tsochantaridis et al. (2005).

Text classification is applied in many contexts, ranging from document filtering to automated metadata generation, word sense disambiguation, and will be employed in more complex tasks discussed in later sections.

2.3. Document clustering

Clustering methods aim at finding groups of documents in a collection with similar content. The result of clustering is typically a partition of the collection also called *clustering* $\mathscr{P} = \{C_1, \ldots, C_m\}$. Each *cluster* C_r consists of a set of documents d. The documents of a cluster should be similar and dissimilar to documents of other clusters. The similarity is measured by a *similarity measure* $S(d_i,d_j) \in \Re^+$ assigning a score to each pair (d_i,d_j) of documents. This will influence the shape of the clusters, as some elements may be close to one another according to one similarity and further away according to another. Again a numerical representation of a document d_i is needed, e.g. the bag-of-words representation $x_i = (x_{i1}, \ldots, x_{iK})$ with the counts of words in a document.

The most popular similarity measure for text document clustering is the *cosine similarity* measuring the angle between the document vectors (Manning et al., 2008, p. 111)

$$S(x_i,x_j) = \frac{x_i\, x_j}{||x_i||\,||x_j||} \qquad (1.7)$$

where the numerator represents the dot product $x_i \cdot x_j = \Sigma_k\, x_{i,k} \cdot x_{j,k}$ (also known as the inner product) of the vectors x_i and x_j, while the denominator is the product of their Euclidean lengths. The cosine similarity can be seen as a method of normalizing document length during comparison.

The simple bag-of-words representation may be enhanced by weighting the terms according to their information content by (e.g. tf-idf). In addition we may include semantic terms, e.g. ontology concepts, leading to similarity measures representing semantic content more closely. Methods discussed in Section 3 assign such concepts to each word or phrase in a text. Hotho et al. (2003) explore different strategies yielding promising results: replacing words by semantic terms or attaching a vector of semantic terms to the bag-of-words representation. Jing et al. (2006) evaluate the WordNet taxonomy to construct semantic annotations yielding improved clustering results.

2.3.1. K-Means clustering

k-means is one of the most frequently used text clustering algorithms (Manning et al., 2008, p.331) assuming a fix number of *k* clusters. It generates a starting solution in which all documents are distributed on *k* clusters. Then it tries to improve the solution by the following steps:

1. Compute the center of each cluster using the similarity measure.
2. Assign each document to the nearest cluster center.

The two steps are repeated until the assignment of documents does no longer change. It can be shown that this procedure leads to a local minimum of the sum of distances to cluster centers (Hastie et al., 2001, p.462). In data mining a large number of variants of these techniques have been developed.

2.3.2. Model-based clustering

Model-based Clustering assumes that the underlying document collection was generated according to a probabilistic model and tries to recover the original model from the data. For instance, we can assume that the model first randomly generates cluster centers and subsequently adds some noise to produce the individual documents. If the noise is normally distributed and the covariance is spherical this approach will result in clusters of spherical shape (Manning et al., 2008, p.338).

In general cluster models are mixtures of component distributions, e.g. Gaussian components

$$p(x|\alpha, \mu, S) = \sum_{i=1}^{k} \alpha_i p(x|\mu_i, S_i) \tag{1.8}$$

with mixture weights α_i. Alternatively we may use discrete component distributions, e.g. multinomial. A commonly used algorithm for solution is the *Expectation-Maximization* (EM) algorithm (Dempster et al., 1977). It estimates optimal values for the parameters α, μ, S. For a new document x_{new} the density $p(x_{new}|\mu_i, S_i)$ with respect to all component distributions may be determined and we may assign it to the component with highest density. Note that $p(x_{new}|\mu_i, S_i)$ may interpreted as a degree of membership producing a soft clustering.

A difficult issue in clustering is the estimation of the number *k* of clusters. To compare the cluster quality for different *k* we can use index numbers like the *purity* (Manning et al., 2008, p.328) which compares the generated clustering with a manually obtained partition according to a gold standard. Alternatively we may use criteria like the *Akaike Information Criterion* which takes into account the number of free parameters (Hastie et al., 2001, p.203).

2.3.3. Hierarchical clustering

An alternative approach to clustering does not provide a flat set of clusters, but a hierarchy of clusters (Hastie et al. 2001, p. 472). By merging clusters of lower level we get a cluster of a higher level. As the relation between clusters is explored the resulting hierarchy is more informative than a simple clustering. Hierarchical clustering does not require a pre-specified number of clusters. There are two main approaches to hierarchical clustering: *hierarchical agglomerative clustering* first treats each document as a single cluster and then merges clusters until it arrives at a single cluster with all elements. *Divisive clustering* starts with all documents in one cluster and successively splits clusters. An extensive discussion of these techniques is given by (Manning et al., 2008, p.346).

2.3.4. Search result clustering

An important application of clustering is the structuring of search results provided by a search engine. For instance, if we search for jaguar we get different web pages discussing the jaguar car, the jaguar cat or the jaguar operating system. Using clustering techniques these documents can be grouped according to their similarity, which should lead to clusters with automotive and biological content. The dynamic nature of the process together with the interactive use of clustering results poses a number of specific requirements (Carpineto et al., 2009):

- Meaningful *labels*. To give the user a fast information of the contents of a cluster there should be cluster labels, which concisely indicate the contents.
- Computational *efficiency*. To be useful for web browsing the clustering has to be computed in a fraction of a second.
- Unknown *number of clusters*. In search result clustering the size as well as the number of clusters have to be determined on the fly.
- *Overlapping* and *hierarchical clusters*. In practice it is useful that a single search result may belong to different clusters. Structuring clusters as hierarchies gives important insights into the relation between clusters.

In principle all the algorithms mentioned before can be used to construct search result clusters. In a subsequent step meaningful labels have to be extracted from the clusters. It turned out that this usually yields a set of un-intuitive labels for the clusters. Usually the extracted keywords are diverse and the content of a cluster is difficult to understand. As a consequence a number of *description-centric clustering algorithms* directly introduce cluster labeling into the procedure of defining clusters. This reflects the view that a cluster, which cannot be described are useless. An example of such an algorithm was implemented in the open source Carrot framework (Osinski and Weiss, 2005).

2.3.5. Topic detection and tracking

Often new documents arrive in a stream, e.g. news stories or new patents, and it is desirable to group them into semantically coherent groups. *Topic detection and tracking* (Allan et al., 2005) aims at dynamically creating clusters. The development of clusters is monitored over time, and the system must decide whether or not two randomly selected stories are on the same topic. It comprises two steps (1) *New event detection* recognizes when a story in the stream discusses a topic that had not been seen previously; (2) *Clustering* places stories into appropriate clusters, such that each cluster includes stories about the same topic. An example of a successful system is the *European Media Monitor* (Steinberger et al., 2009). Every ten minutes and in each of 23 languages it clusters the latest news items and present the largest cluster as the current top-ranking media theme (Top Stories). It allows to access to news content across languages and merges the extracted information.

3. Semantic annotation of words and phrases

3.1. Information extraction

Information extraction (IE) refers to the automatic extraction of meaningful information from unstructured machine-readable text documents in a corpus (Feldman and Sanger, 2007, p.94). Examples are the annotation of phrases as *entities* (names of persons, organizations, etc.) or as *concepts* from an *ontology* (e.g. Wikipedia). Moreover *relationships* between concepts or entities may be extracted. From the sentence `On April 20, 2009, Sun and Oracle announced that Oracle will acquire Sun`. we may among others extract the following information: *Date*(April 20, 2009), *Company*(Sun), *Company*(Oracle) as well as *AcquisionOf*(Oracle, Sun). The extracted information allows much richer forms of queries than possible with keyword retrieval alone, as searches for character strings may be combined with searches for entities, concepts, and their relations.

IE supports the Semantic Web by providing the required information for automatic processing of the contents of documents (see Shadbolt et al. (2006); see also Waltinger & Breuing in this volume). Annotations can be stored with the text, e.g. as eXtensible Markup Language (XML) tags (see Walmsley (2002); see also Rahtz and Stührenberg both in this volume). A typical application is to scan a set of documents written in a natural language and populate a database with the information extracted.

After basic preprocessing has been performed (see Section 2.1) IE is in general performed as a sequence of tasks. The first group of tasks identifies relevant terms for analysis:

- *Linguistic annotation* assigns syntactic categories to single words (*Part-of-speech tagging*), identifies meaningful phrases (*phrase chunking*) or analyzes the syntactic structure relating phrases (*parsing*).
- *Terminology extraction* detects terms (words, phrases, etc.) relevant for documents from a domain, e.g. by combining statistical and syntactic information.

The next group of tasks is focused on the assignment of terms to *sense indicators*

- *Named Entity Recognition* identifies entity names (*person*, *organization*, *location*, etc.), temporal expressions as well as numerical expressions (e.g. *moneyAmount*) (Sang and Meulder, 2003).
- *Topic modeling* derives semantic descriptors for words by grouping words into clusters according to the words in their neighborhood (Blei and Lafferty, 2009).
- *Coreference resolution* detects links between different mentions of the same entity (or concept) in a document or resolves the relation between pronouns and entities (Versley et al., 2008).
- *Word sense disambiguation* distinguishes between different senses of a word, e.g. the `jaguar` car and the `jaguar` cat, usually by comparing the words occurring in the neighborhood of a term (Schütze, 1998). If possible the terms are assigned to the concepts of an external taxonomy or ontology (Sowa, 2000b).

The final group of procedures concentrates on the extraction of relations between concepts and entities.

- *Relation extraction* aims at extracting semantic relationships between concepts or entities, e.g. the *AquisitionOf* relation between two *company* entities (Suchanek et al., 2006a).
- *Ontology Learning* extracts relevant concepts and relations from a given corpus or resources to create or enhance an ontology (Buitelaar and Cimiano, 2008).

Finally the extracted annotations like *person*-entity and *AquisitionOf*-relation are entered into the database. Depending on the analysis goal usually only a some of these tasks are actually performed.

Two different approaches may be used to arrive at automatic extraction methods. For simple tasks explicit rules may be constructed manually. For more

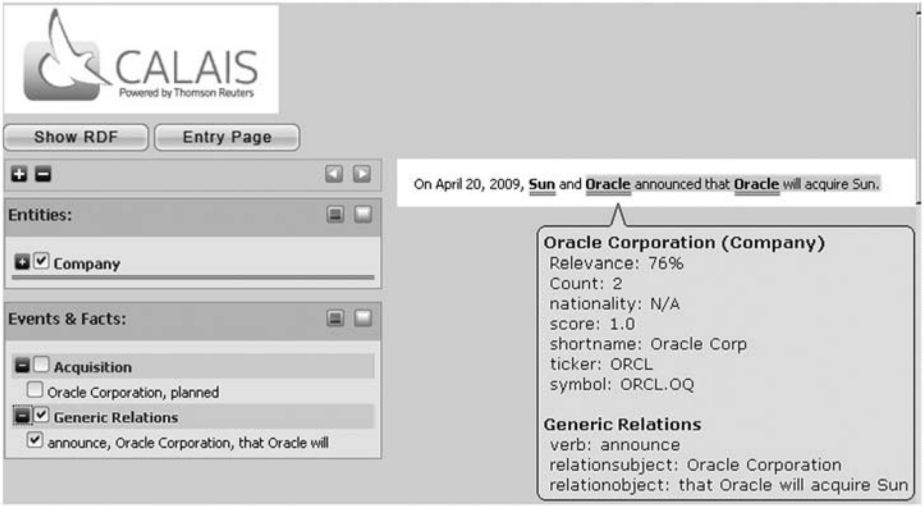


Figure 2. Screenshot of the Open Calais website for the semantic analysis of English text.

complex problems statistical classification and clustering models are determined using training data. In recent years substantial progress has been made in applying data mining techniques to a number of these steps.

There exist several commercial information extraction systems. In Figure 2 the result of a semantic analysis by Open Calais (http://viewer.opencalais.com/) is shown. It is provided as a web service free for commercial and non-commercial use.

3.2. Linguistic annotation

Linguistic processing methods annotate text with syntactic and linguistic categories: *Part-of-speech tagging* (POS) determines the part of speech tag, e.g. noun, verb, adjective, etc. for each term. *Text chunking* aims at grouping adjacent words in a sentence. An example of a chunk is the noun phrase the `strange particle`. *Parsing* produces a full parse tree of a sentence. From the parse, we can find the syntactic relation of each word in the sentence to all the others, and typically also its function in the sentence (e.g. subject, object, etc.).

Linguistic processing either uses lexica and other resources (like WordNet, see Section 3.7) as well as hand-crafted rules. If a set of examples is available the classifiers in Section 2.2 or machine learning methods as described in Section 3.4.2 may be employed to learn the desired tags. Several high quality packages are available, e.g. the Stanford Parser de Marneffe et al. (2006) for English and German.

3.3. Terminology extraction

Before the relevant terminology of a given domain can be represented for text mining purposes, important concepts (or terms) of this domain have to be identified. It is especially important that multiword terms are determined (see Pazienza et al. (2005); see also Sharoff and Hartley in this volume) which have to be treated as single features in subsequent analyses.

Most automatic methods to identify terminology rely on syntactic features, e.g. part of speech patterns and select between candidates by significance tests (Daille et al., 1994; Witschel, 2005). These tests estimate, e.g. the likelihoodratio test (Hastie et al., 2001), if the co-occurrence of the two terms occurs significantly more frequent than random (see Evert (2008) for an overview). Test with given vocabularies show that the approach may extract relevant terms with a high precision. To identify the terminology of a specific domain terminological units can be identified by comparing a word's relative frequency in a given specialized subcorpus to its relative frequency in a large, well-balanced corpus covering many aspects of everyday language (Wong et al., 2007). Kim et al. (2010) report on a competition on terminology extraction from scientific articles discussing the performance of different approaches.

3.4. Named entity recognition

3.4.1. Named entity recognition as a classification problem

Named entity extraction has the task to identify names in a text and assign it to one or more categories, e.g. person, location, organization. This was originally formulated in the Message Understanding Conference (Chinchor, 1997; Sang and Meulder, 2003). One can regard it as a word-based tagging problem: The word, where the entity (e.g. company) starts, gets tag B-C, continuation words get tag I-C and words outside the entity get tag o. This is done for each type of entity of interest. In the following example the annotation for company (*C*) and time (T) is shown:

o	B-T	I-T	B-C	o	o	B-C	o	o	B-C	o
On	April	20	Oracle	announced	that	Oracle	will	acquire	Sun	.

Hence we have a sequential classification problem for the labels of each word, with the surrounding words forming the input feature vector. A frequent way of forming the feature vector is a binary encoding scheme. Each feature component can be considered as a test that asserts whether a certain pattern occurs at a specific position or not. For example, a feature component takes the value 1 if the previous word is the word John and 0 otherwise. Of course we may not only test the presence of specific words but also whether the words starts with a capital letter, has a specific suffix or is a specific part-of-speech. In this way results of previous analyses may be used (Sang and Meulder, 2003).

In the beginning of named entity recognition, approaches were basically list based and performed dictionary look ups. Clearly, these approaches were not able to recognize unseen names. Later, linguists created complicated rule sets for this task (Nadeau and Sekine, 2007). But since this was too time consuming machine learning approaches were created. We may employ any efficient classification method to classify the word labels using a window around the word in the input vector. A good candidate are Support Vector Machines because of their ability to handle large sparse feature vectors efficiently.

3.4.2. Conditional random fields

It is clear that the sequence of words gives important clues for the detection of entities. An approach for modeling sequences of words is called conditional random field (CRF) (Lafferty et al., 2001). Again we consider the observed vector of words x and the corresponding vector of labels y with an underlying sequence or graph structure. For a label y_t let $N(y_t)$ be the indices of neighboring labels. Then a conditional random field assumes that when conditioned on the vector t of all terms the random variables obey the *Markov property*

$$p(y_t | (y_u)_{u \in N(y_t)}, x) = p(y_t | (y_u)_{u \neq t}, x) \tag{1.9}$$

i.e. the whole vector x of observed terms and the labels of neighbors $(y_u)_{u \in N(y_t)}$ are sufficient to describe the distribution of the label y_t. Note that there is no need to specify a model of the distribution $p(x)$ of the observed words, which may exhibit arbitrary dependencies.

We consider the simple case that the words $x = (x_1, \ldots, x_T)$ and the corresponding labels $y = (y_1, \ldots, y_T)$ have a chain structure, i.e. and that y_t depends only on the preceding and succeeding labels y_{t-1} and y_{t+1}. Then the conditional distribution $p(y|x)$ has the form

$$p(y|x) = \frac{1}{Z(x)} \, exp \left(\sum_{i=1}^{T} \sum_{r=1}^{k_j} \lambda_{t,r} f_{t,r}(y_t, y_{t-1}, x) \right) \tag{1.10}$$

where the $f_{t,r}(y_t, y_{t-1}, x)$ are different *features functions* related to the pair (y_t, y_{t-1}). To reduce the number of free parameters $\lambda_{t,r}$ the feature functions are assumed to be time-invariant, i.e for all t we assume $f_{t,r}(y_t, y_{t-1}, x) =: f_r(y_t, y_{t-1}, x)$ and $\lambda_{t,r} =: \lambda_r$. Note that specific feature functions need not depend on all possible arguments, we may for instance have functions $f_r(y_t, x_t)$ or $f_r(x_t)$. CRF models allow arbitrary dependencies in the observation sequence and more complex neighborhood structures of labels.

As for most machine learning algorithms a training sample of words and the correct labels is required. In addition to the identity of words arbitrary properties of the words, like part-of-speech tags, capitalization, prefixes and suffixes, etc. may be used leading to sometimes more than a million features. The unknown parameter values λ_r are usually estimated using conjugate gradient optimization routines (McCallum, 2003; Sutton and McCallum, 2007).

McCallum (2003) applies CRFs with feature selection to named entity recognition and reports the following F-scores for the CoNLL corpus: person names 93 %, location names 92 %, organization names 84 %, miscellaneous names 80 %. CRFs also have been successfully applied to noun phrase identification (McCallum, 2003), part-of-speech tagging (Lafferty et al., 2001), shallow parsing (Sha and Pereira, 2003), and biological entity recognition (Kim et al., 2004).

3.4.3. Weakly supervised approaches

The classic named entity recognition problem assumes a large, high quality set of labeled entities. This works well for relatively small numbers of entity types (people, places, organizations) in relatively uniform, high quality corpora. It does not generalize well to hundreds of entity types in multiple types of corpora, e.g. the web. Therefore methods have been proposed involving the semiautomatic generation of training data, and the simultaneous labeling of many different entity types.

The approach of Whitelaw et al. (2008) starts with a list of seed names (e.g. Persons) with known entity type, e.g. taken from Wikipedia. Then the following steps are performed to find a large set of trusted training examples:

1. Find seed terms on the web by exact string match, e.g. for `Albert Einstein`.
2. Extract context of three tokens on each side of each term found, e.g. `it true that Albert Einstein said that when`.
3. Generate all possible shorter templates by trimming, e.g. `Albert Einstein said`.
4. Filter templates based on quality measures like specificity (> 80 % matches of a single type), and diversity (> 50 names).
5. Match these established templates against the web to find trusted mentions of named entities.

Examples of patterns found in this way are `[person] was born on`, `Prime Minister [prime minister], who`, or `won the [award] for best`. The approach produced about 60 million trusted mentions. By assuming a single sense per page, and trusting all mentions of the same name on the same page as a trusted mention, this is expanded to more than 200 million.

A feature was considered important if it occurs frequently (extremely rare features are not helpful), and if it is highly correlated with any of the type labels.

A *t*-statistic was used to measure these properties and extract appropriate features. Subsequently a fast version of the SVM was used for estimating a named entity classifier. Simultaneously learning templates for all different types provides a major benefit, as it allows to assess the quality of the proposed templates; the different types provide negative examples that would otherwise be missing. The authors achieve the following test set F-scores for different named entities: company: 50,6 %, fish: 51,4 %, person: 92,9 %, location: 86,1 %. These results are very similar to the accuracy achieved by traditional approaches.

Other weakly supervised approaches to named entity recognition are surveyed by (Nadeau and Sekine, 2007) and discussed in the context of relation extraction in Section 4.3.

3.5. Topic modeling

Topic models are probabilistic models for uncovering the underlying semantic structure of a document collection based on a hierarchical Bayesian analysis of the original texts (Hofmann, 2001; Blei and Lafferty, 2009). For example, a topic model might have topics CAT and DOG. The CAT topic has probabilities of generating various words: `milk`, `meow`, `kitten` and of course cat will have high probability given this topic. The DOG topic likewise has probabilities of generating each word: `puppy`, `bark` and `bone` might have high probability. Words without special relevance, like the, will have roughly even probability between classes (or can be placed into a separate category). Topic models have been applied to many kinds of documents, including scientific abstracts (Griffiths and Steyvers, 2004), and newspaper archives.

Latent Dirichlet Allocation (LDA) determines an unsupervised clustering of words which tend to appear together in the documents of a collection The idea behind LDA is to model documents as arising from multiple topics, where a topic is defined to be a distribution over a fixed vocabulary of terms. Specifically, we assume that K topics are associated with a collection, and that each document exhibits these topics with different proportions. This is often a natural assumption to make because documents in a corpus tend to be heterogeneous, combining a subset of main ideas or themes.

LDA assumes a probabilistic generative process to produce the observed data. Let K be a pre-specified number of topics and V the number of words in the vocabulary. To create a new document with m words $w_{i_1}, \ldots, w_{i_m}$ first a distribution of topics t for the document d is generated, which specifies the probability $p(t|d)$ that word w_{i_j} in the document has topic t. In addition for each topic t there is a distribution $p(w|t)$ which specifies for each word w the probability of word w given topic t. Note that each word in a document may belong to a different topic. To avoid overfitting Bayesian Dirichlet prior distributions are specified for $p(t|d)$ and $p(w|t)$.

Table 2. Most important words for five of the 50 topics derived for articles of the Science Journal from 1980–2002.

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
computer	chemistry	cortex	orbit	infection
methods	synthesis	stimulus	dust	immune
number	oxidation	fig	jupiter	aids
two	reaction	vision	line	infected
principle	product	neuron	system	viral
design	organic	recordings	solar	cells
access	conditions	visual	gas	vaccine
processing	cluster	stimuli	atmospheric	antibodies
advantage	molecule	recorded	mars	hiv
important	studies	motor	field	parasite

Although there are no labeled examples the algorithm is able to estimate both distributions $p(t|d)$ and $p(w|t)$ by treating the topic variable for each word as an unobserved latent variable. In this way it yields $p(t|d,w_{ij})$, i.e. the probability that the j-th word w_{ij} in document d belongs to topic t. With these variables, LDA is a type of mixed-membership model as each word of a document may belong to a different topic. These are distinguished from classical mixture models (Nigam et al., 2000), where each document is limited to exhibit one topic. This additional structure is important because documents often exhibit multiple topics. Several computational techniques have been developed for LDA, including mean field variational inference (Blei et al., 2003) and Gibbs sampling (Steyvers and Griffiths, 2006). To give an example for the distribution $p(w|t)$ of words for a topic table 2 shows the most important words for five of 50 different topics estimated for the science journal (Blei and Lafferty, 2009).

For a new document d an estimated topic model assigns a posterior distribution $p(t|d,w_{ij})$ of topics t to each word w_{ij}. The posterior topic proportions gives a sense of which topics the document is about. These vectors can also be used to group articles that exhibit certain topics with high proportions. The posterior topic proportions define a topic-based similarity measure between documents and provide a low dimensional representation of each document. Applications of topic modeling include tasks such as information retrieval, where it is able to improve retrieval performance (Wei and Croft, 2006).

Choosing the number of topics is a persistent problem in topic modeling and other latent variable analysis. One can use cross validation on predictive likelihood, essentially choosing the number of topics that provides the best language model. Paaß et al. (2004) developed a hierarchical topic model which was later extended by (Teh et al., 2007) employing hierarchical Dirichlet processes where the number of topics is automatically selected. A large number of extensions and applications of topic models is discussed in (Blei and Lafferty, 2009).

3.6. Coreference resolution

Coreference resolution, also called *anaphora resolution*, is the process of matching terms, i.e. words or phrases in a text, that refer to the same entity in the real world. There are two main approaches to coreference resolution: one is based on hand-coded rules using linguistic analysis of the sentence, the other is based on machine learning using an annotated training corpus. Bagga and Baldwin (1998) distinguish eleven types of coreference analysis. The most important are (see Clark (1977) for these and related bridging relations; see Vieira and Teufel (1997) for an account of bridging in computational linguistics; see Halliday and Hasan (2000) for a text book on textual cohesion):

– *Proper nouns* referring to the same entity: Sometimes first names or academic titles are omitted. For example `Steve Jobs` and `Mr. Jobs` refer to the same person.
– *Pronouns* referring to an entity: `Steve Jobs` might later be referred to by a pronoun such as he.
– Appositions are constructions like `Bill Gates, the chairman of Microsoft`, …
– Exact Matching of the same term: An example could be multiple mentions of a word like computer.
– Synonymy may refer to each other, e.g. the words `display` and `screen`.

Most algorithms for coreference resolution concentrate on proper nouns and pronouns. A number of rule-based approaches are discussed by Feldman and Sanger (2007, p.109). Machine learning approaches usually requires a labeled training set, where pairs or subsets of phrases are marked as coreferent. Then a classifier, e.g. decision tree (see Section 2.2), is trained to classify each pair of phrases as coreferent or not. This approach with mutually independent pairwise decisions may lead to inconsistent assignments. Therefore later approaches shift attention to more complex models. (Rahman and Ng, 2009) form clusters of coreferencing phrases and use a ranking approach to select clusters. Bengtson and Roth (2008) present an approach based on well selected features achieving an F-score of 80 % on a benchmark dataset. Poon and Domingos (2008) introduce an unsupervised approach based on Markov logic. Versley et al. (2008) present a highly modular toolkit for developing coreference applications. Recasens et al. (2010) report on a competition for coreference resolution involving six different languages and describe the performance of different approaches.

3.7. Word sense disambiguation

A major obstacle of using text documents by computers is the inherent polysemy: words can have more than one distinct meaning. For example, the 121 most frequent English nouns, which account for about 20 % of the words in real text, have on average 7.8 meanings each (Agirre and Edmonds, 2006). *Word sense disambiguation* (WSD) is the process of identifying the sense of a word occurring in an external *mention* of the word (Schütze, 1998). There are different variants of WSD approaches:

– *Supervised WSD* is the problem of selecting a sense for a word from a set of predefined possibilities based on annotated training data. The sense inventory usually comes from a dictionary, a thesaurus or an ontology. The solutions are based on manually constructed rules and supervised learning.
– *Unsupervised WSD* (also called Word sense discrimination (Schütze, 1998)) is the problem of dividing the usages of a word into different meanings, without regard to any particular existing sense inventory. This is especially attractive for specific domains, where the coverage of sense inventories is low.

WSD has to cope with a number of phenomena: *Homonyms* (words with the same spelling but different meaning), *synonyms* (different words with identical or very similar meaning), *hyponyms* (subconcepts), and *hypernyms* (super-concepts).

– For *supervised WSD* word senses are taken from a dictionary, taxonomy or ontology, e.g. *WordNet (Fellbaum, 1998)* an manually annotated mentions are available. Then WSD is essentially a task of classification: word senses are the classes, the context, i.e. the words in the vicinity of the target word, provides the evidence, and each occurrence of a word is assigned to one or more of its possible classes based on evidence. A specific task is the *disambiguation of proper names*.
– *Unsupervised WSD* does not require annotated training data and usually ignores pre-specified sense inventories, which may have a low coverage in specific domains. It clusters the context words of a term, e.g. using topic modeling.

Both approaches are exemplified in the following sections.

3.7.1. Supervised WSD

The approaches aim at assigning a word *mention* in some external text to a sense in a sense inventory. Different types of sense inventory are distinguished:

– Machine readable *dictionaries* with short definitions of the entries (e.g. the Oxford Advanced Learners' Dictionary (Hornby, 2010)).

– *Thesauri* with added synonym information. An example is WordNet *(Fellbaum, 1998)* covering more than 100,000 different meanings (*synsets*, concepts). It contains short definitions of meanings and short usage examples. Each meaning is assigned to a hypernym, a more comprehensive meaning.
– *Encyclopedias* like *Wikipedia* with extensive descriptions of the meaning of the different entries.
– Formal *Ontologies* containing a formal representation of the knowledge by a set of concepts within a domain and the relationships between those concepts. *DBpedia* is an ontology extracted from Wikipedia (Bizer et al., 2009). More details on ontologies are discussed in Section 4.4.

The *Lesk algorithm (Lesk, 1986)* is a very simple approach for WSD based on the assumption that words in a given neighborhood will tend to share a common topic. First a pair of ambiguous words within a neighborhood is chosen. Then all sense definitions for the two words are retrieved from a dictionary. Finally the senses are chosen which maximize the number of common terms in the definitions. A simplified version of the Lesk algorithm is to compare the dictionary definition of an ambiguous word with the terms contained of the neighborhood.

A number of approaches for WSD use classifier models. The input of the classifier for a target word x is a description of the neighborhood $N(x)$, a window of words before and after x. It is necessary to encode the words and derived features as well as their relative position to the target word x. The neighborhood $N(x)$ usually includes local collocations, parts-of-speech tags, and surrounding words as well as topics derived by topic models. Examples of classifiers are naive Bayes classifier (Cai et al., 2007), and the SVM (Chan et al., 2007) yielding good results at the SemEval07 competition, where different approaches were compared on a benchmark task. State-of-the-art systems have an accuracy of around 65 % in the Senseval-3 all-words task, or 60 % (Pradhan et al., 2007) in the SemEval-2007 task.

A major difficulty for high-performance supervised WSD is the fine granularity of available sense inventories. For WordNet with more than 100,000 different meanings annotators have a low agreement, e.g. 72.5 % agreement in the preparation of the English all-words test set at Senseval-3 (Navigli et al., 2007). Note that inter-annotator agreement is often considered an upper bound for the performance of WSD systems. Often it is not necessary to determine all the different senses for every word, but it is sufficient to distinguish the different coarse meanings of a word. It turned out that for a *coarse grained sense* inventory the inter-annotator agreements is more than 90 % Navigli et al. (2007).

A number of WSD procedures employ sequential tagging models like CRFs (see Section 3.4.2). Deschacht and Moens (2006) adapt the CRF to the large number of labels by taking into account the hypernym/hyponym relation and report a marked reduction in training time with only a limited loss in accuracy.

Ciaramita and Altun (2006) use a variant of the Hidden Markov Model, not requiring questionable independence assumptions and achieve F-scores up to 77.2 % for the coarse-grained task. Paaß and Reichartz (2009) use an *unsupervised* CRF as sequence modeling technique generating training information from the structure of WordNet. For the coarse-grained task they report a marked increase of the F-score to 84.4 %. Navigli et al. (2007) and Pradhan et al. (2007) discuss further WSD methods and results in the framework of the SemEval 2007 competition.

3.7.2. Word sense disambiguation of named entities

If there are several persons with the same name or if a persons has several names it is necessary to assign the name to the corresponding real world person. Usually an encyclopedia like Wikipedia is taken as reference information, because the description of the persons in Wikipedia may be taken to define a similarity to some external mention.

Named entity disambiguation is special as there is potentially an unbounded number of entities to be considered. Therefore it is necessary to decide, if an entity is covered at all in the reference information. Cucerzan (2007) gives an overview of supervised methods.

The approach of Cucerzan (2007) relies on a similarity function that incorporates contextual information, both extracted from the external mention of an entity to be disambiguated and Wikipedia, as well as the category tags of the candidate entities. Bunescu and Pasca (2006) use a cosine similarity and a taxonomy based feature set. This is processed by a ranking SVM, which assigns to the name mention in question the candidate names in a ranked order. The model is trained and tested on data automatically extracted from Wikipedia and achieves a disambiguation accuracy of 84.8 %. Pilz and Paaß (2009) extend this approach yielding more than 90 % F-score for the German language. Among others they uses LDA topic models as a criterion for comparing candidates to mentions.

3.7.3. Unsupervised approaches to WSD

Here the underlying assumption is that similar senses occur in similar contexts, and thus senses can be induced from text by clustering word occurrences using some measure of similarity of context. Then, new occurrences of the word can be classified into the closest induced clusters/senses. In principle any cluster technique can be applied to words and their neighborhood.

Boyd-Graber et al. (2007) give a survey on unsupervised WSD methods. They present a generative probabilistic LDA topic model for WSD where the sense of the word is a hidden random variable that is inferred from data (cf. Section 3.5). It defines a probabilistic process of word generation that is based on the hyponymy relationship in WordNet and automatically learns the context in which a word is disambiguated. Rather than disambiguating at the sentence-level or the document-level, the model uses the other words that share the same hidden topic across many documents.

The assignment of a standard `normalized` name to entities is called normalization. The Stanford Entity Resolution Framework (SERF) has the goal to provide a framework for generic entity resolution (Benjelloun et al., 2009). Other techniques for entity resolution employ relational clustering as well as probabilistic topic models Bhattacharya and Getoor (2007). They report very good F-scores and execution times on several benchmark data sets.

Yates and Etzioni (2009) investigate synonym resolution in the context of unsupervised Information Extraction without labeled training data or domain knowledge (cf. Section 4.3.2). They present the *Resolver* system using a probabilistic relational model for predicting whether two strings are co-referential based on the similarity of the statements containing them. The system is capable of processing large corpora. On a set of two million statements extracted from the Web, Resolver matches instances with 72.7 % F-score, and resolves relations with 90 % precision and 35 % recall. An extension of the basic system handles polysemous names with an F-score of 96 % on a data set from the TREC corpus.

4. Relation extraction

4.1. Introduction

In many applications there is an explicit need to relate the entities or concepts determined by information extraction. For example, it is not enough to find occurrences of product names and people names in a news article but also to identify if there is a `was invented by` relationship between a product and a person. A survey is given by Sarawagi (2009, p.55).

Relation Extraction (RE) deals with the problem of finding semantic associations between entities within a text phrase (i. e. usually a sentence). In the simplest case only binary relations are considered. Given a fixed binary relationship of type r, the goal is to extract all instances of entity pairs that have the relationship r between them. More precisely given a text snippet x and two marked entities E_1 and E_2 in x, identify if there is any relationships $r \in R$ such that $r(E_1,E_2)$. The set R of relations includes an alternative relation other if none of the predefined relations holds.

This problem has been studied extensively on natural language text, including news articles (Mitchell et al., 2003), scientific publications (Plake et al., 2006), Wikipedia (Suchanek et al., 2006b) and the general web (Banko et al., 2007). As in entity extraction, much of the impetus to relation extraction research was provided by the various competitions: starting with MUC competitions (Grishman and Sundheim, 1996) and later the ACE task (Doddington et al., 2004) and the BioCreAtIvE II Protein-Protein interaction tasks (Turmo et al., 2006).

For example the ACE task defines five top-level relations `located-at`, `near`, `part`, `role`, and social over pairs from five top level entity types `person`, `organization`, `facility`, `location`, and `geopolitical entity`. The located at relation relates `person` or `organization` with location. The role relation links `person` to an `organization`, etc. The NAGA knowledge base (Kasneci et al., 2007) comprises of 26 relationships such as `isA`, `bornInYear`, `hasWonPrize`, `locatedIn`, `politicianOf`, `actedIn`, `discoveredInYear`, `discoveredBy`, and `isCitizenOf` extracted from sources such as Wikipedia.

A number of different features may be used to solve this prediction problem:

– *Surface* terms: The terms (words) around and in-between the two entities. Often terms are stemmed to their morphological roots.
– *POS tags*: verbs in a sentence are a key to defining the relationship between entities, that are typically noun phrases.
– *Semantic annotations* as determined in Section 3.
– *Syntactic parse tree*s: A parse tree hierarchically represents the interrelation of the constituents of the sentence like noun phrases, verb phrases, etc. A *dependency graph* links each word to the words that depend on it. This is often easier to process than full syntactic parse trees.

Extracted relations can be stored and used in different ways:

– They can be added as annotations to the corresponding text where it was extracted, e.g. as XML-annotations. Section 5.1 surveys methods for retrieval.
– Another approach ignores the underlying text and enters the extracted entities and relations into an ontology. Methods for ontology retrieval are discussed in Section 5.2.

4.2. Supervised relation extraction methods

Procedures for supervised relation extraction require training data where the relation argument entities and the types of relations are manually annotated. These methods can be categorized into one of four main types:

– *Rule-based methods* that create propositional or first order logic rules over structures around two argument entities (Shen et al., 2007).

– *Feature-based methods* extract a flat set of features from the input and then invoke an off the shelf classifier.
– *Kernel-based methods* that design special kernels to capture the similarity between structures such as trees and graphs. Subsequently a kernel classifier is trained to decide if the target relation is present between the two arguments.
– *Probabilistic sequence model*: These set up a model for the dependency of the sequence of words around the two argument entities.

A key task for *feature-based approaches* is the conversion of extraction clues in structures such as sequences, trees, and graphs to a flat set of features for relation extraction. Jiang and Zhai (2007) present a systematic method of designing such features. They report that on eight relationship extraction tasks of the ACE corpus the features obtained with their approach is competitive with other more complex methods. Co-occurrence-based relation extraction is a simple, effective and popular method (Ramani et al., 2005) but usually suffers from a lower recall, since entities can co-occur for many other reasons.

4.2.1. Kernel methods

Among the first who proposed *kernel methods* (see Section 2.2.4) for relation extraction on shallow parse trees were Zelenko et al. (2003). Their kernel was then generalized to dependency trees by Culotta and Sorensen (2004). Later Bunescu and Mooney (2005) suggested a kernel for relation extraction considering the shortest path between the two entity nodes, which nevertheless yielded good performance. Besides work on tree kernels for relation extraction there have been tree kernels proposed for other tasks like question classification (Moschitti, 2006) which have shown improvements over bag-of-words approaches. Considering features associable with words Harabagiu et al. (2005) proposed to use semantic background knowledge from various sources like WordNet, FrameNet and PropBank. They enhanced the dependency tree kernel of Culotta and Sorensen (2004) and achieved good results. Reichartz et al. (2009a) formulate two different extensions of tree kernels which significantly reduce the error of previous approaches. The combination of dependency parse trees and syntactic parse trees leads to additional performance with an F-score of 81 % on the ACE-2003 benchmark data for the `role`-relation (Reichartz et al., 2009b). The approaches may be also applied to German text yielding an F-score of 77 % on a newspaper corpus for the `memberOf`-relation (Reichartz, 2011).

Girju et al. (2007) created a benchmark data set to allow the evaluation of different semantic relation classification algorithms. They chose the following semantic relations: `Cause-Effect` (e.g. `laugh wrinkles`), `Content-Container` (e.g. `the apples in the basket`), `Instrument-Agency` (e.g.

laser printer), Origin-Entity, Part-Whole, Product-Producer and Theme-Tool. Their dataset included part-of-speech tags and Word-Net sense annotations. The best systems used features based on grammatical roles, lexico-syntactic patterns, semantic parses, shallow and deep syntactic information; WordNet synsets and hypernyms. As machine learning approaches kernel methods and SVM were employed achieving an F-value of 72.4 %.

4.2.2. Probabilistic relation extraction

Relation extraction can also be considered in a *probabilistic modeling framework* evaluating statistical dependencies between terms. Bundschus et al. (2008) extend CRFs (Section 3.4.2) for the extraction of semantic relations from text. They focus on the extraction of relations between genes and diseases (five types of relations) in biomedical documents as well as between disease and treatment entities (eight types of relations). Their approach is based on a rich set of textual features. The model is quite general and can be extended to handle arbitrary biological entities and relation types. They apply their approach to a biomedical textual database and generate a network of related genes and diseases in a machine-readable *Resource Description Framework* (RDF) graph (Antoniou and Harmelen, 2008).

There have been many applications of CRFs with more general graphical structures to relation extraction (Sutton and McCallum, 2007). Finkel et al. (2005) augment the chain model with richer kinds of long-distance factors. They report improved results on several other standard benchmark data sets.

4.3. Weakly supervised relation extraction

4.3.1. Pattern-oriented relation extraction

Instead of requiring corpora with hand-tagged relations there are methods using less demanding user input. *Pattern-oriented systems* are given a target relation (e.g. personInventedProduct) and a set of instances of the target relation (person-product seed pairs like Guglielmo Marconi and radio). After submitting a query for the seed pair they find patterns in the text that appear with the seed pair (e.g. the pattern X developed Y extracted from Guglielmo Marconi developed the radio. Then they seek other occurrences of that pattern detecting new instances of the target relation. There are a large number of techniques to extract patterns from text documents, e.g. regular expressions. Most existing systems do not use deep linguistic analysis of the corpus (e.g. parsing) and small variations of the text (e.g. addition of adjectives) prevent a match of the pattern. In addition coreference resolution (Section 3.6) markedly increases the number of valid examples.

The *Leila*-system Suchanek et al. (2006b) follows this strategy extracting facts and relations from text documents. It takes into account counterexamples to identify and discard erroneous patterns. It employs different classifier methods (k-nearest neighbor and SVM) and achieves similar F-values.

The *Espresso* System (Pantel and Pennacchiotti, 2006) is a pattern-based approach taking a few input seed instances $X, Y \in \mathscr{P}$ of a particular relation $r(X,Y)$ and iteratively learns patterns to extract more instances. In a pattern induction phase they infer a set of surface patterns $\mathscr{P}$ that connects as many seed instances X, Y as possible. Then the patterns are ranked according to a reliability measure keeping only the top-ranked patterns. Subsequently the set of instances matching the patterns in the corpus are retrieved and the highest-scoring instances are selected according to a reliability measure. In the *web expansion phase* new instances of the patterns in $\mathscr{P}$ are retrieved from the Web, using the Google search engine. Each instance pair is expanded with a pattern (e.g. the instances `Italy, country` and the pattern `Y such as X` leads to a query `country such as *`). New instances are the constructed from the retrieved sentences (e.g. `Germany,country`) and added to the set. In empirical comparisons with other systems on standard and specialized relations like `is-a`, `part-of`, `succession`, `reaction`, and `production` they achieve higher and more balanced performance than other state of the art systems.

4.3.2. Open information extraction

Most effort in relation extraction has concentrated on a few general relations. In comprehensive text collections like encyclopedia and the web there exist a large number of relations not known beforehand. Etzioni et al. (2008) propose *open information extraction* where the exact meaning of the relations to be extracted are unknown, and where web-scale information extraction is necessary.

TextRunner is an Open IE system (Banko and Etzioni, 2008) working in two phases.

- TextRunner's first phase uses a CRF (Section 3.4.2) learning how to assign labels to each of the words in a sentence denoting the beginning and the end both of entity names and the relationship string.
- In the second phase an extractor rapidly extracts textual triples from sentences that aim to capture (some of) the relationships in each sentence. For example, given the sentence `Goethe, a writer born in Frankfurt, wrote The Sorrows of Young Werther`, the extractor forms the triple (`Goethe, born in, Frankfurt`). The triple comprises of three strings, in which the first and third denotes entities and the second denotes the relationship between them.

To identify synonyms for entities and relations Yates and Etzioni (2009) use the *Resolver* system neither requiring labeled training data nor domain knowledge. TextRunner was applied to more than 120 million web pages extracting more than 500 million tuples. The precision exceeded 75 % on average.

4.4. Ontology learning

An *ontology* is the description of a domain, its concepts their properties and relations by means of a formal language (Gruber, 1993). An important problem is *ontology learning* (Cimiano et al., 2005), the recognition of new concepts and relations that should go into an ontology. This includes the identification of terms, synonyms, homonyms, concepts, concept hierarchies, properties, property hierarchies, domain and range constraints and class definitions. Sometimes the individual instances of concepts and their relations are called *knowledge base* and the addition of new instances to the knowledge base is called *ontology population* (Buitelaar and Cimiano, 2008).

One distinguishes between formal ontologies, terminological ontologies and prototype-based ontologie*s* Sowa (2000a). In *prototype-based ontologies*, concepts are represented by collections of instances hierarchically arranged in subclusters (e.g. the concept `disease` defined by a set of diseases). As these ontologies are defined by instances, they have no definitions and axiomatic grounding. *Terminological ontologies* are described by concept labels and are organized into hierarchies, defined by hypernym or subclass relationships. Typical examples are WordNet (Fellbaum, 1998) and the Medical Subject Headings (MeSH) (Bundschus et al., 2008). A `disease`, for instance, is defined in WordNet as an `impairment of health or a condition of abnormal functioning`. Terminological ontologies usually have no axiomatic grounding. *A formal ontology* such as in the Web Ontology Language (OWL), in contrast, is seen as a conceptualization, whose categories are distinguished by axioms and definitions Antoniou and Harmelen (2008, p.113–156). An OWL ontology is a set of axioms which place constraints on sets of individuals (called *classes*) and the types of relationships allowed between them. The axioms may be evaluated to infer additional semantic information based on the data explicitly provided. An ontology describing `families`, for instance, can include axioms stating that a `hasFather` property is only present between two individuals when `hasParent` is also present. Formal ontologies allow to deduce new facts by reasoning, e.g. to derive conclusions for the purpose of query answering.

Most of the state-of-the-art approaches focus on learning prototype-based ontologies and far less work on learning terminological or formal ontologies (Tresp et al., 2007). Here, the big challenge is to deal with uncertain and often contradicting extracted knowledge, introduced during the ontology learning process.

Ontology learning first identifies key terms in the text (such as named entities and technical terms) by terminology extraction (see Section 3.3) and named entity recognition (see Section 3.4) and then relating them to concepts in the ontology. A frequent application is the attachment of new instances to an existing ontology. Ruiz-Casado et al. (2008) propose an architecture to enrich Wikipedia with semantic information in an automatic way. After extracting new candidate terms they learn the `is-a-relation` (X_{entity} is an instance of $Y_{wikipedia\text{-}class}$) using the hypernym hierarchy of WordNet. Specifically in a Wikipedia article about *A* they look for terms *B* such that *A* and *B* occur in the same sentence, *B* is a hypernym of *A* in WordNet and there is a link pointing to the Wikipedia article about *B*. For learning they use a pattern-based approach, where significant patterns are filtered from a large number of candidates.

YAGO is a large ontology constructed automatically (Suchanek et al., 2008) using Wikipedia as knowledge source and WordNet to define the taxonomic structure. It mainly exploits Wikipedia infoboxes as well as the category pages of Wikipedia. In addition the taxonomic information of WordNet is used to define hypernyms and hyponyms. To improve validity a type checking system investigates, if the arguments of a relation have the correct type.

The basic YAGO ontology can be extended by additional individuals, concepts and relations detected by the *Leila*-system (Suchanek, 2009). Before adding facts to YAGO they were checked by the *Sofie* system Suchanek et al. (2009). In detail it first performs a word sense disambiguation. Then Sofie finds patterns in text documents to extract facts, reasoning on the plausibility of patterns and rejecting patterns if counter evidence becomes available. Finally Sofie exploits the background knowledge from YAGO and takes into account constraints on the relations, links between hypotheses and connections to the existing knowledge. This involves the solution of a Maximum Satisfiability (MAX-SAT) problem using an approximate algorithm (Suchanek, 2009). Evaluation shows that Sofie can extract facts from arbitrary, unstructured Web documents with a precision of over 90 %.

Many reasoning-based applications of ontologies in domains such as bioinformatics or medicine rely on much more complex axiomatizations. Völker et al. (2008) present a system that is able to transform a terminological ontology to a consistent formal OWL ontology. Here, the big challenge is to deal with uncertain and often even contradicting extracted knowledge, introduced during the ontology learning process. Particularly, the treatment of logical inconsistencies becomes a great challenge as soon as we start to learn huge and expressive axiomatizations. The authors propose a reasoning-supported process to guarantee that the learned ontologies are kept consistent whenever the ontology learning procedures modify the ontology. They also suggest ways to handle different types of inconsistencies in learned ontologies, and conclude with a visionary outlook to future ontology learning and engineering environments.

5. Semantic retrieval of documents

Semantic search aims to retrieve documents containing concepts instead of documents containing character strings. To be able to include semantic annotations into search engines specific types of indices and have to be used. The aim is to search documents for keywords and in addition for semantic concepts or relations. Many approaches use techniques from *information retrieval* (IR) known from web search engines.

5.1. Retrieval of semantic annotations

Semantic retrieval approaches use IR-techniques; especially inverted indices (Salton, 1989; Baeza-Yates and Ribeiro-Neto, 1999) are compiled containing all mentions of a word in all documents. In addition the position of all annotations in documents are stored together with the annotated words. Often a standard web search engine like Lucene is enhanced to search for annotated facts and relations, e.g. by Yates and Etzioni (2009) for the TextRunner system.

5.1.1. Topic model retrieval

Topic models can be used to improve retrieval performance. Wei and Croft (2006) combined topic models with a language modeling approach. Topic descriptors determined by LDA describe the thematic area of a word and can be used for disambiguation. They found that the LDA-based document model has performance similar to pseudo-feedback retrieval. However, unlike the latter model, LDA estimation is done offline and only needs to be done once. Therefore LDA-based retrieval can potentially be used in applications where pseudo-relevance feedback would not be possible.

5.1.2. Ontology search engine

Kasneci et al. (2007) present NAGA, a semantic search engine. It builds on a graph-oriented knowledge base, which consists of facts extracted from the Web. NAGA's data model is a graph, in which nodes are labeled with entities (e.g. `Max Planck`) and edges are labeled with relationships (e.g. `bornInYear`). Each edge, together with its end nodes, represents a fact, e.g. `<Max Planck, bornInYear, 1858>` or `<Max Planck, type, physicist>`. Since these facts are derived from Web pages using possibly unreliable Information Extraction techniques, they attach a certainty value to each fact. NAGA's knowledge base is a projection of YAGO Suchanek et al. (2007). It contains about 1 million entities and 6 million facts, partly extracted by LEILA Suchanek et al. (2006b). NAGA's expressive query language can be used to formulate precise queries corresponding to the user's information need. It allows the following types of queries:

– A *discovery query* identifies entities which simultaneously fulfill a set of relations. An example is to find all `physicists` which are born in the same year as `Max Planck`. This exploits the relations `has profession` and `is born in year`. The result may be a list of entities, and – if required – the associated text documents where the relations were found.
– A *relatedness query* looks for all relations between specific entities. An example are all relations which `Max Planck` and `Mihaijlo Pupin` have in common, e.g. `has profession physicist` and `born in year 1858`. The result may be a list of relations and their entity arguments, and – if required – the associated text documents where the relations were found.

The query results are ranked according to a scoring mechanism that takes their certainty values into account. The numbers on the edges of an answer graph represent the certainties of the facts. In order to compute the overall certainty of an answer, the certainties of the edges have to be accumulated.

Codina et al. (2008) discuss a semantic search engine for patent search. Documents are first processed with general purpose language processing tools, which carry out POS tagging, multiword recognition, lemmatization, and dependency parsing. Linguistic annotations are then exploited to recognize frame instances (see FrameNet (Ruppenhofer et al., 2005)), and finally concepts and triples. As ontological framework the Suggested Upper Merged Ontology (SUMO) with mappings to WordNet and several patent domain ontologies are used. In triple-based semantic search the user specifies a target triple by selecting a relation and two concepts filling the subject and object role of the relation. The relation is chosen from a small set of a few tens of significant relations recognized by the system (e.g. hasPart, moves, hasSubstance). Subject and object are selected from a much larger set (over 30.000) of domain specific concepts.

In its basic functionality, the search engine selects all sentences in the corpus containing the target triple, whatever the linguistic form in which the triple is expressed (e.g. `An optical head has a prism or the prism included in the optical head or the prism of the optical head`). However, the user can also choose to expand the search by instructing the system to consider concepts related to the object or subject of the target relation. Alternatively, the user can search for concepts generalizing the concept of `prism`, like `optical_component` (hypernyms). The retrieved sentences can be ordered according to their similarity to the base (non-expanded) target triple. The semantic distance between the target and the retrieved triples is measured according to the length of their connecting path in the ontology.

5.1.3. XML-retrieval

Semantic content can be stored together with the text of the documents using *eXtensible Markup Language* (XML). XML queries are also called *content and structure* (CAS) queries as they enable the user to specify keywords as well as structural and annotation properties (Manning et al., 2008, p.178–200).

The *XQuery* language supports structural queries but only looks for exact matches. Usually, however, also partial matches are important, where the returned documents are sorted by relevance. A survey including first solutions for extending the classical vector space model for structural retrieval is provided by Pal and Mitra (2007). The aim is to find a text or XML-element that is highly relevant. This involves ranking of different XML-elements and the ranking of returned documents. The result may not always be entire documents but nested XML elements. The Initiative for the Evaluation of XML-Retrieval (INEX) provides a platform for evaluating such search algorithms (Fuhr et al., 2002).

5.2. Ontology retrieval

An ontology stores extracted facts and relations as a network of interrelated concepts and entities, e.g. using RDF (Resource Description Framework) and OWL (Web Ontology Language). It usually ignores the underlying text, where the information was extracted (Antoniou and Harmelen, 2008).

SPARQL (SPARQL Protocol and RDF Query Language) is a syntactically-SQL-like language for querying RDF graphs via pattern matching (Saarela, 2007). The language's features include basic conjunctive patterns, value filters, optional patterns, and pattern disjunction. SPARQL queries RDF graphs, which consist of various triples expressing binary relations between resources, by specifying a subgraph with certain resources replaced by variables. Figure 3 shows a SPARQL query returning all country capitals in Asia. This means, SPARQL does return facts and relations, not the documents containing these entities.

```
PREFIX abc: <http://example.com/exampleOntology#>
SELECT ?capital ?country
WHERE {
?x abc:cityname ?capital;
abc:isCapitalOf ?y
?y abc:countryname ?country;
abc:isInContinent abc:Asia
}
```

Figure 3. SPARQL query returning all country capitals in Asia.

SPARQL can also be used to query OWL ontologies. Note that for some queries OWL ontologies require reasoning. The open-source OWL-DL reasoner, Pellet, answers SPARQL queries while considering OWL-DL entailments (Sirin et al., 2010).

Bizer et al. (2009) describe the DBpedia ontology extracted from Wikipedia and represented as a RDF graph. The resulting DBpedia knowledge base at the moment (Jan. 2011) describes more than 3.5 million entities, including 364,000 persons, 462,000 places, 99,000 music albums, 54,000 Films, and 148,000 organizations and contains 672 million RDF triples. They can be queried using a SPARQL interface (http://dbpedia.org/sparql).

This information has mainly been extracted from Wikipedia infoboxes. DBpedia uses English Wikipedia article names for creating unique identifiers. Information from other language versions of Wikipedia is mapped to these identifiers by evaluating the interlanguage links between Wikipedia articles. As Wikipedia attributes do not have explicitly defined datatypes, a further problem is the relatively high error rate of the heuristics that are used to determine the datatypes of attribute values. The current implementation does not consider these uncertainties.

5.3. Multimodal search

Multimodal search refers to the simultaneous retrieval of contents from different media often using different search techniques. The combination of different search paradigm can improve the results in comparison to separate retrieval. In the German *CONTENTUS* project Paaß et al. (2009a) semantically annotate different media like newspaper documents, music pieces, and video news stories. Recorded speech in videos is transcribed to text by speech recognition. Named entities and other terminology are extracted from text or transcribed speech (Paaß et al., 2009b) and matched to the corresponding Wikipedia articles. For music the titles, artists, lyrics, and other metadata is determined. Semantic relations like `bornIn` are identified. The text and the media metadata together with the annotations are indexed in a Lucene index. Finally a search GUI is provided to search text and media metadata and to present text, video and music to the user. The ultimate target is to make the content of the *German National Library* (DNB) accessible in this way.

5.4. Conclusion

In this chapter we described a number of approaches to extract semantic information from text to allow more focussed retrieval. We considered different units of semantic information like classes, named entities, topics, concepts, relations or even complete ontologies. The algorithms for the automatic extraction of these units are usually based on machine learning techniques exploiting text documents as training examples.

For simple types of semantic information, e.g. part-of-speech tagging and named entity recognition, there exist mature approaches which extract the target information with high reliability. For more advanced pieces of information, e.g. relation extraction and ontology learning, the results currently do not yield the quality level necessary for many business application.

To reduce the effort for annotating training data there is a strong trend to weakly supervised or unsupervised machine learning approaches. At the same time many existing resources like WordNet or Wikipedia are exploited. An interesting example is the *Never-Ending Language Learner* (NELL) implemented by Carlson et al. (2010). Starting with an ontology of 500 concepts and relations and with 10–20 seed samples for each this computer program continuously extracts new facts and relations from the web to enhance the ontology. Using similar strategies as TextRunner (Section 4.3.2)it aims at gradually improving its learning strategy, currently achieving a precision of 74 %. NELL is an example of a *Machine Reading* approach aiming at the automatic unsupervised understanding of text. There is a very active research community on this topic and we can expect major improvements in the next years.

Increasingly more semantic techniques are used in real-world applications. Pang and Lee (2008) use machine learning techniques to analyze opinions in online reviews and personal blogs. They focus on methods addressing sentiment-aware applications. Mitra and Mitra (2011) describe how semantic information extracted from newspapers, blogs or fora are used in financial trading. Although information collected from blogs and fora are less reliable than established media they reflect the “collective intelligence” of many players. The German THESEUS research program actively develops and evaluates semantic technologies for applications like corporate control systems, semantic retrieval in multimedia libraries or semantic analysis of business-to-business communication (BMWi, 2011). It is likely that in the next years semantic analyses will get more and more important for economic applications.

Acknowledgments

The work presented here was funded by the German Federal Ministry of Economy and Technology (BMWi) under the THESEUS project.

References

Agirre, Eneko and Phillip Edmonds
2006 Introduction. In *Word Sense Disambiguation: Algorithms and Applications*, pages 1–28, Berlin/New York. Springer.

Allan, James, Stephen Harding, David Fisher, Alvaro Bolivar, Sergio Guzman-Lara and Peter Amstutz
2005 Taking topic detection from evaluation to practice. In *Proceedings of the Proceedings of the 38th Annual Hawaii International Conference on System Sciences (HICSS'05) – Track 4 – Volume 04*.

Antoniou, Grigoris and Frank van Harmelen
2008 *A Semantic Web Primer*. The MIT Press, Cambridge, MA, USA, 2 edition.

Baeza-Yates, R. and B. Ribeiro-Neto
1999 *Modern Information Retrieval*. Addison Wesley, New York.

Bagga, Amit and Breck Baldwin
1998 Entity-based cross-document coreferencing using the vector space model. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics*, pages 79–85, San Francisco, California.

Banko, M., M.J. Cafarella, S. Soderland, M. Broadhead and O. Etzioni
2007 Open information extraction from the web. In *Proceedings of IJCAI*, pages 2670–2676.

Banko, Michele and Oren Etzioni
2008 The tradeoffs between open and traditional relation extraction. In *Proceedings of ACL-08: HLT*, pages 28–36.

Bengio, Yoshua and Yves Grandvalet
2004 No unbiased estimator of the variance of *k*-fold cross-validation. *Journal of Machine Learning Research*, 5:1089–1105.

Bengtson, Eric and Dan Roth
2008 Understanding the value of features for coreference resolution. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP '08, pages 294–303, Stroudsburg, PA, USA. Association for Computational Linguistics.

Benjelloun, Omar, Hector Garcia-Molina, David Menestrina, Qi Su, Steven E. Whang and Jennifer Widom
2009 Swoosh: A generic approach to entity resolution. *The VLDB Journal*, 18:255–276.

Berners-Lee, Tim, James Hendler and Ora Lassila
2001 The semantic web. *Scientific American*, 284:34–43.

Bhattacharya, Indrajit and Lise Getoor
2007 Collective entity resolution in relational data. *ACM Transactions on Knowledge Discovery from Data*, 1.

Bizer, Christian, Jens Lehmann, Georgi Kobilarov, Sören Auer, Christian Becker, Richard Cyganiak and Sebastian Hellmann
2009 DBpedia – a crystallization point for the web of data. *Web Semantics: Science, Services and Agents on the World Wide Web*, 7:154–165.

Blei, D. and J. Lafferty
2009 Topic models. In *Text Mining: Theory and Applications*. Taylor and Francis, London.

Blei, David M., Andrew Y. Ng and Michael I. Jordan
2003 Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022.

BMWi
2011 The THESEUS research program. URL http://theseus-programm.de/en-US/home/. Retrieved on April 10th, 2011.

Boyd-Graber, Jordan, David Blei and Xiaojin Zhu
2007 A topic model for word sense disambiguation. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 1024–1033.

Brants, Thorsten
2000 TnT: a statistical part-of-speech tagger. In *Proceedings of the sixth conference on Applied natural language processing*, ANLC '00, pages 224–231, Stroudsburg, PA, USA. Association for Computational Linguistics.

Breiman, Leo, J. H. Friedman, R. A. Olshen and C. J. Stone
1984 *Classification and Regression Trees*. Wadsworth, Belmont, CA.

Buitelaar, Paul and Philipp Cimiano (eds.)
2008 *Ontology Learning and Population: Bridging the Gap between Text and Knowledge*. IOS Press, Amsterdam.

Bundschus, Markus, Mathaeus Dejori, Martin Stetter, Volker Tresp and Hans-Peter Kriegel
2008 Extraction of semantic biomedical relations from text using conditional random fields. *BMC Bioinformatics*, 9(207).

Bunescu, Razvan C. and Raymond J. Mooney
2005 A shortest path dependency kernel for relation extraction. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 724–731.

Bunescu, Razvan C. and Marius Pasca
2006 Using encyclopedic knowledge for named entity disambiguation. In *Proceedings of EACL*, pages 9–16.

Cai, Jun Fu, Wee Sun Lee and Yee Whye Teh
2007 NUS-ML: Improving word sense disambiguation using topic features. In *Proceedings of SemEval-2007*, pages 249–252.

Carlson, Andrew, Justin Betteridge, Bryan Kisiel, Burr Settles, Estevam R. Hruschka Jr. and Tom M. Mitchell
2010 Toward an architecture for neverending language learning. In *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence (AAAI-10)*.

Carpineto, Claudio, Stanisiaw Osinski, Giovanni Romano and Dawid Weiss
2009 A survey of web cluster engines. *ACM Computing Surveys*, 41:17:1-17:38.

Caruana, Rich and Alexandru Niculescu-Mizil
2006 An empirical comparison of supervised learning algorithms. In *Proceedings of the 23rd international conference on Machine learning*, ICML '06, pages 161–168, New York, NY, USA. ACM.

Cesa-Bianchi, Nicolò, Claudio Gentile and Luca Zaniboni
2006 Hierarchical classification: Combining Bayes with SVM. In *Proceedings of the 23rd international conference on Machine learning*, ICML '06, pages 177–184, New York, NY, USA. ACM.

Chan, Yee Seng, Hwee Tou Ng and Zhi Zhong
2007 NUS-PT: Exploiting parallel texts for word sense disambiguation in the english all-words tasks. In *Proceedings of SemEval-2007*, pages 253–256.

Chinchor, N.
1997 MUC-7 named entity task definition version 3.5. Technical report, NIST, ftp.muc.saic.com/pub/MUC/MUC7-guidelines.

Ciaramita, M. and Y. Altun
2006 Broad-coverage sense disambiguation and information extraction with a supersense sequence tagger. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, EMNLP '06, pages 594–602, Stroudsburg, PA, USA. Association for Computational Linguistics.

Cimiano, Philipp, Andreas Hotho and Steffen Staab
2005 Learning concept hierarchies from text corpora using formal concept analysis. *Journal of Artificial Intelligence Research*, 24:305–339.

Clark, H.
1977 Bridging. In P. N. Johnson-Laird and P. C. Wason, editors, *Thinking: Readings in Cognitive Science*, pages 411–420. Cambridge University Press, Cambridge.

Codina, Joan, Emanuele Pianta, Stefanos Vrochidis and Symeon Papadopoulos
2008 Integration of semantic, metadata and image search engines with a text search engine for patent retrieval. In *Proceedings of SemSearch 2008*.

Cover, T. M. and P. E. Hart
1967 Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13:21–27.

Cucerzan, Silviu
2007 Large-scale named entity disambiguation based on Wikipedia data. In *Proceedings of 2007 Joint Conference on EMNLP and CNLL*, pages 708–716.

Culotta, Aron and Jeffrey Sorensen
2004 Dependency tree kernels for relation extraction. In *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, ACL '04, Stroudsburg, PA, USA. Association for Computational Linguistics.

Daille, B., E. Gaussier and J. Lang
1994 Towards automatic extraction of monolingual and bilingual terminology. In *Proceedings of COLING 94*, pages 515–521.

de Marneffe, Marie-Catherine, Bill MacCartney and Christopher D. Manning
2006 Generating typed dependency parses from phrase structure parses. In *5th International Conference on Language Resources and Evaluation (LREC 2006)*, pages 449–454.

Dempster, A. P., N. M. Laird and D.B. Rubin
1977 Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, 39(1):1-38.

Deschacht, Koen and Marie-Francine Moens
2006 Efficient hierarchical entity classifier using conditional random fields. In *Proceedings of 2nd Workshop on Ontology Learning and Population*, pages 33–40.

Diewald, Niels
2012 Decentralized social network sites. In Alexander Mehler, Laurent Romary, and Dafydd Gibbon, editors, *Handbook of Technical Communication*. De Gruyter, Berlin/Boston.

Doddington, G., A. Mitchell, M. Przybocki, L. Ramshaw, S. Strassel and R. Weischedel
2004 Automatic Content Extraction (ACE) program – task definitions and performance measures. In *LREC 2004: Fourth International Conference on Language Resources and Evaluation*.

Etzioni, Oren, Michele Banko, Stephen Soderland and Daniel S. Weld
2008 Open information extraction from the web. *Communication of the ACM*, 51:68–74.

Evert, Stefan
2008 Corpora and collocations. In Anke Ludeling and Merja Kyto, editors, *Corpus Linguistics. An International Handbook of the Science of Language and Society*, pages 1212–1248. Mouton de Gruyter, Berlin/New York.

Federal Ministry of Economics and Technology
2009 ID2010. Germany: building the information society. http://www.bmwi.de/English/Navigation/Technology-policy/the-information-society,did=79428.html. Retrieved on February 22

Feldman, Ronen and James Sanger
2007 *The Text Mining Handbook – Advanced Approaches in Analysing Unstructured data*. Cambridge University Press, Cambridge.

Fellbaum, Christine
1998 *WordNet: An Electronic Lexical database*. The MIT Press, Cambridge, MA, USA.

Finkel, J., T. Grenager and C. D. Manning
2005 Incorporating non-local information into information extraction systems by Gibbs sampling. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, ACL '05, pages 363–370, Stroudsburg, PA, USA. Association for Computational Linguistics.

Fuhr, Norbert, Norbert Gövert, Gabriella Kazai and Mounia Lalmas
2002 INEX: INitiative for the Evaluation of XML retrieval. In *Proceedings of SIGIR*.

Gärtner, Thomas
2003 A survey of kernels for structured data. *SIGKDD Explorations*, 5:49–58.

Girju, Roxana, Preslav Nakov, Vivi Nastase, Stan Szpakowicz, Peter Turney and Deniz Yuret
2007 SemEval-2007 task 04: Classification of semantic relations between nominals. In *Proceedings of SemEval-2007*.

Goyvaerts, Jan and Steven Levithan
2009 *Regular Expressions Cookbook*. O'Reilly, Sebastopol, CA.

Griffiths, T. and M. Steyvers
2004 Finding scientific topics. *Proceedings of the National Academy of Sciences*, 101(suppl. 1):5228–5235.

Grishman, Ralph and Beth Sundheim
1996 Message understanding conference – 6: A brief history. In *Proceedings of the 16th International Conference on Computational Linguistics (COLING)*, pages 466–471.

Gruber, Thomas R.
1993 A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5:199–220.

Halliday, Michael A. K. and Ruqaiya Hasan
1976 *Cohesion in English*. Longman, London.

Halliday, Michael A. K. and Ruqaiya Hasan
2000 System and text: Making links. *Text*, 20(2):201–210.

Harabagiu, Sanda, Cosmin Adrian Bejan and Paul Morarescu
2005 Shallow semantics for relation extraction. In *Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence (IJCAI '05)*.

Hastie, T., R. Tibshirani and J. Friedman
2001 *The Elements of Statistical Learning*. Springer, Berlin/New York.

Hofmann, T.
2001 Unsupervised learning by probabilistic latent semantic analysis. *Machine Learning*, 42:177–196.

Hornby, A. S. (ed.)
2010 *Oxford Advanced Learners' Dictionary*. Oxford University Press, Oxford.

Hotho, A., S. Staab and G. Stumme
2003 Ontologies improve text document clustering. In *Proceedings of the IEEE International Conference on Data Mining (ICDM 03)*, pages 541–544.

Hotho, Andreas, Andreas Nürnberger and Gerhard Paaß
2005 A brief survey of text mining. *LDV Forum*, 20(1):19–62.

Jiang, J. and C. Zhai
2007 A systematic exploration of the feature space for relation extraction. In *Proceedings of the Annual Conference of the North American Chapter of the Association for Computational Linguistics Rochester*, pages 113–120, Stroudsburg, PA, USA. Association for Computational Linguistics.

Jing, Liping, Lixin Zhou, Michael K. Ng and Joshua Zhexue Huand
2006 Ontology-based distance measure for text clustering. In *SIAM Workshop on Text Mining*.

Joachims, Thorsten
1998 Text categorization with support vector machines: learning with many relevant features. In *Proceedings of ECML*, pages 137–142, Berlin/New York. Springer.

Joachims, Thorsten
2002 Optimizing search engines using clickthrough data. In *Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2002)*.

Kasneci, Gjergji, Fabian M. Suchanek, Maya Ramanath and Gerhard Weikum
2007 How NAGA uncoils: Searching with entities and relations. In *Proceedings of the 16th international conference on World Wide Web*, WWW '07, pages 1167–1168, New York, NY, USA. ACM.

Kim, J., T. Ohta, Y. Tsuruoka, Y. Tateisi and N. Collier
2004 Introduction to the bio-entity task at JNLPBA. In N. Collier, P. Ruch, and A. Nazarenko, editors, *Proceedings of Workshop on Natural Language Processing in Biomedicine and its Applications*, pages 70–76.

Kim, Su Nam, Olena Medelyan, Min-Yen Kan and Timothy Baldwin
2010 SemEval-2010 task 5: Automatic keyphrase extraction from scientific articles. In *SemEval 2010: 5th InternationalWorkshop on Semantic Evaluation*.

Konya, Iuliu Vasile, Christoph Seibert, Sebastian Glahn and Stefan Eickeler
2008 A robust front page detection algorithm for large periodical collections. In *Proceedings of ICPR 2008*, pages 1–5.

Lafferty, John, Andrew McCallum and Fernando Pereira
2001 Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of 18th International Conference on Machine Learning*, pages 282–289.

Lesk, Michael
1986 Automatic sense disambiguation using machine readable dictionaries: how to tell a pine cone from an ice cream cone. In *SIGDOC '86: Proceedings of the 5th annual international conference on Systems documentation*, pages 24–26, New York, NY, USA. ACM.

Liu, Tie-Yan, Yiming Yang, Hao Wan, Hua-Jun Zeng, Zheng Chen and Wei-Ying Ma and
2005 Support vector machines classification with a very lage-scale taxonomy. *SIGKKK Explorations*, 7:36–43.

Manning, C. D., P. Raghavan and H. Schütze
2008 *Introduction to Information Retrieval*. Cambridge University Press, Cambridge.

McCallum, Andrew
2003 Efficiently inducing features of conditional random fields. In *Proceedings of the 19th Conference on Uncertainty in Artificial Intelligence (UAI)*, page 403410.

Menke, Peter
2012 Evaluation of technical communication. In Alexander Mehler, Laurent Romary, and Dafydd Gibbon, editors, *Handbook of Technical Communication*. De Gruyter, Berlin/Boston.

Mika, Peter
2007 Ontologies are us: A unified model of social networks and semantics. *Journal of Web Semantics*, 5(1):5–15.

Mitchell, Alexis, Stephanie Strassel, Mark Przybocki, JK Davis, George Doddington, Ralph Grishman, Adam Meyers, Ada Brunstein, Lisa Ferro and Beth Sundheim
2003 ACE-2 Version 1.0; corpus LDC2003T11. Linguistic Data Consortium, Philadelphia. http://www.ldc.upenn.edu.

Mitchell, Tom
1997 *Machine Learning*. McGraw-Hill, New York.

Mitra, Gautam and Leela Mitra (eds.)
2011 *The Handbook of News Analytics in Finance*. Wiley, New York.

Moschitti, Alessandro
2006 Efficient convolution kernels for dependency and constituent syntactic trees. In *Proceedings of the ECML*, pages 318–329.

Nadeau, David and Satoshi Sekine
2007 A survey of named entity recognition and classification. *Lingvisticae Investigationes*, 30(1):3–26.

Navigli, Roberto, Kenneth Litkowski and Orin Hargraves
2007 SemEval-2007 task 07: Coarse-grained english all-words task. In *Proceedings of Workshop on Semantic Evaluations (SemEval-2007)*, pages 30–35.

Nguyen, Minh Hoai and Fernando De la Torre
2010 Optimal feature selection for support vector machines. *Pattern Recognition*, 43:584–591.

Nigam, Kamal, Andrew McCallum, Sebastian Thrun and Tom Mitchell
2000 Text classification from labeled and unlabeled documents using EM. *Machine Learning*, 39:103–134.

Osinski, Stanislaw and Dawid Weiss
2005 Carrot2: Design of a flexible and efficient web information retrieval framework. In *Proceedings of the third International Atlantic Web Intelligence Conference (AWIC 2005)*, pages 439–444.

Paaß, Gerhard and Frank Reichartz
2009 Exploiting semantic constraints for estimating supersenses with CRFs. In *Proceedings of the SIAM International Conference on Data Mining, SDM 2009, April 30 – May 2, 2009, Sparks, Nevada, USA*, pages 485–496. SIAM.

Paaß, Gerhard, Jörg Kindermann and Edda Leopold
2004 Learning prototype ontologies by hierachical latent semantic analysis. In *Proceedings of ECML/PKDD 2004 Workshop on Knowledge Discovery and Ontologies*, pages 49–60.

Paaß, Gerhard, Stefan Eickeler and Stefan Wrobel
2009a Text mining and multimedia search in a large content repository. In *Proceedings of Sabre Conference on Text Minng Services (TMS '09) 2009, Leipzig*, pages 15–22.

Paaß, Gerhard, Anja Pilz and Jochen Schwenninger
2009b Named entity recognition of spoken documents using subword units. In *Proceedings of Third IEEE International Conference on Semantic Computing Berkeley, CA, USA – Sept. 2009*, pages 529–534.

Pal, Sukomal and Mandar Mitra
2007 XML retrieval: A survey, technical report, cvpr, 2007. Technical Report TR/ISI/CVPR/IR07-01, Computer Vision and Pattern Recognition Unit, Indian Statistical Institute, Kolkata.

Pang, Bo and Lillian Lee
2008 Opinion mining and sentiment analysis. *Journal Foundations and Trends in Information Retrieval*, 2:1–135.

Pantel, Patrick and Marco Pennacchiotti
2006 Espresso: leveraging generic patterns for automatically harvesting semantic relations. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics Year of Publication: 2006*, pages 113–120.

Pazienza, Maria Teresa, Marco Pennacchiotti and Fabio Massimo Zanzotto
2005 Terminology extraction: an analysis of linguistic and statistical approaches. In *Knowledge Mining. Proceedings of the NEMIS 2004 Final Conference*. Springer, Berlin/New York.

Pilz, Anja and Gerhard Paaß
2009 Named entity resolution using automatically extracted semantic information. In *Proceedings of the Workshop on Knowledge Discovery, Data Mining and Machine Learning 2009*, pages 84–91.
Plake, C., T. Schiemann, M. Pankalla, J. Hakenberg and U. Leser
2006 Alibaba: Pubmed as a graph. *Bioinformatics*, 22:2444–2445.
Poon, Hoifung and Pedro Domingos
2008 Joint unsupervised coreference resolution with Markov logic. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, pages 650–659.
Porter, Martin F.
1980 An algorithm for suffix stripping. *Program*, 14(3):130–137.
Pradhan, Sameer S., Edward Loper, Dmitriy Dligach and Martha Palmer
2007 SemEval-2007 task 17: English lexical sample, SRL and all words. In *Proceedings of the Workshop on Semantic Evaluations (SemEval-2007)*, pages 87–92.
Rahman, Altaf and Vincent Ng
2009 Supervised models for coreference resolution. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pages 968–977.
Rahtz, Sebastian
2012 Representation of documents in technical communication. In Alexander Mehler, Laurent Romary, and Dafydd Gibbon, editors, *Handbook of Technical Communication*. De Gruyter, Berlin/Boston.
Ramani, A. K., R. C. Bunescu, R. J. Mooney and E. M. Marcotte
2005 Consolidating the set of known human protein-protein interactions in preparation for large-scale mapping of the human interactome. *Genome Biology*, 6(R40). doi: 10.1186/gb-2005-6-5-R40.
Recasens, Marta, Lluis Marquez, Emili Sapena, M. Antonia Marti, Mariona Taulé, Véronique Hoste, Massimo Poesio and Yannick Versley
2010 SemEval-2010 task 1: Coreference resolution in multiple languages. In *SemEval 2010: 5th InternationalWorkshop on Semantic Evaluation*.
Reichartz, Frank
2011 *Automatic Relation Extraction from Text*. PhD thesis, Mathematisch-Naturwissenschaftliche Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn.
Reichartz, Frank, Hannes Korte and Gerhard Paaß
2009a Dependency tree kernels for relation extraction from natural language text. In *Proceedings of the European Conference on Machine Learning, ECML 2009*, pages 270–285.
Reichartz, Frank, Hannes Korte and Gerhard Paaß
2009b Composite kernels for relation extraction. In *Proceedings of ACL-IJCNLP 2009*, pages 365–368.
Reichartz, Frank, Hannes Korte and Gerhard Paaß
2010 Semantic relation extraction with kernels over typed dependency trees. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining. Washington, DC, USA*, pages 773–782.

Ruiz-Casado, Maria, Enrique Alfonseca, Manabu Okumura and Pablo Castells
2008 Information extraction and semantic annotation of Wikipedia. In *Proceeding of the 2008 Conference on Ontology Learning and Population: Bridging the Gap between Text and Knowledge*, pages 145–169, Amsterdam, The Netherlands, The Netherlands. IOS Press.

Ruppenhofer, Josef, Michael Ellsworth, Miriam R. L. Petruck and Christopher R. Johnson
2005 FrameNet II: extended theory and practice. Technical report, ICSI, Berkeley.

Saarela, Janne
2007 *The SPARQL Handbook*. Academic Press, New York.

Salton, G. and C. Buckley
1988 Term weighting approaches in automatic text retrieval. *Information Processing & Management*, 24(5):513–523.

Salton, Gerard
1989 *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer*. Addison Wesley, Reading, Massachusetts.

Tjong Kim Sang, Erik F. and Fien De Meulder
2003 Introduction to the CoNLL-2003 shared task: language-independent named entity recognition. In *Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003*, pages 142–147, Morristown, NJ, USA. Association for Computational Linguistics.

Sarawagi, Sunita
2009 *Information Extraction*. Now Publishers Inc., Boston.

Schmid, Helmut
2010 Decision trees. In Alexander Clark, Chris Fox, and Shalom Lappin, editors, *The Handbook of Computational Linguistics and Natural Language Processing*, chapter 7, pages 180–196. Wiley-Blackwell, Chichester.

Schütze, Hinrich
1998 Automatic word sense discrimination. *Computational Linguistics*, 24:97–123.

Sebastiani, Fabrizio
2002 Machine learning in automated text categorization. *ACM Computing Surveys*, 34:1–47.

Sha, Fei and Fernando Pereira
2003 Shallow parsing with conditional random fields. Technical report, CIS TR MS-CIS-02-35, University of Pennsylvania.

Shadbolt, Nigel, Tim Berners-Lee and Wendy Hall
2006 The semantic web revisited. *IEEE Intelligent Systems*, 21:96–101.

Sharoff, Serge and Anthony Hartley
2012 Lexicography, terminology and ontologies. In Alexander Mehler, Laurent Romary, and Dafydd Gibbon, editors, *Handbook of Technical Communication*. De Gruyter, Berlin/Boston.

Shawe-Taylor, John and Nello Cristianini
2004 *Kernel Methods for Pattern Analysis*. Cambridge University Press, Cambridge.

Shen, W., A. Doan, J.F. Naughton and R. Ramakrishnan
2007 Declarative information extraction using datalog with embedded extraction predicates. In *Proceedings of VLDB*, pages 1033–1044.

Sirin, Evren, Bijan Parsia, Bernardo Cuenca Grau, Aditya Kalyanpur and Yarden Katz
2010 Pellet: A practical OWL-DL reasoner. Technical report, Department of Computer Science, University of Maryland, College Park, MD.

Sowa, John F.
2000a Ontology, metadata, and semiotics. In *Proceedings of the Linguistic on Conceptual Structures: Logical Linguistic, and Computational Issues*, pages 55–81, London, UK. Springer.

Sowa, John F.
2000b *Knowledge Representation: Logical, Philosophical, and Computational Foundations*. Brooks/Cole, Pacific Grove.

Steinberger, Ralf, Bruno Pouliquen and Erik van der Goot
2009 An introduction to the Europe Media Monitor family of applications. In *Proceedings of the SIGIR Workshop Proceedings Information Access in a Multilingual World*.

Steyvers, M. and T. Griffiths
2006 Probabilistic topic models. In T. Landauer, D. McNamara, S. Dennis, and W. Kintsch, editors, *Handbook of Latent Semantic Analysis*, pages 427–448. Laurence Erlbaum, Mahwah, New Jersey.

Stührenberg, Maik
2012 Foundations of markup languages. In Alexander Mehler, Laurent Romary, and Dafydd Gibbon, editors, *Handbook of Technical Communication*. De Gruyter, Berlin/Boston.

Suchanek, F. M., G. Ifrim and G. Weikum
2006a Combining linguistic and statistical analysis to extract relations from web documents. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '06, pages 712–717, New York, NY, USA. ACM.

Suchanek, F. M., G. Kasneci and G. Weikum
2008 YAGO – a large ontology from Wikipedia and WordNet. *Journal of Web Semantics*, 6:203–217.

Suchanek, Fabian M.
2009 *Automated Construction and Growth of a Large Ontology*. PhD thesis, Faculties of Natural Sciences and Technology of Saarland University Saarbrücken, Germany.

Suchanek, Fabian M., Georgiana Ifrim and Gerhard Weikum
2006b Combining linguistic and statistical analysis to extract relations from web documents. Technical Report MPI-I2006-5-004, Max-Planck-Institut für Informatik, Saarbrücken.

Suchanek, Fabian M., Gjergji Kasneci and Gerhard Weikum
2007 YAGO: A core of semantic knowledge. In *Proceedings of the 16th international conference on World Wide Web*, WWW '07, pages 697–706, New York, NY, USA. ACM.

Suchanek, Fabian M., Mauro Sozio and Gerhard Weikum
2009 SOFIE: a self-organizing framework for information extraction. In *Proceedings of the 18th international conference on World wide web*, WWW '09, pages 631–640, New York, NY, USA. ACM.

Sutton, Charles and Andrew McCallum
2007 An introduction to conditional random fields for relational learning. In Lise Getoor and Ben Taskar, editors, *Introduction to Relational Statistical Learning*. The MIT Press, Boston.

Teh, Y., M. Jordan, M. Beal and D. Blei
2007 Hierarchical Dirichlet processes. *Journal of the American Statistical Association*, 101:1566–1581.

Tresp, Volker, Yi Huang, Emanuele Della Valle, Daniele Braga, Davide Barbieri and Stefano Ceri
2007 Survey of relevant literature on abstraction and learning. Technical report, FP7 LarKC project.

Tsochantaridis, Ioannis, Thorsten Joachims, Thomas Hofmann and Yasemin Altun
2005 Large margin methods for structured and interdependent output variables. *Journal of Machine Learning Research*, 6:1453–1484.

Turmo, J., A. Ageno and N. Catala
2006 Adaptive information extraction. *ACM Computer Services*, 38.

van Rijsbergen, Cornelis J.
1975 *Information Retrieval*. Butterworths, London/Boston.

Vapnik, Vladimir
1995 *The Nature of Statistical Learning Theory*. Springer, Berlin/New York.

Versley, Yannick, Simone Paolo Ponzetto, Massimo Poesio, Vladimir Eidelman, Alan Jern, Jason Smith, Xiaofeng Yang and Alessandro Moschitti
2008 Bart: A modular toolkit for coreference resolution. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies: Demo Session*.

Vieira, Renata and Simone Teufel
1997 Towards resolution of bridging descriptions. In *Proceedings of the 35th Joint Meeting of the Association for Computational Linguistics*, pages 7–12.

Völker, Johanna, Peter Haase and Pascal Hitzler
2008 Learning expressive ontologies. In Paul Buitelaar and Philipp Cimiano, editors, *Ontology Learning and Population: Bridging the Gap between Text and Knowledge*, pages 45–69. IOS Press, Amsterdam.

Walmsley, Priscilla
2002 *Definitive XML Schema*. Prentice Hall, Upper Saddle River.

Wei, X. and B. Croft
2006 LDA-based document models for ad-hoc retrieval. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR ’06, pages 178–185, New York, NY, USA. ACM.

Whitelaw, Casey, Alex Kehlenbeck, Nemanja Petrovic and Lyle Ungar
2008 Web-scale named entity recognition. In *Proceeding of the 17th ACM conference on Information and knowledge management*, CIKM ’08, pages 123–132, New York, NY, USA. ACM.

Witschel, Hans-Friedrich
2005 Terminology extraction and automatic indexing – comparison and qualitative evaluation of methods. In *Proceedings of Terminology and Knowledge Engineering (TKE)*.

Wong, Wilson, Wei Liu and Mohammed Bennamoun
2007 Determining termhood for learning domain ontologies in a probabilistic framework. In *Proceedings of the Sixth Australasian Data Mining Conference (AusDM 2007)*.
Yates, Alexander and Oren Etzioni
2009 Unsupervised methods for determining object and relation synonyms on the web. *Journal of Artificial Intelligence Research (JAIR)*, 34:255–296.
Zelenko, Dmitry, Chinatsu Aone and Anthony Richardella
2003 Kernel methods for relation extraction. *Journal of Machine Learning Research*, 3:1083–1106.

6. Multimodal and Speech Technology

Jean-Claude Martin and Tanja Schultz

This chapter introduces into technologies for speech processing and multimodal interfaces. 'Multimodal' stands for a channel of communication involving more than one human output-input combination, e.g. oral-auditory, gesture-visual (Gibbon et al. 2000). 'Multimedia' stands for a channel of communication involving more than one technical subchannel, e.g. sound and video. Multimodal communication may not necessarily be multimedia communication, and vice versa (e.g. visual multimedia communication with video and text). Challenges and solutions are presented to overcome today's limitations of speech and language processing technologies, particularly in the context of the vast number of languages which are not supported yet. The chapter explains the goals and target applications of multimodal interfaces. It describes models, tools, methodologies, and standards for the multimodal input side and focuses on the interaction with Virtual Characters and on the corpus-based approaches that can inspire the design of such systems.

6.1. Speech technology for multilingual and multimodal interaction

The performance of speech and language processing technologies has improved dramatically over the past decade, with an increasing number of commercial and research systems being deployed in a large variety of applications, such as spoken dialog systems, speech summarization and information retrieval systems, voice-operated cell-phone or car navigation systems, and speech translation assistance, as well as applications in military and security domains (a good overview can be found in the special section on "Spoken Language Technology", IEEE Signal Processing Magazin, May 2008). Among the most important trends in modern speech technology is the need to support multiple input and output languages, particularly since an increasing number of applications are intended for international markets and linguistically diverse user communities (Schultz and Kirchhoff 2006). The challenge of rapidly adapting existing speech processing systems to new languages is currently one of the major bottlenecks in the development of multilingual speech technology (Schultz 2004). While most efforts in speech technologies to date are focused on acoustic speech signals as input modality, latest research in the area of brain computer and silent speech interfaces indicates that a variety of other modalities may serve as input signal for speech processing systems (Denby et al. 2010). This section gives a brief overview of the state-of-the-art in speech technology in the context of multilingual and multimodal interaction. It summarizes the major challenges and presents possible solutions for multilingual speech processing and gives a brief survey of significant ongoing research, trends and open research issues in the area of silent speech interfaces. Furthermore, multimodal interaction strategies (see Alwood and Ahlsén 2012 in this volume) are discussed with an emphasis on spoken dialog systems and the recognition of user states from multimodal traits. A subsection on multimodal speaker recognition complements our short introduction to speech technologies.

6.1.1. Multilingual speech processing

Most research efforts in speech and language processing to date were focused on a very small number of languages spoken by a large number of speakers in countries of great economic potential, and a population with immediate information technology needs. However, speech technology has a lot to contribute to those languages that do not fall into this category. Languages with a small number of speakers and few linguistic resources may suddenly become of interest for humanitarian, economical or military reasons. Also, a large number of languages are in danger of becoming extinct (Crystal 1999), and ongoing projects for preserving them could benefit from speech technologies. We will introduce the major challenges for speech and language processing posed by the various language peculiarities and will describe possible solutions.

6.1.1.1. Challenges and language peculiarities

The availability of speech technology for different languages today closely mirrors the level of research effort and funding in different countries over the past two or three decades. Of all languages, English has probably had the largest share in speech technology research and development. This is not only due to the amount of research funding in English-speaking countries, but also due to the fact that early efforts towards standardization of language resources concentrated on the English language. Another factor is the impact of benchmark evaluations in a range of different speech processing applications organized by the U.S. National Institute for Standards and Technology (NIST, http://www.nist.gov). These evaluations have been held annually since the early nineties and regularly attract many participating teams from all over the world. Large R&D efforts also exist for major European and Asian languages, such as German, French, Chinese, Arabic, or Japanese.

When trying to determine which languages are going to play a significant role in speech processing for the foreseeable future, the most important predictors are the number of speakers, the economic potential of the countries where those languages are spoken, and the information technology needs of the population. Clearly, the level of speech technology will never be the same for all languages; commercial developers of speech technology will only invest in a given language when there is the prospect of investment return, which in turn depends on market demands and cost effectiveness. Based on purely economic considerations, it is probably safe to say that current "mainstream" languages will continue to be of major interest to speech researchers and developers, and that Asian and Southeast Asian languages will become increasingly more important. However, speech technology research has much to contribute even to those languages that do not fall into this category. First, languages with a relatively small number of speakers and few linguistic resources may suddenly become of interest for military reasons or because of a catastrophic event requiring international rescue efforts, such as the 2004 tsunami in Southeast Asia. Second, a large number of languages are in danger of becoming extinct. Ongoing projects for preserving them (e.g. http://emeld.org, http://www.oralliterature.org) could benefit from speech and language technology. Yet, at the time of writing, state-of-the-art speech technology is not even available for the twenty most widely spoken languages in the world (see Table 1). If speech processing systems could be ported to new languages rapidly and at a reasonable cost, speech applications could be made available for a much larger number of languages.

Language characteristics and peculiarities:
How many languages exist in the world? The perhaps surprising answer is that the exact number is not known; any published figures only represent estimates rather than precise counts. Generally, the number of languages in the world is estimated to be between 4000 and 8000. The most recent edition of the Ethnologue (http://www.ethnologue.com), a database describing all known living languages (Gordon 2005), lists 6,912 known living languages, a classification established based on mutual intelligibility and ethnolinguistic identity. Table 1 shows the number of speakers for the twenty most widely spoken languages – note that this only includes *first-language* speakers, not second or third-language speakers. According to the Ethnologue, 5 % of the world's languages are spoken by 94 % of the world's population, whereas the remaining 95 % are distributed over only 6 % of the population. Asia and Africa have the largest *number* of languages, while the most *widely distributed* languages are those of Asia and Europe.

Table 1. The twenty most widely spoken languages in the world according to the number of first-language speakers (Gordon 2005).

#	Language	Speakers (in Mio)	#	Language	Speakers (in Mio)
1	Mandarin Chinese	867.2	11	Wu Chinese	77.2
2	Spanish	322.3	12	Javanese	75.6
3	English	309.4	13	Telugu	69.7
4	Arabic	206.0	14	Marathi	68.0
5	Hindi	180.8	15	Vietnamese	67.4
6	Portuguese	177.5	16	Korean	67.0
7	Bengali	171.1	17	Tamil	66.0
8	Russian	145.0	18	French	64.8
9	Japanese	122.4	19	Italian	61.5
10	German	95.4	20	Urdu	60.5

When building speech processing systems, various challenges can occur at different linguistic levels. These include differences in:

- *sound system*, e.g. small, simple systems versus those that contain complex consonantal clusters, diphthongs and triphthongs, as well as word discrimination systems using tone, stress, or duration;
- *writing system*, such as logographic scripts (Hanzi, Kanji) versus segmental (Roman, Cyrillic), syllabic (Devanagari), or featural (Hangul, Kana) phonographic scripts; complications such as those posed by the Semitic writing systems (Arabic, Hebrew) that mostly omit long vowels; and, finally, languages without any written form at all;
- the relationship between sound and writing system, i.e. the ease of mapping *letters to sounds* and thus generating pronunciations;
- *word segmentation* system, i.e. languages that lack a segmentation of sentence-level character strings into individual words (e.g. Chinese, Japanese, Thai) and thus require pre-processing to extract units suitable for text and speech processing;
- *morphology*, contrasting languages with a very rich morphology (Korean, Turkish, German) with languages having a small number of derivative forms (English) or no derivation at all (Chinese). The morphological structure has a large impact on the vocabulary growth within a language and thus on the out-of-vocabulary rate.

Each of these factors (as well as several other factors not discussed here, such as syntax or word usage) may have a large impact on model quality, as well as on the amount of resources, time, and effort needed for system development (see Schultz and Kirchhoff 2006 for more details).

Algorithms and lack of resources:
For more than 20 years speech recognition has been dominated by statistically based modeling schemes, such as *Hidden Markov Models* (HMM) and n-gram language models. Both modeling schemes have turned out to be very successful in speech recognition despite their limits. Recently, statistical models have also become the predominant paradigm in speech synthesis and machine translation.

The core algorithms applied in statistical modeling are mostly language independent and have indeed proven to work reasonably well for a variety of languages. On the flip side, statistical modeling relies heavily on vast amounts of training data for robust parameter estimation. For high-end performance, thousands of hours of (quickly) transcribed audio data and hundreds of Millions words text data based on hundreds of thousand word forms are typically used.

Lack of data:
As the public interest turns towards less widespread languages, it becomes clear that the traditional approach to system development is prohibitive for all but the most widely spoken, widely read, and economically viable languages. To date, large-scale data resources have been systematically collected and distributed for less than 50 languages. Major collections are provided by consortia such as Linguistic Data Consortium (LDC, http://www.ldc.upenn.edu) or the European Language Resources Association (ELRA, http://www.elra.info). To foster and support research on multilingual speech technologies, uniform collection scenarios across languages are crucial. Unfortunately, only very few data collections take these requirements into account. One of the few exceptions is GlobalPhone, a standardized multilingual text and speech database that currently covers 20 languages Arabic, Bulgarian, Chinese (Shanghai, Mandarin), Croatian, Czech, French, German, Hausa, Japanese, Korean, Russian, Polish, Portuguese, Spanish, Swedish, Tamil, Thai, Turkish, and Vietnamese (Schultz 2002) and is available from ELRA.

Lack of language conventions:
A surprisingly large number of languages or dialects do not have a writing system, or only an ill-defined (i.e. unsystematic) writing system. Omniglot (http://omniglot.com) lists around 780 languages that have scripts, according to Ager (2007), the number of languages with writing systems might be closer to 1000. For languages without any written form it is not possible to harvest large amounts of text corpora or even create word lists, pronunciation dictionaries,

and lexicons. Even for many languages that are written, the lack of a standardized form significantly hinders web-based harvesting or the consistent construction of dictionaries and lexicons. For example, to enable speech processing for Arabic dialects, writing systems had to be standardized and transcribers had to be trained to apply them consistently. This procedure has been used, for example, in recent U.S. government funded projects for Egyptian Arabic, Pashto, and Iraqi Arabic (Precoda et al. 2007).

Gap between language and technology expertise:
With a suitable amount of data and well-defined standards, the development of speech processing components might be considered to be a rather straightforward task. However, it turns out to be surprisingly time- and cost intensive to build well-performing speech processing systems for new languages. As pointed out above, this is partly due to language-specific peculiarities, but another reason is the lack of language experts, especially for low-density languages. The "TransTac 100-day challenge", a current U.S. government project on rapid development of an English-Iraqi Arabic speech translation system revealed that the most crucial factor for timely development was the lack of language and cultural expertise. It turned out to be extremely difficult to find native speakers who simultaneously have enough insight into their own language *and* the necessary technical background. Without such a skilled language expert, speech processing developers face the time-consuming situation of having to either familiarize themselves with the language or train an unskilled language expert. Social and cultural aspects may further complicate the process, e.g. native speakers may not wish to have their voices recorded, or communication between native informants and non-native developers may be difficult. We have also experienced an interesting intercultural issue: native speakers may be eager to help but are reluctant to identify flaws in the system. Consequently, one of the central issues in building systems for new languages is to *bridge the gap between language and technology expertise*.

6.1.1.2. Rapid language adaptation

With more than 6900 languages in the world and the need to support multiple input and output languages, the most important challenge today is to port or adapt speech processing systems to new languages rapidly and at reasonable costs (Schultz 2004). Major bottlenecks are the sparseness of speech and text data, the lack of language conventions, and the gap between technology and language expertise. Data sparseness is a critical issue due to the fact that today's speech technologies heavily rely on statistically based modeling schemes, such as Hidden Markov Models and n-gram language modeling. Although statistical modeling algorithms are mostly language independent and proved to work well for a variety of languages, the parameter estimation requires vast amounts of training data. Large-scale data resources are currently available for less than 50 languages and the costs for these collections are prohibitive to all but the most widely spoken and economically viable languages. The lack of language conventions concerns a surprisingly large number of languages and dialects. The lack of a standardized writing system for example hinders web harvesting of large text corpora or the construction of dictionaries and lexicons. Last but not least, despite the well-defined process of system building, it is very cost- and time consuming to handle language-specific peculiarities, and requires substantial language expertise. Unfortunately, it is extremely difficult to find system developers who simultaneously have the necessary technical background and significant insight into the language in question. As a result, one of the central issues in developing speech processing systems in many languages is the challenge of bridging the gap between language and technology expertise.

To date, the standard way of building speech applications for an unsupported language is to collect a sizable training corpus and to train statistical models for the new language from scratch. Considering the enormous number of languages and dialects in the world, this is clearly a suboptimal strategy, which highlights the need for more sophisticated modeling techniques. It would be desirable to develop models that can take advantage of similarities between dialects or languages of similar type and models that can share data across different varieties. This would have two benefits, first leading to truly multilingual speech processing which can handle code switching, and second providing models that are likely to be more robust toward dialectal and cross-lingual pronunciation variations. These multilingual shared models can then be used as seed models to jump-start a system in an unsupported language by efficiently adapting the seeds using limited data from the language in questions. We refer to this development strategy as *rapid language adaptation.*

The Rapid Language Adaptation project of the Cognitive Systems Lab (CSL) at the Karlsruhe Institute of Technology and the Language Technologies Institute at Carnegie Mellon has successfully launched a set of web-based tools, namely the **R**apid **L**anguage **A**daptation **T**oolkit (RLAT) and the SPICE Toolkit (**S**peech **P**rocessing – **I**nteractive **C**reation and **E**valuation), which aim to significantly reduce the amount of time and effort involved in building speech processing systems for new languages. RLAT and SPICE provide innovative methods and interactive web-based tools to enable users to develop speech processing models, to collect appropriate speech and text data to build these models, as well as to evaluate the results allowing for iterative improvements (Schultz et al. 2007). In particular, the toolkits allow the user to (1) design databases for new languages at low cost by enabling the user to record appropriate speech data along with transcriptions as well as to harvest and process massive amounts of text data from the web, (2) to select appropriate phone sets for new

Figure 1. SPICE Audio Collection Interface.

languages efficiently, (3) to create vocabulary lists and generate pronunciation dictionaries, (4) to apply these resources by developing acoustic and language models for speech recognition and (5) to develop models for text-to-speech synthesis, and (6) to finally integrate the built components into an application and evaluate the results using online speech recognition and synthesis in a talk-back function. Figure 1 shows the Java-based Audio Collection Interface to support the recording and archiving of speech data and transcripts. Visual feedback is provided to reduce audio quality issues.

RLAT and SPICE leverage the mature projects GlobalPhone (Schultz 2002) and FestVox (Black and Lenzo 2000) and implements bootstrapping techniques that are based on extensive knowledge and data sharing across languages, as well as sharing across system components (Schultz et al. 2007). Examples for data sharing techniques are the training of multilingual acoustic models across languages based on the definition of global phone sets. Sharing across components happens on all levels between recognition and synthesis, including phone sets, pronunciation dictionaries, acoustic models, and text resources.

RLAT (http://csl.anthropomatik.kit.edu/rlat-dev) and SPICE (http://cmuspice.org) are online services and the interface to the web-based tools has been designed to accommodate all potential users, ranging from novices to experts.

Novice users are able to read easy-to-follow, step-by-step instructions as they build a speech processing. Expert users can skip past these instructions. In addition, file-uploading routines allow for feeding the bootstrapping algorithms with available data and thus shortcut the process. The result is that RLAT and SPICE can collect information from the broadest array of people: a general audience of Internet users who may have little experience with speech tools, and a specific audience of speech and language experts, who can use data they already have. The tools allow for speech and massive text data harvesting and archiving, soliciting basic language information, logging activity, user profiles, development projects, and the automatic construction of speech recognition and speech synthesis including acoustic model bootstrap from multilingual models, text cleaning and normalization, vocabulary selection, language model generation, and pronunciation generation. By keeping the users in the developmental loop, the RLAT and SPICE tools can learn from their expertise to constantly adapt and improve.

User studies were carried out to indicate how well speech systems can be built, how well the tools support development efforts, and what must be improved to create even better systems. While the SPICE tools have been used previously to bootstrap speech recognition systems in Afrikaans, Bulgarian, and Vietnamese within a 60-day timeframe, they recently targeted the parallel development of speech processing components for a broader range of ten languages within a six-week hands-on lab course at Carnegie Mellon University and Karlsruhe Institute of Technology. Students were required to rely solely on the SPICE and RLAT tools and report back on problems and limitations of the system. Results indicate that it is feasible to build various end-to-end speech processing systems in new languages for small domains within the framework of a six-week course. The rapid language adaptation toolset will hopefully revolutionize the system development process in the future. Archiving the data gathered on-the-fly from many cooperative native users will significantly increase the repository of languages and resources. Data and components for under-supported languages will become available at large to let everyone participate in the information revolution, improve the mutual understanding, bridge language barriers, and thus foster educational and cultural exchange.

6.1.2. Silent speech interfaces

While most efforts in the area of speech-driven technologies and applications to date focus on the acoustic speech waveform as input modality, a significant body of research and technologies emerges on brain computer and silent speech interfaces that leverage off a variety of modalities arising from spoken communication: sensory data are acquired from the human speech production process ranging from the activity of articulators, their neural pathways, and the brain itself

(Denby et al. 2010). These types of modalities allow the development of "Silent Speech Interfaces", i.e. systems which enable the communication via speech even when audible acoustic signals are corrupted, not available, or not desired.

Silent Speech Interfaces have the potential to overcome major limitations of conventional speech-driven interfaces. Firstly, since the speech signals are captured before getting airborne, they are not prone to ambient noise and thus provide a solution to robust speech recognition in very noisy environments. Secondly, conventional speech interfaces rely on audibly uttered speech, which has two major drawbacks: it jeopardizes confidential communication in public and it disturbs bystanders. Services which require the search, retrieval, and transmission of private or confidential information, such as PINs, passwords, and security or safety information are particularly vulnerable. Silent Speech Interfaces allow to utter speech silently and thus overcome both limitations (Schultz and Wand 2010). Last but not least, Silent Speech Interfaces might give hope to people with certain speech disabilities as the technologies allow the building of virtual prostheses for patients without vocal folds (Denby et al. 2006). Also, elderly and weak people may benefit since silent articulation can be produced with less effort than audible speech.

Silent Speech Interfaces are still in the experimental stage, so we briefly summarize the types of technology that have been described in the literature so far (see Denby et al. 2010 for a detailed review of Silent Speech Technologies).

Non-Audible Murmur (NAM), a term coined by Nakajima et al. (2003), refers to very low amplitude sounds generated by laryngeal airflow noise and its resonance in the vocal tract. While the sound can barely be perceived by nearby listeners, it can easily be detected using a type of stethoscopic microphone attached to the skin (Nakajima et al. 2003). Nakajima, Heracleous, and others successfully applied NAM for robust speech recognition in noisy environments (Nakajima et al. 2006; Heracleous et al. 2007). Tran investigated speech transformation techniques to produce more natural sounding NAM-based speech synthesis, particularly useful for speakers with voice pathologies due to laryngeal disorders (Tran et al. 2010).

Analysis of glottal activity using electromagnetic (Titze et al. 2000; Ng et al. 2000; Tardelli 2004; Preuss et al. 2006; Quatieri et al. 2006), or vibration (Bos and Tack 2005; Hansen et al. 2010) sensors with the goal of obtaining glottal waveforms that can be used for de-noising by correlation with an acoustic signal captured by standard close-talk microphones. The required waveforms may be obtained either via detectors which are directly sensitive to vibrations transmitted through tissue such as throat microphones, or from the interaction of glottal movement with an imposed electromagnetic field. Research in this area was propelled by DARPA's Advanced Speech Encoding Program, which provided funding to develop non-acoustic sensors for low bit rate speech encoding in harsh acoustic environments.

Electromagnetic Articulography (EMA): EMA systems aim to track the precise Cartesian coordinates of fixed points within the vocal tract as its shape is the most crucial part of speech production. Many researchers investigated the method of implanted coils that are electrically connected to external equipment and are electromagnetically coupled to external excitation coils (Carstens 2008; Schönle et al. 1987; Hummel et al. 2006). While conventional EMA is very inconvenient for the speaker, Fagan et al. (2008) recently developed a less intrusive system which consists of permanent magnets glued to some points in the vocal apparatus that are coupled with magnetic sensors mounted on a pair of spectacles worn by the user. In laboratory conditions promising recognition results could be achieved on a small vocabulary (Fagan et al. 2008).

Ultrasound (US): Real-time characterization of the vocal tract using ultrasound and optical imaging of the tongue and lips (Denby and Stone 2004; Denby et al. 2006; Hueber et al. 2007; Hueber et al. 2010) developed a system that captures articulatory activity via ultrasound and video imaging and then transforms these multimodal observations of articulatory gestures into an audio speech signal based on a direct mapping between visual and acoustic features using Gaussian mixture models and alternatively based on an intermediate HMM-based phonetic decoding step to take advantage of *a priori* linguistic information. The resulting "silent vocoder" is particularly helpful to patients who have undergone a laryngectomy.

Surface electromyography (sEMG) of the articulator muscles or the larynx (Jorgensen et al. 2003; Maier-Hein et al. 2005; Jou et al. 2006; Jorgensen and Dusan 2010; Schultz and Wand 2010); Surface ElectroMyoGraphy (sEMG) is the process of recording electrical muscle activity captured by surface electrodes (see Figure 2, Schultz et al. 2010). As speech is produced by the activity of human articulatory muscles, the resulting myoelectric signal patterns provides a means of recovering the speech corresponding to it. Furthermore, speech can be recognized even when produced silently, i.e., without any vocal effort, and the signal cannot be corrupted or masked by ambient noise transmitted through air. While (Jorgensen et al. 2003) was first to prove the applicability of myoelectric signals for non-audible speech recognition on a small set of words, recent studies tackle large vocabulary speech recognition by removing the restriction of words or commands spoken in isolation and evolve toward a less limited, more user-friendly continuous speaking style (Maier-Hein et al. 2005); allowing for acoustic units smaller than words or phrases (Walliczek et al. 2006; Schultz and Wand 2010); study the effects of electrode re-positioning (Maier-Hein et al. 2005) and more robust signal preprocessing (Jorgensen and Binsted 2005; Jou et al. 2006); investigate real-life applicability, by augmenting conventional speech recognition systems (Chan 2003), implementing alternative modeling schemes such as phonetic features to enhancing phoneme models (Jou et al. 2007); and to address coarticulation effects (Schultz and Wand, 2010); by

Figure 2. EMG-based Silent Speech Recognition (CeBIT Press).

examining the impact of speaker dependencies on the myoelectric signal (Wand and Schultz 2009); and by addressing size, attachment, and mobility of the capturing devices (Manabe et al. 2003; Manabe and Zhang 2004). Most recently, Schultz and Wand (2010) showed first experiments on speaker independent and speaker adaptive sEMG-based speech recognition based on a large collection of EMG data recorded from 78 speakers reading sentences in both, audible and silent speaking mode that gives 10 % word error rate on the best performing speakers and thus comes within reach for practical applications.

Brain-Computer Interfaces: Interpretation of signals from electroencephalographic (EEG) sensors (Porbadnigk et al. 2009) and from implants in the speech-motor cortex (Brumberg et al. 2010). EEG-based brain computer interfaces (BCI) have become an increasingly active field of research. Overviews can be found in Dornhege et al. (2007) and in Wolpaw et al. (2002). Prominent examples of current BCIs, where the aim is to translate thoughts or intentions into control signals for devices, include the Thought Translation Device (Birbaumer 2000) and the Berlin Brain Computer Interface (Blankertz et al. 2006). To avoid the time consuming learning process, and develop more intuitive silent speech interfaces, Wester and Schultz (2006) investigated a new research approach that directly recognizes unspoken speech from EEG signals, i.e. speech imagined to be spoken but without actually producing any sound or moving any articulatory muscle. Also Suppes showed that isolated words can be recognized based on EEG and MEG recordings (Suppes et al. 1997) while a study by Porbadnigk et al. (2009) indicates that temporally correlated brain activities may superimpose the signal of interest. DaSalla and colleagues (DaSalla et al. 2008) proposed a control scheme for a silent speech brain computer interface using neural activities associated with vowel speech imagery. Recently, attempts have been made to utilize intracortical microelectrode technology and neural decoding techniques which have the potential to restore speech communication to paralyzed individuals (Brumberg et al. 2010; Bartels et al. 2008), or to restore written communication through development of mouse cursor control BCIs for use with virtual keyboards (Kennedy et al. 2000) and other augmentative and alternative communication devices (Hochberg et al. 2008).

While substantial progress has been made over the last 5 years, current Silent Speech technologies are not quite ready for the market yet, where usability and cost are of major concern. Among other factors, future progress will likely benefit from advances in dedicated sensory equipment but also in signal processing techniques to limit the impact of artifacts resulting for example from sensor positioning, environmental conditions, and various user characteristics.

6.1.3. Cognitive interaction systems

Computers play an important role in our everyday lives for communication, work, gaming, and entertainment. However, despite significant research efforts and substantial progress in computational power, the way humans operate machines has not dramatically changed since the upcoming of graphical user interfaces in the 1980s. Human-Machine interaction is still surprisingly unnatural and non-intuitive (see also Kopp and Wachsmuth 2012 in this volume). However, in our modern and global society humans have to handle dynamic and complex processes with overwhelming amounts of information and the need for smart systems to augment human capabilities is constantly rising. Several studies like the media equation (Nass et al. 1994) and the Computers Are Social Actors paradigm (Reeves and Nass 1996) state that human-machine interaction follows the same rules as human-human interaction and has a fundamentally social nature. What seems to be lacking so far are human-centered technologies, which constantly analyze, recognize, and adapt to natural human thinking, feeling, and behaving in order to integrate machines into our everyday life in a seamless and efficient way. This is the main target of research in the area of affective and ubiquitous computing (see also Lauth et al. 2012 in this volume). Klien et al. (2004) formulates the vision as to overcome the logical and rational way in which computers are operated, and integrate them into our lives as proactive assistants and team members.

When it comes to spoken interactions, today's systems are mostly incapable to carry out a natural communication with the user. For example, dialog systems can still not cope with the many ways to express any given content. Therefore, interaction systems are necessary which systematically observe the cognitive/ inner states and traits of a user, which are able to understand and model the user's needs and, which constantly adapt its interaction strategies to changing conditions. These Cognitive Interaction Systems will make use of cognitive modeling components, such as memory and emotional models of its user and combine the information to derive robust estimates of the user's state. This section on cognitive interaction strategies emphasizes speech as input modality. It surveys research on spoken dialog systems and introduces components to detect a user's state, in order to provide more information to the computer for intuitive spoken interaction systems. Particularly, it focuses on emotions, personality, and mental task demand, which fundamentally coin many aspects of human behavior in a particular situation and can provide the basis for adequate adaptation of spoken communication systems.

6.1.3.1. Spoken dialog systems and interaction strategies

Spoken dialog systems have come a long way since the first voice-activated interfaces and today are widely used in speech-driven applications, such as automated call-centers (Gorin et al. 1997), information portals (Raux et al. 2005), in-car interfaces (Becker et al. 2006), and tutoring systems (Litman and Silliman 2004). The gathered experience has been used to generalize to more generic systems with multimodal input and output (Oviatt et al. 2004; Benoit et al. 2000; Wahlster et al. 2006) which allow a more natural and robust interaction between man and machine. However, many challenges still remain, making dialog systems an active and growing field of research.

A dialog manager offers a flexible management component which integrates all peripheral components for interaction with the user, like speech recognition, language understanding, language generation and speech synthesis. It also contains the dialog strategy which controls the actions of the system. State-of-the-art dialog systems implement elaborated discourse and dialog theories and are able to handle generic meta-communication and dialog phenomena like confirmations and error recovery. A number of existing dialog system frameworks allow rapid bootstrapping of dialog systems for new domains and enable by their modular structure the easy integration of new components (Larsson and Traum 2000; Bohus et al. 2007; Holzapfel 2008).

A variety of approaches is available for the implementation of the dialog strategy, ranging from simple but efficient finite state machines, over information state based approaches (Traum and Larsson 2003) to statistical, automatically learned strategies which now transit from academic research into industrial application (Paek 2006). VoiceXML manifests a standardized specification language for dialog systems (VoiceXML 2007) (for XML-based representation models see Stührenberg 2012 in this volume). The goal of this standard is to offer building blocks for voice-driven dialog systems analogous to the components which are already well established for graphical user interfaces, i.e. selections from a set of given options, links, and input fields. It also offers ready-to-use options to handle barge-ins, repair requests, or confirmations.

For a more natural interaction, many modern dialog systems are not restricted to speech-only input and output but allow several modalities for both directions. The most mature approach is the combination of speech and gestures, which is especially interesting because the two modalities can both disambiguate and complement each other. This development begun with R. A. Bolt's Media Room (Bolt 1980), in which spoken commands like "Create a blue square there" combined with a pointing gesture would create the appropriate symbol on the screen. Since then, the naturalness of speech and gesture input as well as the quality of fusion has been vastly improved. This quality is demonstrated in realistic applications like crisis room control (Rauschert et al. 2002; Krahnstoever et al. 2002) or human-robot interaction (Stiefelhagen et al. 2004). Section 6.2 provides a more detailed survey of research in this area.

While the foundations of dialog management are well established with the existing tools, designing a system for a good user experience is still challenging and remains an open research issue. Recently, several user studies showed that systems which are tailored towards the specific needs of a user perform better in all aspects of usability. In (Nass et al. 2005) the authors present a user study with a "virtual passenger" in the car which indicates that adapting to the driver's emotion by the dialog system in terms of voice characteristics improves both subjective impression and objective criteria, such as driving quality. The authors show that there is no single optimal emotional stance for the virtual passenger but that the system has to match the current (usually changing) emotional state of the user. Another study (Nass and Lee 2000) shows that similarity attraction is also valid for personality expression in man-machine-interaction as matching the users' degree of extraversion strongly influenced their trust and assigned intelligence in a spoken dialog based book shop. Studies investigating the interaction behavior of users in different emotional states and situations show the feasibility of detecting the affective state. Motivated by these observations, dialog systems emerged which model the users' traits and states by observing them during interaction. These models then influence the system's behavior: Instead of a fixed interaction pattern for all users, the behavior depends on several attributes of the discourse and the user's behavior. This can be measured indirectly, i.e. by observing the development of dialog parameters, such as recognition rate, average turn length etc. This approach is used in Bohus et al. (2006) to locally learn optimal recovery strategies from non-understandings using linear regression models with input features derived from the discourse.

Hassel and Hagen (2006) use a heuristic approach to derive the user's technical expertise from a user model that dynamically estimates the user's capability to cope with the system by counting the number of help requests or timeouts. It then dynamically adapts the output in terms of explicitness or length of help statements. Similar adaptation techniques have been applied to non-verbal interactions, e.g. in Microsoft's Lumiere project (Horvitz et al. 1998) which assessed a user's task and competence state of an office application using dynamic Bayesian networks to deduce the best supporting action for the current situation. Several features of the user's behavior are extracted from the application's help menu to identify the most probable task and the need for assistance, which is then weighed against the costs for interrupting the current work flow.

Another approach is to directly observe the user in a multimodal way and base the adaptation on this information. This is for example used by Mairesse and Walker (2005), who use lexical and prosodic features from dialog speech to derive the speaker's personality, characterized on five dimensions. The authors propose language generation parameterization techniques to adapt a dialog system in order to mirror the expressed personality of its current user. The authors of Gnjatović and Rössner (2008) describe a dialog system that bases its handcrafted dialog strategy for a gaming interface on the user's emotional state, derived from prosody, language and visual features. Together with the history of interaction, the current user command and other discourse features, the user state can be accessed by the dialog strategy that is implemented in form of a decision tree which can branch on different user states. Nasoz and Lisetti (2007) describe a user modeling approach for an intelligent driving assistant. This model is based on a Bayesian network which allows to deriving the most useful system action in terms of driving safety given the estimated driver state. It consists of emotional state, personality and other features and is partially derived from physiological sensory data like the user's heart rate. The score for each action is calculated using a utility node which measures the probability of safety improvement given the current user state. Similar decision-theoretic user models system action evaluation approaches are used in Li and Ji (2005), which also include an active sensor selection mechanism. Conati (2002) presents an educational dialog system that can decide for different user assistance options, given the user's emotional state derived from multiple modalities. This work bases its network on the cognitive OCC (by Ortony, Clore and Collins) appraisal theory, which relates the users' emotions with their goals and expectations.

All approaches presented above have in common that they perform their adaptation locally, i.e. they decide on a turn level with limited context and strategic foresight. Recently, Reinforcement Learning (Sutton and Barto 1998) has become a promising approach for automatically training long-term optimal dialog strategies in dialog systems. This is especially important if the state spaces and the modeled behavior become more complex, leading to less straight-forward decision making. Reinforcement Learning regards the dialog discourse as a Markov decision process, where typically each transition between two dialog states marks one speech act by either the system or the user. Walker et al. (1998)

and Singh et al. (2002) used Reinforcement Learning to build a dialog strategy for simple information presentation systems where the dialog space consisted of the state of the relevant information slots and the action set was formed by the available speech acts. Singh et al. (2002) shows that their strategy outperforms handcrafted ones in a user evaluation, indicating that automatically learned strategies can discover optimal behavior that improves heuristics of human developers. Since the first trained interaction strategies, many improvements have been achieved and applications developed, as briefly summarized in the following paragraphs.

Instead of limiting the input to speech and speech acts, there exist interaction strategies which allow multimodal input such as gestures or multimodally acquired user identities (Prommer et al. 2006) to enrich the state space and the set of possible user actions by additional dimensions. In Bui et al. (2007), the authors describe a dialog system that uses Reinforcement Learning in partially observable Markov models which include information on the user's affective state, such as stress. In simulations, they show that their system can handle stress-induced user errors better than handcrafted strategies. Both works show that modality-independent representations of information concerning the user or the interaction can be integrated into generic learning frameworks. Reinforcement Learning has also been applied to adaptive dialog systems. Tapus et al. (2007) presents a speech-enabled therapist robot which adapts the conveyed personality (manifested by the maintained proximity to the patient and the therapy style, ranging from nurturing to challenging) to the different personalities of its rehabilitation patients. The system uses online Reinforcement Learning to train its behavior. This allows the system to adapt to the preferences of individual users on runtime. The authors of Janarthanam and Lemon (2008) demonstrate that Reinforcement Learning can be applied successfully to learn not only the dialog act (what to present) but also the style of presentation (how to present). Their troubleshooting system can select the degree of presentation detail based on an automatically estimated level of expertise of the user.

Early systems usually used reward functions that only regarded the pure dialog success. However, further research suggested to use different reward functions that better represent the user's needs. Walker et al. (1997) and Walker (2005) proposed to use the predictions of the PARADISE framework as reward function. PARADISE takes simple features from the past discourse, e.g. the number of turns or the number of time-outs, and uses a linear regression model to predict user-satisfaction. As most features can be extracted automatically at runtime, the predicted satisfaction can be used as reward function. Recently, Rieser and Lemon (2008) extended this work by evaluating different learned reward functions on data from real users. PARADISE is not only helpful in the context of Reinforcement Learning, but for automated usability testing in the early stages of development of a new dialog system in general. In contrast to many other works in the area of cognitive dialog systems, the models of the READY project (Schäfer et al. 1997) make use of psychological knowledge of mental processes, e.g. capacity-limited working memory or non-perfect language understanding. In a dynamic Bayesian Network, this is integrated in a framework to model the user's cognitive capacity. This framework allows the user to evaluate different actions and their chance of success. In Bohnenberger et al. (2002), the authors extend this work by modeling a speech-driven man-machine-interaction with resource limitations (i.e. time pressure on the user). Using this model, they generate policies optimizing the expected dialog success rate. The strategy can balance between efficiency (giving much information at once) and safety (asking for confirmation after each piece of information), based on the user's cognitive load.

An intuitive continuation of the development is to design cognitive interaction systems (Putze and Schultz 2009a) that systematically model and observe the user's states and traits and are able to adapt their interaction strategies to changing conditions. Cognitive interaction systems make use of cognitive modeling components, for example memory models (Putze and Schultz 2009b) or emotional models which use psychologically sound theories and are combined with observational data to derive a robust estimation of the user's inner state. This information is employed by adaptive interaction strategies which are trained using a Reinforcement Learning approach within a user simulation framework which is again driven by cognitive modeling components.

6.1.3.2. Recognition of (cognitive) user states

In the following subsection we briefly survey systems that attempt to recognize human emotions, personality, and mental task demand, which fundamentally coin many aspects of human behavior in a particular situation and might serve as the basis for adequate adaptation of communication systems to particular situations in user interaction.

Emotion recognition

Emotions are intensively researched in the psychological community and known for their strong influence on human-human communication. Therefore, the automatic recognition of emotions by computers received substantial attention in the research community over the last years. Several systems for emotion recognition from multimodal biosignals have been proposed. In Picard et al. (2001) a recognition rate of 81 % could be achieved by collecting electromyographic data, blood volume pressure, skin conductance, and respiration information from one person during several weeks for eight emotion categories. To handle daily variations, a special feature set was developed. Recognition was performed by Fisher Projection with the results of Sequential Floating Forward

Search. Kim et. al. (2004) studied user-independent information from skin temperature, skin conductance, and heart rate, and achieved a recognition rate of 61.76 % on four emotion categories by applying *Support Vector Machines* (SVM) (for machine learning methods in text mining see Paaß 2012 in this volume). In Haag et al. (2004) a differentiation of 96.58 % on arousal ratings, and 89.93 % on valence ratings could be achieved using data from one single subject. Features were extracted from electrocardiographic and electromyographic signals, blood volume pressure, skin conductance, respiration, and temperature. Classification was done with a neural network. (Takahashi 2004) described a user-independent study where features were extracted from skin conductance, heart rate and electroencephalographic signals. By using SVMs as a classifier, a recognition rate of 41.7 % could be achieved, differentiating among the five emotions joy, anger, sadness, fear, and relax. When using only joy, anger, and sadness, the recognition rate increased to 66.7 %.

Electroencephalography (EEG) is a promising field for the assessment of human emotions as it uses information directly from the brain, which can be viewed as the source of emotions. EEG sensors are applied directly to the head, and therefore do not restrict the choice of clothes and allow building hands-free systems. Furthermore, the EEG signal is hard to manipulate intentionally, while speech and facial expressions are more prone to deception. Additionally, bioelectrical signals are emitted continuously and are independent of lighting conditions, or background noise, in contrast to visual and auditory signals. Schultz and Schaaff (2009) investigated different feature sets to build an emotion recognition system based on EEG-signals. They used pictures from the International Affective Picture System to induce three emotional states: pleasant, neutral, and unpleasant. A headband with four build-in electrodes at the forehead was used to record data from five male subjects. Compared to standard EEG-caps, the headband is more comfortable to wear and easy to attach, which makes it more suitable for everyday life conditions. They compared two approaches to derive features from the raw data, i.e. Fast Fourier Transform (FFT) frequency components between 5 and 40 Hz are obtained and reduced by averaging over adjacent spectral coefficients versus Peak alpha-frequency, alpha power, and cross-correlation among electrodes within the alpha-band. The authors applied normalization by subtracting each feature's mean and dividing by its standard deviation. For both approaches a correlation-based feature reduction was used to reduce the feature set. Classification was done by SVMs. The system achieves a mean accuracy of 44.00 % with a maximum of 47.78 % for the first feature extraction method and a mean accuracy of 48.89 % with a maximum of 66.67 % for the second feature extraction on subject dependent recognition (for evaluation methods in text mining see Paaß 2012 in this volume).

Automatic recognition of human emotions remains a challenging task and many problems have to be overcome. Future works will most likely leverage advanced machine learning and signal processing techniques to further improve recognition accuracies. Also, it will require engineering effort to develop systems that work more robustly under real-life conditions.

Personality recognition

Personality is an abstract psychological concept and there is no consensus on a generally accepted definition of personality. Pervin and John (2001) define personality as "Personality represents those characteristics of the person that account for consistent patterns of feeling, thinking and behaving". The most important property of personality is its consistency over long periods of time. This is what distinguishes personality from volatile affective states like emotions. Still, there is a strong relationship between both: Emotions are actually dependent on the person's personality and may provide cues to determine it or can be predicted for a known personality in a given situation. Over the past few years, psychologists identified five main dimensions of human personality (Goldberg 1990), the so called "Big Five" model of personality traits. This model consists of the five bipolar scales Neuroticism, Extraversion, Openness to experience, Agreeableness, and Conscienciousness.

To date, only a few automatic systems have been proposed in the literature, which can assess personality information of the user. Mairesse and Walker (2006) were the first to present statistically significant results for the recognition of personality in conversations. They used acoustic data from an essay corpus and the EAR data (Mehl et al. 2006), which consist of random snippets of conversation extracts from the participants' daily life over two days recorded by an Electronically Activated Recorder (EAR). They derived features using Praat (Boersma et al. 2004), utterance type features, MRC Psycholinguistic Database, and the Linguistic Inquiry and Word Count (LIWC) database and applied multiple techniques for classification, regression and ranking to recognize Big Five personality traits. They compared self reports with observer ratings and found that observed personality was easier to model, at least in conversational data. The authors propose to integrate this automatic personality recognition (APR) system into an adaptive dialog framework which modifies parameters of the speech synthesis and language generation modules to create a system personality that fits the user's one. Figure 3 shows the high level architecture of their system, where personality recognition is an integral part of input understanding.

To our knowledge there is only the APR system of (Pianesi et al. 2008) that includes visual information in the multimodal personality recognition system. The authors used acoustic features based on the speech feature extraction toolbox from the Human Dynamics Group at MIT Media Lab (Human Dynamics Group 2004) and visual cues related to fidgeting energy of head, hands and body. These visual features are calculated on the basis of Motion History Images using skin and activity pixel classification and describe the amount of

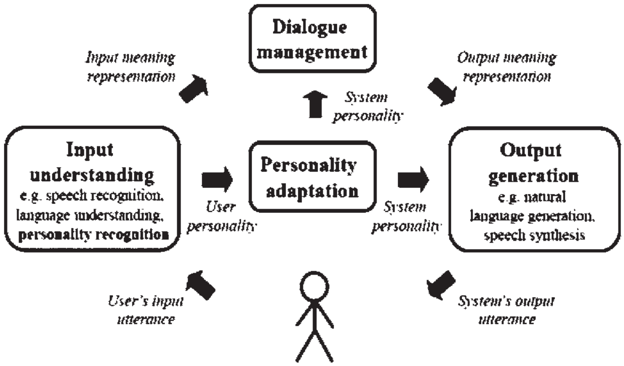


Figure 3. Adaptive Dialog Framework (Mairesse & Walker 2006).

energy associated with body gestures. They used the MS-2 corpus (Mana et al. 2007) consisting of 13 meetings, in which four-person groups discuss on a strategy for survival in a disaster scenario. Their APR system recognizes three classes of intensity (low, middle, and high) based on self reports for extraversion and locus-of-control. Classification is done using windows of one minute duration. The authors used SVMs for classification and report respectable accuracies up to 94 % using 15-fold stratified cross-validation (see Paaß 2012 in this volume), whereby due to the distribution of ground truth classes 66 % of the data belongs to one class.

Automatic personality recognition is a very interesting research discipline, which might have a strong impact on future communication systems. Currently, systems are investigated that exploit information from multimodality by using various biosignals such as EEG and facial expressions in addition to acoustic data.

Workload recognition

The terms workload or mental task demand define the amount of mental resources required to execute a current activity. Task demand information can be helpful in various human-machine communication situations, e.g. while driving a car, operating machines, or performing other critical tasks. Intelligent systems should optimize user performance by adapting themselves depending on the predicted level of demand and cognitive load. For example distraction arising from electronic devices such as text messages, incoming phone calls, traffic or navigation information, etc. can be filtered or delayed in demanding situations. Additionally, the analysis of task demand during the interaction with a system allows to asses usability information.

Several groups reported research on the assessment of task demand based on multimodal data, including muscular activity (Pleydell-Pearce et al. 2003), blood hemodynamics (Izzetoglu et al. 2004), and pupil diameter (Iqbal et al. 2004). Reasonable results could be achieved with all three modalities. However, correlations between pupil diameter and task demand could only be shown for one interactive task out of a group of various cognitive tasks. Healey and Picard (2005) developed a classifier to monitor the stress levels in daily life car driving tasks. They collected data from twenty-four drives of at least fifty minute duration and used the biosignals electromyography, electrocardiography and skin conductance for their system. In the experiment, all participants drove along a pre-defined route where they encountered different traffic conditions that should generate different levels of workload. To assess ground truth of the work load for training and evaluation of the system, a continuous stress metric was derived by analysis of recorded videos and counting the number of stress indicators at each second of driving. Additionally the authors used subjective ratings of the workload obtained by questionnaires. Three different workload intensities were derived from five minute intervals of data. Linear discriminant analysis (LDA) was used for dimensionality reduction of the feature vector consisting of various features of the different modalities. A classifier using a linear decision function was able to discriminate the three classes with accuracies of 100 % (low workload), 94.7 % (medium workload), and 97.4 % (high workload).

Other work focused on the EEG data recorded while varying the difficulties of tasks (Smith et al. 2001; Pleydell-Pearce et al. 2003; Berka et al. 2004). The features extracted from the EEG data represented mostly the frequency content of the signals. Positive correlations between predictions and references or predictions and self-estimates of task demand (Smith et al. 2001) are reported throughout these studies. Pleydell-Pearce (2003) achieved a classification accuracy of 72 % for the discrimination of low versus high task demand in subject and session dependent experiments and 71 % in subject independent experiments. Honal and Schultz (2008) conducted various experiments to evaluate levels of task demand from EEG measured by 4 electrodes using a comfortable headband, as well as an electrode cap with 16 electrodes. Their focus was evaluation in a lecture and meeting scenario, but the described methods are widely applicable to other real-life situation. They utilized *Independent Component Analysis* (ICA) to identify and drop artifact components in the signal, which has been shown to be very efficient for artifact removal in EEG data (Jung et al. 1997). Features are derived by *Short Time Fourier Transform* (STFT). For two second long segments overlapping by one second, features were calculated representing the content of frequency bands with 0.5Hz width. Since the dimensionality of the feature vector may be large compared to the amount of training data, the authors used a straightforward approach to average over adjacent frequency bands and a correlation-based feature reduction method, which selects those features that correlate best with the variable to be predicted. Finally, *Support Vector Machine*s (SVMs) and *Artificial Neural Network*s (ANNs) for classifi-

Figure 4. Online User State Detection from Brain Activity (EEG).

cation and regression are applied to obtain task demand predictions. For the prediction of low versus high task demand during a presentation the authors obtained 92 % accuracies in session dependent experiments, 87 % in subject dependent but session independent experiments, and 80 % in subject independent experiments. To make brain activity measurements less cumbersome, they built a comfortable headband with which they achieved 69 % classification accuracy for low versus high task demand discrimination. For the reduced number of electrodes, the classification accuracies for half of the subjects are at least 86 % or better, while for the other half they are around chance. This implies that the feasibility of task demand estimation based on four electrodes might be strongly subject dependent or more task dependent, as presentations were tailored towards the educational background of the subjects. The authors also developed an online system that derives user states from brain activity using the headband (Honal and Schultz 2005) that is shown in Figure 4.

6.1.4. Multimodal speaker recognition

With the development of multimedia technologies and improving capabilities of storage devices, the amount of multimedia data which becomes available in large databases and on the Internet increases dramatically – and with it the challenge to archive, maintain, search, and retrieve multimedia data. Recognition technologies provide effective solutions to overcome this problem, for example automatic speech recognition technologies can provide transcriptions of audio samples, image recognition can provide information on peoples' faces, and speaker recognition can help to index and retrieve multimedia data by using peoples' voices and/or faces.

Speaker tracking

Speaker tracking is the task of identifying data sections according to given peoples' voices. This task takes a target speaker and audio files as input and outputs those regions of the audio files that have been uttered by the target speaker. In order to output the identity of a speaker, the target speakers have to be enrolled in the speaker tracking system beforehand, and some training data of the target speakers has to be available for training. Speaker tracking has many potential applications, such as the automatic structuring and indexing of audio data on the web without any manual effort. Such data can then be searched and retrieved by submitting a search query, which may either consist of typing in a person's name or of providing a voice example of the targeted person. Speaker tracking also allows to automatically gather audio snippets of particular speakers, to measure their speaking time contribution, or in combination with automatic speech recognition, to provide a summary on the topics people speak, create speaker-dependent acoustic models for advanced speaker-adaptive speech recognition, and alike.

One interesting domain for which several systems have been proposed is TV broadcasting (Istrate et al. 2005; Moraru et al. 2005; Zibert et al. 2008; Antoni et al. 2006). State-of-the-art systems are traditionally limited to one modality, i.e. the identity of speakers is recognized based either solely on their voices or based solely on their faces. However, using multiple modalities could vastly improve the robustness and the performance of the end system. For example, audio based people tracking performance may suffer severe degradation if the audio is corrupted by environmental noise, channel variability, mismatching training and test conditions, or insufficient audio training material. In contrast, visual based people tracking may suffer from the fact that the face of the person is not always clearly been seen or not visible at all. Thus, fusing information from audio and visual clues has a large potential to improve the overall system (Jain et al. 1999; Ekenel et al. 2007; Erzin et al. 2006). In this subsection we will first introduce a baseline speaker tracking system and then show improvements by fusing this information with video-based face identification, a task which is currently investigated in the project QUAERO, a European multimedia and multilingual program funded by OSEO, French State Agency for Innovation (http://www.quaero.org).

A typical speaker tracking system consists of two main components: speaker segmentation and open-set speaker identification. Given an audio file, speaker segmentation is first applied to remove non-speech regions, followed by segmenting the speech regions based on speaker turn changes. Speech segments are then passed to the open-set speaker identification (SID) system to detect the corresponding speakers. A speaker diarization (SPKD) system serves as a first component, which is to split the speech data into homogeneous segments according to speaker identity. Our example system includes an iterative process and applies speaker identification to perform better segmentation (Li et al. 2009; Hsiao et al. 2008). The full system looks as follows:

- Audio Segmentation finds the regions of speech in the audio stream. This is often implemented by Hidden-Markov-Model (HMM) based segmenters, discriminating main categories, i.e. speech, noise, silence, and music.
- Speaker Change Detection aims at detecting the seamless instant speaker turn-changes which are unable to be detected by the HMM segmenter (Jin and Schultz 2004).
- Bayesian Information Criterion (BIC) Clustering groups the segments which are uttered by the same speakers. It is realized by a hierarchical, agglomerative clustering technique based on Gaussian Mixture Models (GMM) and BIC stopping criterion (Scott and Gopalakrishnan 1998).
- Speaker Model Training uses the output of the BIC clustering. A Universal Background Model (UBM) is trained by pooling the whole test speech segments instead of using external data. Speaker models are Maximum-Aposterior adapted from the UBM on the segments of the corresponding speakers.
- Windowing splits the segments output by the Speaker Change Detection into smaller pieces.
- SID Labeling performs speaker identification on the splitted segments to assign a speaker label to each segment.
- Segment Merging merges adjacent segments with same speaker labels.

Training data that does not contain speech of the target speakers is used to train the UBM models. The Expectation Maximization algorithm is applied to estimate the model parameters. Target speaker models are adapted using the Maximum A Posteriori (MAP) approach (Reynolds et al. 2000) on the speakers' training data from the corresponding UBM model. T-Norm score normalization (Auckenthaler 2000) is applied to the raw scores. Channel and gender classification are typically first applied to the test segments before the detection. Each test segment is scored against both the speaker model and the UBM model during detection, matching the gender and channel condition. Often, *Frame-base Score Competition* (FSC) is performed. The goal of FSC is to combine information from multiple models, and multiple mismatched models are assumed to have the potential of better covering the unknown test space. (Jin et al. 2007) showed significant improvements over the training and testing mismatched case.

Multimodal speaker tracking

In multimodal speaker tracking systems the information available from audio is fused with visual cues in order to improve the overall tracking performance. (Sanderson and Paliwal 2004) categorizes the information fusion into 3 groups, which are refered to as pre-mapping fusion, midst-mapping fusion, and post-mapping fusion. Pre-mapping fusion, also known as "early fusion", is carried out on the feature level in which the features extracted from data of different modalities are concatenated into one feature vector. This technique is rather problematic for audio and video fusion since the frame rates for speech and face recognition systems usually differ substantially. Apart from this, combined features may have high dimensionalities and thus result in the "curse of dimensionality" problem. Midst-mapping aims to exploit the temporal synergies between different data streams. Post-mapping, also known as "late fusion", fuses different systems on the score and decision level. It is preferred by most of multimodal speech and face systems. (Ekenel et al. 2007) presented a multimodal person identification system for smart room environments. The system combines a speaker identification system and a video-based face recognition system. Late fusion is performed on the score level which consists of 3 steps score normalization, modality weighting, and modality combination. Different schemes are investigated and compared for each step. In order to map the confidence scores generated by the speaker and face system into the same distribution, min-max and hyperbolic tangent normalization are applied. In the modality weighting step, two adaptive weighting schemes are introduced which are based on the observations that the difference of the confidence scores between the first hypothesis and second hypothesis is generally smaller in the case of a false classification than in the case of a correct classification. Finally, both weighted modalities are combined. Experimental results show that by fusing the two complementary biometric traits, i.e. speech and face, the overall performance improves significantly over uni-modal systems (Ekenel et al. 2007). In Brady (2008), a GMM based and an SVM based speaker identification system are fused with a face identification system. Score-level fusion is carried out. Normalized scores generated by speaker and face systems are combined respectively, using fixed weights which are estimated on the development set. Again, experiments confirm that the results obtained by the multimodal system outperform uni-modal systems.

Challenges of multimodal speaker recognition

As described above, the fusion of audio and visual cues may improve the overall recognition performance over single-modal systems. However, it also poses several scientific challgenges and raises new issues, such as the question on how to handle those portions of data in which the face of one person appears in the video while the voice of another person is heard. This scenario of non-aligned multimodal information happens frequently, e.g. in broadcast TV when parts of a foreign-language contribution are translated for the audience. In the CLEAR evaluation for the multimodal person recognition task for example, speech had been manually segmented and non-aligned portions were discarded for benchmarking (Ekenel et al. 2007; Brady 2008). However, in reality, multimedia data does not follow these constraints. In order to perform fusion on unconstrained multimedia data for the speaker tracking task, it first has to be detected whether the face and speech are associated to a same person or not. If they are from the same person, the system will usually benefit from fusion. We can simply perform the fusion by using the approaches mentioned above. However, if the speech does not belong to the person appearing in the video scene, combining information from both video and audio will impair the overall performance. One solution to this problem is to perform lip movement detection prior to fusion, and fuse results only when continuous lip movements are detected. Unfortunately, this approach will only allow to determining if somebody speaks at all but not if the speaker is the same person as the one displayed. In the broadcast news domain, one can think of two cases which may fail this approach:

(1) One person speaks in the foreground while another speaker speaks in the background at the same time. This scenario is observed when the person appearing on scene gives a talk or an interview in a foreign language, while the background speaker provides the translation.
(2) The person on screen displays continuous lip movements but no voice of this speaker is heard, meanwhile another talks in the background. This typically occurs when the anchor speaker reports on a meeting which is displayed on screen.

In summary, today's speaker recognition techniques provide effective approaches to index, search, and retrieve vast amounts of multimedia data. Typical challenges are noise and channel effects. Fusing speech with complementary biometric traits, i.e. facial features, improves the overall performance. In this section some work on multimodal systems was reviewied. It showed that multimodal systems benefit from fusion and outperform single-modal systems. However, significant challenges remain for the application of multimodal approaches to unconstrained data, which exhibit non-aligned information.

6.2. Multimodal interfaces

6.2.1. Goals and target applications

Human-Computer Interaction (HCI) is a discipline concerned with the design, evaluation and implementation of interactive computing systems for human use and with the study of major phenomena surrounding them (Hewett et al. 1996). Current human-computer interfaces are quite limited compared to human communication. For example while talking with someone about objects or when reacting emotionally, we use subtle combinations of speech, hand gestures, facial expressions, posture, gaze and proxemic behaviors. The interfaces we use every day are most of the time limited to the use of a keyboard, a mouse and a screen via a classical *Graphical User Interface* (GUI). Moreover, the communication between the user and the computer is mono-directional: the user sends a command to the computer, then the computer responds.

In contrast, human-human communication is *bidirectional, symmetrical and intertwined*. The signals exchanged between two people are interlinked and their relations involve intricate patterns of conversation enabling a continuous management of communicative functions such as feedback and turns (Goodwin 2000; Allwood et al. 2006).

Multimodal interaction seems relevant for a wide range of applications: gaming, education, assistance, web services, interactive maps, graphical editing, and alike.

6.2.2. Background definitions

According to a long-term view, *multimodal human-computer interfaces* should feature the following properties:

– *Bidirectional*: The computer understands the user's combined input of several modalities such as speech and gestures; and the computer generates multimodal behaviors via animated characters; this subsumes perception, interaction and generation capacities in the multimodal system.
– *Symmetrical*: Both the computer and the user are able to use the full range of modalities of human communication (this might enable for example imitation behaviors since the modalities are the same)
– *Intertwined*: Not only the two directions of communication should alternate, but modalities on both sides should be active in a simultaneous interwoven fashion at multiple levels of abstraction as in human communication.

Multimodal interfaces aim at an intuitive (since no training is required) interaction by combining several human-like modalities such as speech and gestures (see also Kopp and Wachmuth 2012 in this volume). The last 20 years have seen the development of research on multimodal interfaces via numerous projects and dedicated conferences. Yet, the community and the field are still quite recent. The first International Conference on Multimodal Interfaces (ICMI) was organized in 1996. The first *Intelligent Virtual Agents* (IVA) international workshop was held in 1998. The first AAMAS Embodied Agents related workshop was held in 2001. The first *Language Resources and Evaluation Conferences* (LREC) workshop on Multimodal Corpora was held in 2000.

6.2.3. Multimodal input: models, tools, methodologies and standards

In a *multimodal input* interface, the computer has to recognize and understand the user's speech and gestures by integrating them at a semantic level which is called multimodal fusion (Coutaz and Caelen 1991) (Figure 5). Multimodal

Figure 5. In future multimodal input systems, users should be able to combine their speech and their gestures to interact in a multimodal way with graphical user interfaces (a prototype of a multimodal map application on the left (Kehler 2000; Cheyer et al. 2001), a prototype for multimodal search on the right (Paek et al. 2008)); a prerequisite to the design of such systems is for the designers to have a representation of the multimodal behaviors that can be expected from the future users.

input interfaces have been developing since the seminal work of Bolt's "Put-That-There" demonstrator (Bolt 1980) which enabled the user to combine a set of spoken commands with pointing gestures of the hand as an interface to a graphical editor.

Several studies and research prototypes have demonstrated the advantages that can bring multimodal input to interactive systems in applications such as tourist maps or mobile contexts (e.g. fusion of user's speech and pen gestures on a PDA). A multimodal interface enables the choice among the available modalities, the one which suits best the environmental constraints, task requirements, and user's preferences or sensory-motor capacities (Carbonell 2003). In spatial tasks, gesture can be used for referring to graphical objects and locations, whereas speech can be used for uttering commands. The joint interpretation of several modalities helps to solve ambiguities in their individual interpretation through "mutual disambiguation" enabling robust and efficient interfaces (Cohen et al. 1997; Oviatt 2002; Kaiser, Olwal et al. 2003). It was also observed that when users were able to speak and gesture with a pen, their utterances contained less disfluencies and were thus easier to process than those observed during a spoken only condition (Oviatt 2002).

The design of these future multimodal input interfaces requires knowledge on how these modalities are related. In order to define appropriate algorithms and software architectures, designers have to know precisely how users combine their speech and gestures when interacting with the system. Designers also need to know if users will be willing to combine their speech and gestures, and if they find it friendly and efficient. One way to collect such knowledge is to build a corpus of users' multimodal behavior with a simulated version of the system.

Multimodal Human-Computer interfaces aim at integrating several modalities into interactive systems. First, we describe related work on multimodal input systems. Second, we survey some *Embodied Conversational Agent* (ECA) systems (Wachsmuth et al. 2008; Rehm and André 2008). Surveys describing both multimodal input human-computer interfaces and ECAs can be found in Benoit et al. (2000) and in Buisine (2005).

Multimodal input prototypes

The 'Put That There' demonstrator (Bolt 1980) enabled editing graphical shapes by combining speech and gesture inputs using various devices such as a joystick, a touchpad or a Polhemus sensor. The user could utter commands like "create a blue square here <gesture>", "move that <gesture> to the right of the green triangle", "put that <gesture> there <gesture>". Since then, several multimodal input prototypes have been designed in various application fields such as map-based applications (Oviatt 2003), crisis management (Sharma et al. 2003), bathroom design (Catizone et al. 2003), logistic planning (Johnston et al. 1997; Johnston 1998), tourist maps (Almeida et al. 2002; Johnston and Bangalore 2004), real estate (Oviatt 1997), graphic design (Milota 2004) or intelligent rooms (Gieselmann and Denecke 2003; Juster and Roy 2004). Some use 2D gesture input (pen, tactile screen), other use 3D input gesture (dataglove, camera-based recognition) (Koons et al. 1993). Several of these systems are task-oriented and sometimes use predefined combinations of modalities (Kaiser et al. 2003).

Time

The temporal relation between two user events detected on two different modalities (e.g. speech and 2D gestures) is often used as a criterion for fusion. Allen defined thirteen temporal relations (Allen 1983; Allen and Ferguson 1994) which are displayed in Figure 6. Such relations enable identification of which part of the gesture is done before, during or after that the whole or part of a vocal utterance. As mentioned by (Buisine 2005), this level of details is not always necessary for representing users' multimodal behaviors.

This temporal fusion needs to consider the possible differences of processing durations between the modalities. Let's take the example of a user uttering a command and then doing a 2D circling gesture with a pen. Although the gesture was done after the spoken utterance, the recognition of its shape will be finished before the recognition of the speech command (Bellik 2001).

This dimension of time is used in a "time and fusion" classification proposed by (Nigay and Coutaz 1993) for describing multimodal input systems resulting in four types of multimodality (exclusive, alternate, synergistic, concurrent).

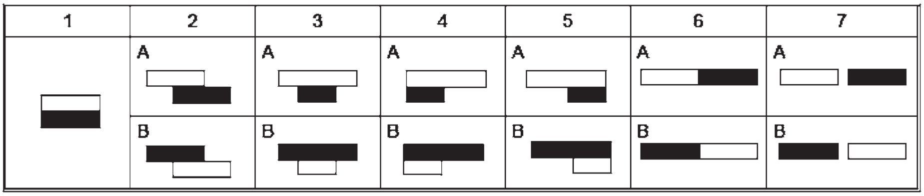


Figure 6. Temporal relations between two elements (Allen 1983).

Multimodal processing

Different fusion criteria are used at different levels in multimodal input systems. These systems often use mono-modal processing as a first step (e.g. recognition of the 2D gesture shape). Such low-level recognition provides information which can carry meanings for some applications: the shape of the 2D gesture (pointing, circling, line, …), the size of the gesture (a short line means that the user wants to select a single graphical object, whereas a long line means that the user wants to connect two graphical objects) (Martin et al. 2006).

The next steps of multimodal processing are system dependent. Some systems use an *early fusion* approach. Such systems do not wait for all the required information (e.g. results of requests sent to other modules such as dialog history). Instead, they merge the results of low-level processing in each modality. This fusion hypothesis might be cancelled later when the system receives a delayed input contradicting the hypothesis. Other systems use a *late fusion* approach and process individually each modality at a semantic level. For example the meaning of a gesture shape can be interpreted with respect to the graphical objects which are close to the gesture location. Such a late fusion is used in the Quickset system (Cohen et al. 1997) and in the dialog controller defined by (Nigay et al. 1993). The Smartkom system uses an early integrated syntactic and semantic processing of the different modalities (Wahlster et al. 2001).

Several representations and fusion techniques are used: frames (Koons et al. 1993) and feature structures (Cohen et al. 1997). Early systems using these techniques were limited (for example a spoken utterance could be related to at most one gesture). A more sophisticated grammar and multidimensional chart parser was defined by (Johnston 1998). It uses extended feature structures including not only the detected events and their evaluation but also the rules to be used for fusion (e.g. logical combinations of temporal and spatial constraints). One of the most successful fusion techniques is *unification*. It is used for fusion in several task-based systems (Johnston 1998). For example, in a graphical design application, if the user utters ("erase this rectangle") and produces a pointing gesture on an object having a rectangular shape, the unresolved speech referent is unified with the gestured object due to their shape compatibility. Other techniques for multimodal fusion were designed such as weighted sum of salience values assigned to all objects in all modalities (Huls et al. 1995) and statistical methods (Wu et al. 1999).

Some multimodal input systems combine several n-best hypotheses produced by speech recognition with several hypotheses produced by gesture recognition. This leads to several possibilities of fusion, each with a score computed as a weighted sum of the recognition scores provided by individual modalities. Mutual disambiguation is a mechanism used in multimodal input systems in which a modality can help a badly ranked hypothesis to get a better multimodal ranking (Kaiser et al. 2003). Thus, multimodality enables the strength of one modality to compensate for weaknesses of others.

Multimodal input systems also face a segmentation problem. For example the system should wait for a given delay before deciding that the previously recognized utterance was mono-modal and will not be followed by an associate gesture.

In (Johnston 1998), *multiple commands* such as " multiple floods zone " can be followed by any number of drawn areas. Each gesture is assigned a timeout feature which indicates how long it can persist. This time is pushed forward every time another gesture is drawn so that multiple commands can persist for as long as the user continues to draw further edges. Such plural behaviors during multimodal input have been studied by (Landragin 2003; Landragin 2006) who suggested using the notion of "perceptual groups" from Gestalt Theory (a perceptual group is defined as several graphical objects sharing common features such as their shapes or their spatial organization).

One problem for a multimodal input system when interpreting a multimodal command is to decide what to do when one of the parameters of the command has not been specified by the user. In Bolt (1980), default values were specified for some parameters (e.g. the size in "create a blue square here") but not for others (e.g. the shape of object has to be specified using either modality). After a certain delay, the system can either ask the user the value of this parameter or cancel the pending interpretation of the current command. In the same system, interpretation of speech was driving the assignment of meaning to the gesture. Early and continuous feedbacks were required to inform the user of the current state of multimodal interpretation (e.g. the color of the cursor changed when the word 'there' was recognized). As pointed to by Cassell et al. (Cassell et al. 1999), gestures in these multimodal input interfaces are mostly used as support for referring expressions, or gesture commands. Several systems are listed in Table 1.

Table 2. Some prototypes managing multimodal fusion and the techniques they use.

System	**Application**	**Modalities**	**Fusion**
Put That There (Bolt 1980)	Manipulation of 2D shapes in a graphical design application	• Speech • Gesture (joystick + pad), sensor on wrist	• Recognition of reference in speech (“that”) is used to launch gesture interpretation • Default value for some attributes (size) and none for others (color). • Example: move that to the right of the green square
CUBRICON (Neal et al. 1989)	Tactical air force mission planning	• Speech input / output • NL text • Maps, tables • Pointing gestures	• The “ Input Coordinator ” module fused the input streams into a single input stream, maintaining the temporal order of tokens in the original streams. This stream is then passed to the multi-media parser/interpreter that produces an interpretation of the input. The parser was a *Generalized Augmented Transition Network* (GATN) that included gesture constituents. The user has to interact sequentially with the system such as in the following example: “What is the status of this <pointing> air base”. The user has to pause before pointing.
XTRA (eXpert TRAnslator) (Wahlster 1991)	Interface to an expert system (fill out a tax form)	• Typed natural language • Mouse gestures • Graphics	• Limited to sequential combinations • Input is interpreted through a unification-based chart parser where pointing gestures are treated as terminal edges in the input stream. • The form has no predefined mouse-sensitive regions, instead regions can overlap or be embedded in other regions.
MATIS (Nigay and Coutaz 1993)	Airline Travel Information System	• Speech • Direct manipulation • Keyboard • Mouse	• Data fusion is performed within a component called dialogue controller which receives objects from another component called the Presentation Techniques Component (NL parser and graphical object manager which manages mouse events). The objects passed by the PTC to the dialog controller obey a uniform format called the melting pot. • Melting pot (structure × time) for merging complementary information. The engine attempts three types of fusion in the following order: micro-temporal fusion for parallel info, macro-temporal fusion for sequential info, contextual fusion for no consideration of temporal constraints.
(Koons et al. 1993)	Interaction with 3D block world (move, modify orientation)	• Speech • Gaze • Hand gesture	• Use gestures as deictic but also as iconic (used for shape of objects, spatial relations and actions) • Parallel information; Integration through frame-like representations • Each incoming stream is assigned a time-stamp, which is later used to realign data from the different sources. • For gestures, hand movements are described according to three features: posture (for each finger), orientation and motion (“gestlet”). Gestures are integrated only if there is evidence in the spoken language that a relevant gesture was made. • The values of a frame can be other frames. Evaluation methods are assigned to different frames (e.g. “below”). Each property of object is represented (color, shape, position).
Ötraf (Johansson 2001)	Local bus traffic and time table	• Speech • Pointing gestures (mouse, pen) • Maps, forms, tables	• Unification of feature structures. • Grammar for specifying multimodal integration patterns.
QuickSet (Cohen et al. 1997)	Simulation set up and control, medical application	• Speech • Pen gestures and writing	• The system uses typed feature structures as a common representational format. • Both the speech recognition agent and the gesture recognition agent generate N-best lists with associated probabilities. • Unification-based. If speech is temporally compatible with gesture (overlap or if gesture slightly precedes speech) the integrator agent takes the sets of interpretations for both speech and gesture, and for each pairing in the product set attempts to unify the two feature structures.

Representation languages for multimodal input interfaces

Whereas early multimodal systems did not use any declarative descriptions of the multimodal commands and of the fusion criteria to apply, more recent multimodal input systems involved the specification of the multimodal combinations to be recognized by the system. These systems also define an XML language for representing messages exchanged between the software modules, for example M3L in the Smartkom system (Wahlster et al. 2001). The EMMA (Extensible MultiModal Annotation markup language) representation language is being defined by the W3C[1] (see also Stührenberg 2012; Trippel 2012 as well as Sasaki and Lommel 2012 all in this volume). It enables the representation of concepts, which are often used: "composite" information (resulting from the fusion of several modalities), confidence scores, timestamps, incompatible interpretations ("one-of").

Collecting and representing users' multimodal behaviors

Although generic principles of architectures are proposed (Carbonell 2005), experimental studies assessing the multimodal behaviors of users are necessary.

The design of a multimodal input interface requires to model:

- the application (often the list of commands and their parameters),
- the spoken behavior of the users
- the gesture behavior of the users (e.g. gesture shapes, singular or plural behaviors)
- the combinations of gesture and speech that the user will display (e.g. timing patterns, semantic relations)

Whereas some multimodal input demonstrators rely on a grammar which does not result of user studies (Kaiser et al. 2003), designing a system for users require a methodological approach for collecting and modeling the above mentioned pieces of information. Several user studies involve a *Wizard of Oz* protocol (Dahlbäck et al. 1993). This protocol consists in designing a simulated version of the system. A hidden experimenter, who is aware of the multimodal behaviors of the user, executes the commands using a classical graphical user interface. The user is video-taped, his actions and results of system's processing are collected in log files. This body of data are then annotated, resulting in a corpus of behaviors which can be used for the design of the system (Mignot and Carbonell 1996; Cheyer et al. 2001). An adaptation of this protocol was defined for the design of mobile multimodal applications (Salembier et al. 2005). Oviatt et al. observed in a pen and speech map application that speech and gesture overlap, or that gesture precedes speech (Oviatt et al. 1997). Table 2 illustrates some of these experimental studies.

The potential malleability of users' multimodal integration patterns is explored in (Oviatt et al. 2003), as well as variation in these patterns during system error handling and tasks varying in difficulty.

Table 3. Some experimental studies of cooperation between modalities in human-computer interaction.

References	**Application**	**Modalities**	**Some results**
(Huls et al. 1995)	File manager	Typed text Mouse pointing gestures	Variety in use of referring expressions
(Oviatt et al. 1997)	Map task	Speech Pen gestures 2D graphics	60 % of multimodal constructions do not contain any spoken deictic
(Oviatt and Kuhn 1998)	Map task	Speech Pen gestures 2D graphics	Most common deictic terms are "her", "there", "this", "that".
			Explicit linguistic specification of definite and indefinite reference is less common compared to speech only.
(Kranstedt et al. 2004)	Assembly task	Speech 3D gestures 3D graphics	Temporal synchronization between speech and pointing Influence of spatio-temporal restrictions of the environment (density of objects) on deictic behavior

Carbonell (Carbonell 2003) reports three wizard of Oz studies in a furniture arrangement application. She compares the collected multimodal behaviors of users (speech and tactile screen) in two conditions: spontaneous versus constrained interaction. The results show that the constrained condition decreased speech disfluencies. The spontaneous condition revealed a stronger use of gesture and multimodality, although subjects in the constraint condition complied rapidly with the constraints. A high inter-individual variability was observed with respect to the quantity of multimodal commands, possibly due to a stimulation of individual creativity in such a design task.

Some of these studies on multimodal behaviors of users are surveyed in (Buisine 2005). The main results concern: 1) the temporal relations between the speech and the gestures displayed by the users, 2) the semantic relations between these modalities, 3) the modalities that users select for each of the commands, 4) the subjective preference of users with respect to the modalities and their combination.

Several studies use video corpora for studying users' multimodal behaviors when interacting with a computer (Table 3).

Table 4. Some corpora of human-computer multimodal interactions.

Reference	Project	Annotated modalities	Application
(Steininger 2000; Steininger et al. 2002)	Smartkom	Users' gestures and emotional expressions (user states)	Information kiosk
(Kvale et al. 2004)	MUST	Users' pen-based pointing gestures using a PDA	Tourist application
(Vuurpijl et al. 2004)	COMIC	Users' pen-based pointing gestures	Bathroom design
(Höysniemi and Hämäläinen 2004)	Wizard of Oz	Children body movement	Game

6.2.4. Multimodal input for interacting with virtual characters

Multimodal output interfaces enable the combination of several audio and visual modalities when presenting information to the user. When they combine media such as text generation and graphic generation, they are called "Intelligent Multimedia Interfaces" (Maybury 1993). We are interested in one of these types of multimodal output interface: the *Embodied Conversational Agent* (ECA) (see also Kopp and Wachsmuth 2012 in this volume). An ECA is a virtual character capable of interacting with users using both verbal and nonverbal behaviors enabling a *face-to-face* interaction with the user. An ECA is thus an interface represented on the screen by a human or cartoon like body, on the audio channel via speech output, and aiming at being conversational in its human-like behaviors: generation of verbal and nonverbal output, management of turn-taking, feedback and repair functions, and also recognition and response to verbal and nonverbal input (Cassell et al. 2000) (Figure 7). Current systems do not address all these ideal features and focus instead on some of them. One of the expected positive features of these ECAs is that they should enable an intuitive communication (since the users know how to interpret speech and gestures in everyday life). They are also supposed to enable the communication of subtle signs (e.g. several non-verbal means to say "No" via speech and facial expressions instead of a mere textual display of "No").

One of the first ECA systems simulated a conversation between two agents in a bank scenario by combining the joint generation of facial expressions, gestures and spoken intonation (Cassell, Pelachaud et al. 1994). Since then, several research prototypes have addressed some of the challenges of ECAs such as the definition of representation languages (Kopp et al. 2006; Krenn et al. 2006) and the generation of expressive behaviors (Pelachaud 2005; Vinayagamoorthy et al. 2006).

The design of these future multimodal output interfaces requires cooperation between several areas of research (Gratch et al. 2002) and the collection of

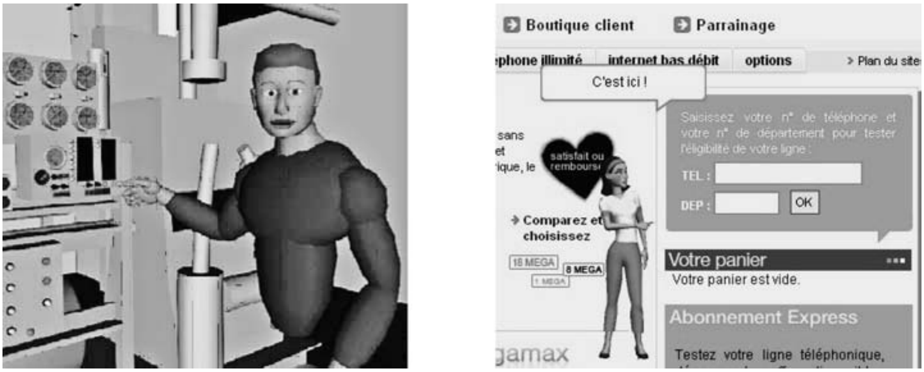


Figure 7. Future multimodal output systems should be able to communicate with users via intuitive modalities (Steve, a prototype of an e-learning application on the left (Johnson et al. 2000); a web assistant on the right); a requisite to the design and acceptability of such systems is for the designers to have a representation of the intuitive multimodal behaviors that the characters should display (e. g. how they should combine gaze, posture and hand gestures for example during deictic behaviors or emotional expressions).

knowledge on how the ECAs should combine the modalities. In order to define appropriate behavioral specifications, algorithms, and software architectures, designers have to know precisely how the ECA should combine its speech, gesture, and facial expression, so that the communication between the user and the system is efficient and smooth. One way to collect such knowledge is to build a corpus of human multimodal behaviors.

6.2.5. Evaluation

Although it was observed that users tend to behave socially with computers displaying social behaviors (Reeves and Nass 1996), multimodal interfaces raise specific questions about their evaluation and the claimed naturalness of the interaction. For example, with respect to multimodal input, the use of a computer media such as the pen, since far from the natural use of hand gestures during human-human communication, might induce cognitive overload (Rugelbak and Hamnes 2003). Furthermore, most of the current multimodal input prototypes are task-oriented as opposed to more 'natural' conversational systems.

Several evaluation studies have demonstrated the advantages that ECAs can bring to interactive systems in applications such as e-learning, edutainment (e.g. the COHIBIT system for a museum exhibit with tangible interaction (Ndiaye et al. 2005)), or educational games for cognitive handicap (Golan and Baron-cohen 2006). In e-learning, pedagogical agents may provide feedback to the student, point to objects in a lesson, or motivate students via expressive behaviors so that it might increase the time children seek to spend with the educational software. Experimental studies about communication and perception are also an application as such since ECAs provide the possibility to turn certain dialogue actions on and off at will (something that is much difficult to do even with a skilled actor). Some other experimental studies did not find any positive advantage of the use of an animated character compared to a classical graphical user interface. Evaluation of ECAs raises several issues such as the methodology, the interaction between the user and the ECA, micro-evaluation studies and applications. Surveys of evaluations of ECAs (including some of the ECAs described below) can be found in (Dehn and van Mulken 2000; Ruttkay and Pelachaud 2004; Buisine 2005). For example, some studies have shown that users can be sensitive to inconsistencies in the ECA's behavior (Nass et al. 2000).

6.2.6. Corpus-based approaches

Although there is a very rich literature on non-verbal communication in the social sciences, it does not provide us with enough contextual and application specific details on the *situated* (i.e. in a specific environmental interactional and application settings) interactive behaviors to display (Goodwin 2000). The goal of multimodal corpora is to provide this situated contextual and detailed knowledge required for the design of multimodal interactive system in a given application context.

A *multimodal corpus* features video recordings of several communication modalities such as speech, hand gesture, facial expression, head movements, body postures, either during human-human or human-computer interaction (Wegener Knudsen et al. 2002; Martin, den Os et al. 2004). A multimodal corpus does not only contain these video recordings, but is an organized set of files including meta-data (recording conditions, context), media files (video / audio), trackers/sensors data, logged files from recognition modules, documentation (annotation guide, protocol description and annotation examples, subject's answers to pre and post-experimental questionnaires) (Figure 8).

Several conferences deal with multimodal corpora: LREC (workshops on multimodal corpora in 2000, 2002, 2004, 2006), Measuring Behavior, Intersensory forum, Interacting Bodies 2005, and Gesture Workshops.

The building of such corpora requires several time-consuming steps such as (Bakeman and Gottman 1997; Montada 2005): identifying the goals of the corpus, studying the related work in social sciences, defining the collection protocol, defining a coding scheme, writing an annotation guide, testing the annotation with a few samples, recruiting annotators, setting the software and hardware architecture (cameras, lights, microphones, network connections), recording the sessions, digitizing the videos, segmenting and formatting these

Figure 8. Multimodal corpora of human-computer interactions integrate several media, manual and automatic annotations for informing the design of multimodal human-computer interfaces: the example of the Smartkom kiosk corpus (Schiel et al. 2002). Questions arise on how to annotate and represent the multimodal behaviors of users in order to design multimodal systems. The same applies to multimodal corpora of human-human interactions which can be used for the design of ECAs.

videos, annotating at multiple levels both manually and automatically, validating these annotations (intra-coder agreement, inter-coder agreement, combination of both), automatic or manual processing of the annotations, computing statistical measures and models. Iterations are required at multiple levels during this process.

Originally, studies of nonverbal communication involved the manual textual annotation of nonverbal behaviors (Goodwin 1981). Gesture studies defined classification of gesture phases and gesture types (McNeill 2005; Kendon 2004). Several computer-based annotation tools have been designed for annotating behaviors observed in videos[2] anchored in time on multiple layers such as Anvil (Kipp 2004), the Observer (Cadée et al. 2002), EXMARaLDA (Schmidt 2009), ELAN (Lausberg and Sloetjes 2009), Vista (Shi et al. 2004). Annotation tools are surveyed in (Dybkjær et al. 2001; Dipper et al. 2004; Dybkjær and Bernsen 2004; Dybkjær and Bernsen 2004; Kipp 2004; Loehr et al. 2005). Annotation can be done with these different tools at several temporal levels: time-based, segments, global video. Annotation models have been defined such as annotation graphs (Bird and Liberman 2001) and the Abstract Corpus Model (Wittenburg et al. 2002). A three level architecture is defined for such annotation tools (physical, logical, application levels) (Bird and Liberman 2001). Systems for coding gestures have been proposed (Gibbon et al. 2004, Gibbon 2009).

Table 5. Illustrative examples of digital corpora of human multimodal communication (LREC Workshops 02, 04, 06).

Source	**Reference**	**Situation**	**Annotations of behaviors (+ speech)**	**Manual annotation**	**Automatic annotation (image processing of motion capture)**	**Goal and means**
Lab	(Campbell and Suzuki 2006)	Meetings	Movements	X		Detect speaker
Lab	(Pianesi, Zancanaro et al. 2006)	Meetings	Functional role Head position/orientation, fidgeting	X	X	Understand group dynamics
Lab	(Caridakis, Raouzaiou et al. 2006)	Acting emotions	Head and hand tracking, Upper-body movement expressivity		X	Image processing of movement and design of expressive ECAs
Lab	(Bailly, Elisei et al. 2006)	Conversation	Facial movements		X	Modeling of conversational speech using motion capture
Lab	(Quek, McNeill et al. 2002; Shi, Rose et al. 2004)	Meetings, conversation about house	Head orientation, gaze		X	Tool for interactive multimodal automatic annotation
Lab	(Beskow, Cerrato et al. 2004)	Elicited dialogues, emotional expressions, phonetic expressions	Lip, eyebrow, cheek, chin, eyelid		X	Audiovisual speech for ECA design
Lab	(Bänziger, Pirker et al. 2006)	Emotions portrayal	Upper body is recorded Emotion	X	X	Controlled study of multimodal expressions of emotion
TV	(Kipp 2004; Kipp, Neff et al. 2006)	Talk show	Conversational gestures	X		Individual ECAs. Economic capture of timing and form of gestures
TV	(Allwood, Cerrato et al. 2006)	Interviews	Feedback, turn management, sequencing Facial display, hand gestures, gaze	X		Multimodal behaviors for the management of interaction.
TV	(Magno Caldognetto, Poggi et al. 2004)	Commercial spot	Gesture, facial (mouth, gaze, eyes, eyebrows), head and body posture, speech, prosody	X		Multilayer musical score method
TV	(Kettebekov, Yeasin et al. 2002)	Weather Narration Broadcast	Pointing gestures, prosody		X	Prosody based co-analysis
Teaching	(Merola and Poggi 2003)	Lecture	Belief, goals, emotion, function (repetitive, …) gesture categories	X		Study communicative style of different teachers
Teaching	(Martell, Osborn et al. 2002; Martell and Kroll 2006)	Lecture	Gestures	X	X	Annotation of gesture kinematics; inter-coder agreement and motion capture

A multimodal musical score

Magno Caldognetto et al. (2004) defined a musical score coding scheme grounded on her model of meanings and goals. It enables the transcription and classification of signals in five modalities: verbal modality (words and sentences uttered), prosodic-intonational modality (speech rhythm, pauses, intensity, stress, intonation), gesture modality (movements of the hands, arms and shoulders), facial modality (head and eye movements, gaze, smile and other facial expressions), body modality (trunk and leg movements, body posture, orientation and movements in space). Five levels of analysis are annotated: signal description (e.g. hand shape and hand movement), signal type (e.g. gesture type such as deictic), a verbal formulation of meaning description, meaning type (according to the meaning classification, e.g. speaker's belief), function (e.g. repetitive). Meaning description, meaning type and function involve two layers of annotation each: literal and indirect. The multimodal score has been implemented in Anvil (Kipp 2004) and applied to video samples of different types: interview, TV news, Italian Sign Language, laboratory experiment, commercial spot.

Gesture annotation for individual gesture generation

Kipp (2004) aims at generating individual gesture behaviors for agents involved in a team of agents. He develops a corpus-based approach for the observation, modeling and generation of conversational gestures. In order to annotate the data he collected, Michael Kipp designed the Anvil annotation tool (Kipp 2001). A word-based speech transcription scheme is defined (including hesitation and non identifiable word). Segments are also coded as well as part of speech categorizing words according to their syntactic role. Discourse annotations include the theme (linking the utterance to the previous discourse, corresponds to a question or topic that is presupposed by the speaker, specifies what the utterance is about), the rheme (relates to the theme, specifies something novel or interesting about the theme), focus (intonationally marked words which enable to distinguish the theme or the rheme from other alternatives that the context makes available). A coding scheme is also defined for annotating gestures on descriptive and interpretative levels (Kipp et al. 2006). Both arms are treated in a single track. A lexicon of gestures is defined according to the function and form of observed gestures. The following classes are used for gesture function: adaptors, emblems, deictics, iconics, metaphorics and beats. A decision tree is used for classification in case of ambiguities between two classes. Inside each of these classes, lemmas (groups of gesture occurrences that are closely related in surface form) are defined for classifying gestures according to their form. The form dimensions used for identifying the lemmas are: hand shape, hand location, hand orientation, movement, bi-handedness, concomitant shoulder movement, concomitant facial expression. Other gesture properties are also annotated using objective description (handedness) and interpretative information (lexical affiliates, and temporal relationship between the gesture and the lexical affiliates). Annotating handedness involves identifying the hand that is an integral part of gesture and finding out if the other hand complements the gesture (e.g. is necessary for the gesture function) or duplicates the gesture. Four descriptive timing patterns are observed and annotated: direct timing (temporal overlap between lexical affiliate and stroke (or multiple beats or an independent hold)), indirect timing (the gesture co-occurs with words that are not lexical affiliates but bear some adjective or negation relation with the lexical affiliates), span timing (a gesture with multiple beats or an independent hold phase covers the length of a whole utterance) and init timing (a short gesture starts with the utterance). High frequency gesture lemmas (i.e. whose instances make up more than 1 % of all occurrences) reveal to cover 85 % of the gesture occurrences.

Gestures and politeness strategies

Rehm and André (2005) collected a video corpus of seminar talks featuring staged criticism given by the audience. Their goal is to study the relations between nonverbal behaviors (including gesture categories) and politeness strategies. They found that gestures were used to strengthen the effect of verbal acts of politeness. Direct strategies were correlated with concrete gestures (deictic and iconic gestures), while metaphoric gestures were more frequent in off record strategies (e.g. vague and indirect form to address the interlocutor).

Image processing of multimodal behavior

Several researchers have developed image processing techniques for analyzing human behavior. Pianesi et al. collected a corpus of consensus decision making meeting (Pianesi et al. 2006). They used manual annotation of the participants' functional role (according to a scheme they propose for task area roles and for socio-emotional roles) and automatic annotation of body activity, head position and orientation, and fidgeting activities. Caridakis et al. (Caridakis et al. 2006) have developed image processing techniques for head and hand tracking, as well as algorithm for computing expressivity parameters (overall activation, spatial extent, temporal extent, fluidity, power, repetitivity). These representations can be used for replaying the gestures by the Greta agent (Hartmann et al. 2005). Grammer et al. (1997) designed a motion energy detection software and applied it to several video corpora to study movement quality during interactions between strangers, and self presentation.

Meetings corpora

There has been a recent increase of interest in the processing of multimodal behaviors during meeting (projects CHIL, AMI, VACE, CALO; ICMI 05 workshop on multimodal multiparty meetings processing[3]). Contributions feature the tracking of meeting participants, meeting recording and playback, annotation of human behaviors.

Campbell and Suzuki analyzed audio and video meeting data (Campbell and Suzuki 2006). They manually annotated the movement of upper body parts. They observed that people do tend to move more when they speak and that there was a steady rise in the amount of movement of all participants preceding the onset of speech.

Looking for natural behaviors

A key issue of multimodal and emotional corpora is the availability of so-called "real-life" data (Batliner et al. 2000; Douglas-Cowie et al. 2003). According to the observer paradox (Grammer et al. 1997), whenever the videotaped persons are aware of the recording device, the naturalness of the collected behaviors is limited. Some comic TV shows use hidden camera but with a reduced context (e.g. scaring people in the street with a false animal), leading mostly to a limited set of nonverbal emotional expressions (e.g. surprise, fear, anger, relief). Grammer et al. (Grammer et al. 1997) use a hidden camera to study social interactions between strangers in a waiting room.

Acting protocols

Clips from movies have been used as behavioral data in several studies about emotional expressions (Clavel et al. 2006; Poggi 2006).

As we described above, there has been a lot of psychological researches on emotional behavior in individual modalities such as speech (Banse and Scherer 1996; Cowie 2000; Schröder 2003), facial expressions (Ekman 1999), and body movements (DeMeijer 1989; Newlove 1993; Boone and Cunningham 1998; Wallbott 1998). Most of these studies about emotion are using "acting" in a restricted sense. Most of them consist in recording, in a laboratory environment, actors who are expressing emotions instructed via single labels of emotions (e.g. anger, disgust, fear, joy, sadness, surprise), sometimes using scripts (Wallbott 1998).

Recent studies in the affective computing area also use similar restricted acting protocol: use of markers on body to recognize four acted basic emotions (Kapur et al. 2005), use of motion capture of static postures during acting of two nuances of four basic emotions (e.g. sad and depressed) (De Silva et al. 2005), use of video processing of facial expressions and upper body gestures during six acted emotional behaviors (Gunes and Piccardi 2005).

Digital multimodal emotional corpora have recently been collected enabling the high resolution and controlled collection of a wide variety of directed and scripted acted gestures, facial expressions and speech emotional behaviors. This is the case of the GEMEP corpus (Bänziger et al. 2006).

Enos and Hirschberg (Enos and Hirschberg 2006) analyze objections and arguments for emotion portrayals. They have collected a corpus of emotional

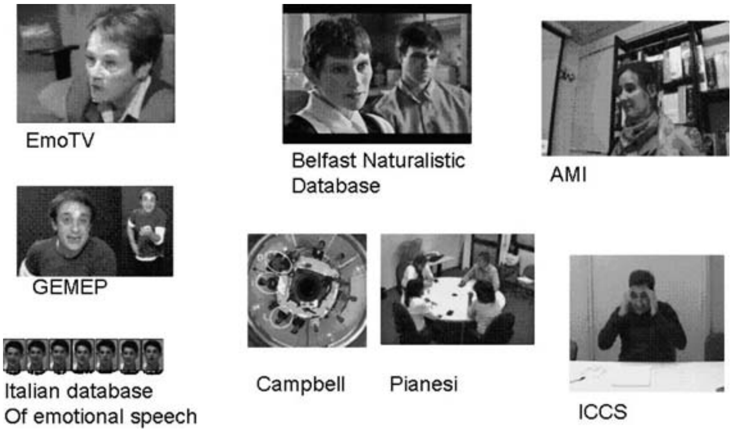


Figure 9. Sample frames some emotional corpora.

behaviors by adding context to emotion portrayals. They suggest two approaches for eliciting emotional speech via acting. The scenario approach uses a description of the character, the situation, the actor's goals and the goals of other characters. The script approach makes direct use of the actor's training and skills via the acting of original scenes taken from theatre plays.

Another mean of involving acting in the study of emotional expression is to have actors replay videotaped behaviors and compare the perceived naturalness of both. In Frigo (2006), naturalistic behaviors collected from TV interviews are replayed by actors. The genuiness and validity of acting is evaluated by naive subjects in order to study the differences in speech between naturalistic versus acted conditions. Several emotional corpora have been described in a dedicated LREC workshop (Devillers et al. 2006). Some emotional corpora are illustrated in Figure 9.

The frequent presence of blends of emotion in such real-life video data was also observed in the Belfast Naturalistic Database (Douglas-Cowie et al. 2005).

Merola and Poggi (2006) collected a video corpus in which athletes were requested to talk about best and worst past performances. At the end of the imagery session, each athlete listed three emotions that characterized the imagined competition. The video were analyzed with the following dimensions: verbal, signal (description of gesture), meaning and type, signal meaning relation, expressivity. They observed that iconic gestures were used to represent visual elements (tracks). There were differences between best and worse performances. Self manipulators and closure gesture (crossing the arms) were more frequent in worse performances.

These different corpora enable the study of several dimensions of emotional corpora that have an impact on the observed behaviors:

- location of recording (laboratory / TV on stage / TV in street / natural environment,e.g., at home);
- instructions staging / spontaneous / portrayed / posed;
- timescale (minutes (TV clip) / hours (meeting) / days);
- awareness of being filmed, awareness of emotion study;
- subjects (actors / experts on emotion / ordinary people);
- interactivity; monologue / multiparty / social constraints on interactivity;
- human-human or human-computer or mixed;
- number of recorded modalities, resolution and quality of collected data; intrusiveness of recording media

Thus, researchers are collecting various corpora that provide complementary insights on multimodal emotional behaviors.

6.3. Conclusions

Multimodal and speech technologies aim at intuitive communication but raise several challenges that we surveyed in this chapter. Multilingual but also multicultural aspects impact on verbal and nonverbal communication. Adaptation, cognitive states processing, and evaluation protocols are required to enable multimodal communication. All these challenges will require substantial advanced and integrated multimodal corpora collection which at the same time will benefit the automatic processing of human behaviors. Future developments include the consideration of affective states of users, use across a variety of a mobile context and consideration of individual differences such as age and culture. A list of recommendations for 'best practices' is difficult to compile due to the wide nature of the multimodality domain. Early consideration of users' behaviors within and across multiple modalities is of main importance.

Acknowledgements

The author of the first section would like to thank Dominic Heger, Felix Putze, Tim Schlippe, and Qian Yang for their valuable contributions.

Notes

1. http://www.w3.org/TR/emma/
2. http://www.ldc.upenn.edu/annotation/gesture/
3. http://www.idiap.ch/ICMI05/programpapers/document_view

References

Ager, S.
2007 Omniglot – writing systems and languages of the world. www.omniglot.com. December 2009.

Allwood, Jens and Elisabeth Ahlsén
2012 Multimodal communication. In: Alexander Mehler, Laurent Romary and Dafydd Gibbon (Eds.): Handbook of Technical Communication. Berlin/Boston: De Gruyter.

Allwood, J. L., Cerrato, L., Jokinen, K., Navaretta, C. and P. Paggio
2006 A Coding Scheme for the Annotation of Feedback, Turn Management and Sequencing Phenomena. In: *Workshop Multimodal Corpora. 5th International Conference on Language Resources and Evaluation (LREC'2006)*, Genova, Italy.

Antoni, W., Fredouille, C. and J.-F. Bonastre
2006 On the use of linguistic information for broadcast news speaker tracking. In: *Proceedings of Acoustics, Speech and Signal Processing (ICASSP 2006).*

Auckenthaler, R.
2000 Score normalization for text-independent speaker verification systems. *Digital Signal Processing* 10: 42–54.

Bailly, G., Elisei, F., Badin, P. and C. Savariaux
2006 Degrees of freedom of facial movements in face-to-face conversational speech. In: *Workshop Multimodal Corpora: from Multimodal Behaviour Theories to Usable Models in Association with the 5th International Conference on Language Resources and Evaluation (LREC2006)*, Genova, Italy.

Bakeman, R. and J. M. Gottman
1997 *Observing Interaction. An introduction to sequential analysis*. Second edition. Cambridge: Cambridge University Press.

Banse, R. and K. Scherer
1996 Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology* 70(3): 614–636.

Bänziger, T., Pirker, H. and K. Scherer
2006 GEMEP – Geneva Multimodal Emotion Portrayals: A corpus for the study of multimodal emotional expressions. In: *Workshop Corpora for research on emotion and affect. 5th International Conference on Language Resources and Evaluation (LREC'2006)*, Genova, Italy.

Bartels, J. L., Andreasen, D., Ehirim, P., Mao, H., Seibert, S., Wright, E. J. and P. R. Kennedy
2008 Neurotrophic electrode: method of assembly and implantation into human motor speech cortex. *Journal of Neuroscience Methods* 174(2): 168–176.

Batliner, A., Fisher, K., Huber, R., Spilker, J. and E. Noth
2000 Desperately seeking emotions or: Actors, wizards, and human beings. In: *Proceedings of ISCA Workshop on Speech and Emotion: A Conceptual Framework for Research*, Newcastle, Northern Ireland.

Becker, T., Blaylock, N., Gerstenberger, C., Kruijff-Korbayov, I., Korthauer, A., Pinkal, M., Pitz, M., Poller, P. and J. Schehl

2006 Natural and intuitive multimodal dialogue for in-car applications: The SAMMIE system. In: *Proceedings of the ECAI Sub-Conference on Prestigious Applications of Intelligent Systems*, Riva del Garda, Italy, 612.

Benoit, C., Martin, J.-C., Pelachaud, C., Schomaker, L. and B. Suhm
2000 Audio-visual and multimodal speech systems. In: D. Gibbon, R. Moore and R. Winski (eds.), *Handbook of Standards and Resources for Spoken Language Systems*. Berlin: Mouton de Gruyter, 102–203.

Berka, C., Levendowski, D. J., Cvetinovic, M. M., Petrovic, M. M., Davis, G., Lumicao, M. N., Zivkovic, V. T., Popovic, M. V. and R. Olmstead
2004 Real-Time Analysis of EEG Indexes of Alertness, Cognition, and Memory Acquired with a Wireless EEG Headset. *International Journal of HCI* 17(2): 151–170.

Beskow, J., Cerrato, L., Granström, B., House, D., Nordstrand, M. and G. Svanfeldt
2004 The Swedish PF-Star Multimodal Corpora. In: *Proceedings of the Workshop Multimodal Corpora: Models Of Human Behaviour for the Specification and Evaluation of Multimodal Input and Output Interfaces in Association with the 4th International Conference on Language Resources and Evaluation (LREC2004)*, Centro Cultural de Belem, LISBON, Portugal.

Birbaumer, N.
2000 The thought translation device (TTD) for completely paralyzed patients. *IEEE Transactions on Rehabilitation Engineering* 8(2): 190–193.

Bird, S. and Liberman, M.
2001 A Formal Framework for Linguistic Annotation. *Speech Communication* 33(1–2): 23–60.

Black, A. W. and K. A. Lenzo
2000 The FestVox Project: Building Synthetic Voices. http://festvox.org/bsv/. December 2009.

Blankertz, B., Dornhege, G., Krauledat, M., Müller, K.-R., Kunzmann, V., Losch, F. and G. Curio
2006 The Berlin brain-computer interface: EEG-based communication without subject training. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 14(2): 147–152.

Bohnenberger, T.; Brandherm, B.; Großmann-Hutter, B.; Heckmann, D. and F. Wittig
2002 Empirically Grounded Decision-Theoretic Adaptation to Situation-Dependent Resource Limitations. In: *Künstliche Intelligenz* 16(3): 10–16.

Bohus, D. and Raux, A. and Harris, T. and Eskenazi, M. and A. Rudnicky
2007 Olympus: an open-source framework for conversational spoken language interface research. In: *Proeceedings of HLT-NAACL (2007) Workshop on Bridging the Gap: Academic and Industrial Research in Dialog Technology*, 32–39.

Bohus, D., Langner, B., Raux, A.; Black, A. W.; Eskenazi, M. and A. Rudnicky
2006 Online supervised learning of non-understanding recovery policies. In: *Proceedings of Spoken Language Technology Workshop*, 170–173.

Bolt, R. A.
1980 Put-that-there: Voice and gesture at the graphics interface. In: *Proceedings of 7th annual International Conference on Computer Graphics and Interactive Techniques*. Seattle, Washington, United States, ACM, 262–270.

Boone, R. T. and Cunningham, J. G.
1998 Children's decoding of emotion in expressive body movement: The development of cue attunement. *Developmental Psychology* 34(5): 1007–1016.
Bos, J. C. and D. W. Tack
2005 Speech input hardware investigation for future dismounted soldier computer systems. DRCD Toronto CR 2005–064.
Brady, K.
2008 MIT Lincoln laboratory multimodal person identification system in the clear 2007 evaluation. In: *Proceedings of Multimodal Technologies for Perception of Humans: International Evaluation Workshops CLEAR 2007 and RT 2007*, Baltimore, MD, USA, May 8–11, 2007, Revised Selected Papers, 240–247.
Brumberg, J.S, Nieto-Castanon, A., Kennedy, P. R. and F. H. Guenther
2010 Brain-Computer Interfaces for Speech Communication, Speech Communication. In: *Special Issue on Silent Speech Interfaces*, April 2010, in press.
Bui, T., Poel, M., Nijholt, A. and J. Zwiers
2007 A tractable DDN-POMDP Approach to Affective Dialogue Modeling for General Probabilistic Frame-based Dialogue Systems. In: *Proceedings of the 5th IJCAI Workshop on Knowledge and Reasoning in Practical Dialogue Systems*, 34–37.
Buisine, S.
2005 *Conception et Évaluation d'Agents Conversationnels Multimodaux Bidirectionnels*. Doctorat de Psychologie Cognitive – Ergonomie, Paris V. 8 avril 2005. Direction J.-C. Martin & J.-C. Sperandio.
Cadée, N., Meyer, E., Theuws, H. and L. Noldus
2002 The Observer Video-Pro: a Versatile Tool for the Collection and Analysis of Multimodal Behavioral Data. In: *Proceedings of Workshop on Multimodal Resources and Multimodal Systems Evaluation, Conference on Language Resources and Evaluation (LREC'2002)*, Las Palmas, Canary Islands, Spain.
Campbell, N. and Suzuki, N.
2006 Working with Very Sparse Data to Detect Speaker and Listener Participation in a Meetings Corpus. In: *Proceedings of Workshop Multimodal Corpora: from Multimodal Behaviour Theories to Usable Models in Association with the 5th International Conference on Language Resources and Evaluation (LREC2006)*, Genova, Italy.
Carbonell, N.
2003 Recommendations for the design of usable multimodal command languages. Universal Access in the Information Society Journal (UAIS). Special issue of *Multimodality, a step towards universal access* 2(2): 143–159.
Caridakis, G., Raouzaiou, A., Karpouzis, K. and S. Kollias
2006 Synthesizing Gesture Expressivity Based on Real Sequences. In: *Proceedings of Workshop Multimodal Corpora: from Multimodal Behaviour Theories to Usable Models in Association with the 5th International Conference on Language Resources and Evaluation (LREC2006)*, Genoa, Italy.

Caridakis, G., Raouzaiou, A., Karpouzis, K. and S. Kollias
2006 Synthesizing Gesture Expressivity Based on Real Sequences. In: *Proceedings of Workshop Multimodal Corpora. From Multimodal Behaviour Theories to Usable Models. 5th International Conference on Language Resources and Evaluation (LREC'2006)*, Genova, Italy.

Carstens
2008 *Carstens Medizinelektronik*. http://www.articulograph.de/ (accessed 6 November 2008).

Cassell, J., Pelachaud, C., Badler, N., Steedman, M., Achorn, B., Becket, T., Douville, B., Prevost, S. and M. Stone
1994 Animated conversation: rule-based generation of facial expression, gesture and spoken intonation for multiple conversational agents. In: *Proceedings of ACM SIGGRAPH'94*.

Cassell, J., Sullivan, J., Prevost, S. and E. Churchill
2000 *Embodied Conversational Agents*. Cambridge: MIT Press.

Chan, A. D. C.
2003 Multi-expert automatic speech recognition system using myoelectric signals. Ph.D. Dissertation, Department of Electrical and Computer Engineering, University of New Brunswick, Canada.

Chen, S. and P. S. Gopalakrishnan
1998 Speaker, environment and channel change detection and clustering via the Bayesian information criterion. In: *Proceedings of the DARPA Broadcast News Transcription and Understanding Workshop*, February 1998.

Cheyer, A., Julia, L. and J.-C. Martin
2001 A Unified Framework for Constructing Multimodal Experiments and Applications. In: H. Bunt, R. J. Beun, T. Borghuis (eds.): *Cooperative Multimodal Communication*. Berlin/New York: Springer: 234–242.

Clavel, C., Vasilescu, I., Devillers, L., Richard, G. and T. Ehrette
2006 The SAFE Corpus: illustrating extreme emotions in dynamic situation. In: *Workshop Corpora for research on emotion and affect. 5th International Conference on Language Resources and Evaluation (LREC'2006)*, Genova, Italy.

Cohen, P. R., M. Johnston, D. McGee, S. Oviatt, J. Pittman, I. Smith, L. Chen and J. Clow
1997 Quickset: Multimodal interaction for distributed applications. In: *Proceedings of the Fith International ACM Multimedia Conference (Multimedia '97)*, Seattle, WA, ACM Press.

Conati, C.,
2002 Probabilistic assessment of user's emotions during the interaction with educational games. *Applied Artificial Intelligence* 16: 555–575.

Coutaz, J. and J.Caelen
1991 A taxonomy for multimedia and multimodal user interfaces. In: *Proceedings of the 1st ERCIM Workshop on Multimodal HCI*, Lisbon.

Cowie, R.
2000 Emotional states expressed in speech. In: *Proceedings of ITRW on Speech and Emotion: Developing a Conceptual Framework for Research*, Newcastle, Northern Ireland.

Crystal, D.
1999 Death sentence. In: *The Guardian G2, 25 October 1999, 2–3*. Adaptation of Prospect Article by The Guardian; further adaptation by Spotlight (March 2000), 14–20.

Dahlbäck, N., Jönsson, A. and L. Ahrenberg
1993 Wizard of Oz studies – Why and how. In: *Proceedings of IUI'93*.

DaSalla, C., Kambara, H., Sato, M. and Y. Koike
2008 Personal communication on EEG classification of vowel speech imagery using common spatial patterns.

De Silva, P. R., Kleinsmith, A. and N. Bianchi-Berthouze
2005 Towards unsupervised detection of affective body posture nuances. In: *Proceedings of the 1st Int. Conf. Affective Computing and Intelligent Interaction (ACII'2005)*, Beijing, China, Springer.

Dehn, D. M. and S. van Mulken
2000 The impact of animated interface agents: a review of empirical research. *International Journal of Human-Computer Studies* 52: 1–22.

DeMeijer, M.
1989 The contribution of general features of body movement to the attribution of emotions. *Journal of Nonverbal Behavior 13*: 247–268.

Denby, B., Schultz, T., Honda, K., Hueber, T. and J.Gilbert
2010 Silent Speech Interfaces. In: *Speech Communication Vol. 52, Issue 4,* April 2010, Pages 270–287.

Denby, B. and M. Stone
2004 Speech synthesis from real time ultrasound images of the tongue. In: *Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '04)*, Montréal, Canada, 17–21 May 2004, volume 1, I685-I688.

Denby, B., Y. Oussar, Y., Dreyfus, G. and M. Stone
2006 Prospects for a Silent Speech Interface Using Ultrasound Imaging. In: *Proceedings of IEEE ICASSP*, Toulouse, France, I365-I368.

Devillers, L., Martin, J.-C., Cowie, R., Douglas-Cowie, E. and A. Batliner
2006 Workshop Corpora for research on emotion and affect. In: *Proceedings of 5th International Conference on Language Resources and Evaluation (LREC'2006)*, Genova, Italy.

Dipper, S., Götze, M. and M. Stede
2004 Simple Annotation Tools for Complex Annotation Tasks: an Evaluation. In: *Proceedings of the Workshop on XML-based richly annotated corpora held during LREC'2004*.

Dornhege, G., del R. Millan, J., Hinterberger, T., McFarland, D. and K.-R. Müller (eds.)
2007 *Towards brain-computer interfacing*. Cambridge: MIT Press.

Douglas-Cowie, E., Campbell, N., Cowie, R. and P. Roach
2003 Emotional speech; Towards a new generation of databases. *Speech Communication* 40: 33–60.

Douglas-Cowie, E., Devillers, L., Martin, J.-C., Cowie, R., Savvidou, S., Abrilian, S. and C. Cox
2005 Multimodal Databases of Everyday Emotion: Facing up to Complexity. In: *9th European Conf. Speech Communication and Technology (Interspeech'2005)*, Lisbon, Portugal.

Dusan, S., Gadbois, G. J. and J. Flanagan
2003 Multimodal Interaction on PDA's Integrating Speech and Pen Inputs. In: *Proceedings of Eurospeech'2003*, Geneva, Switzerland.

Dybkjær, L. and N. O. Bernsen
2002 Data Resources and Annotation Schemes for Natural Interactivity: Purposes and Needs. In: *Workshop on Multimodal Resources and Multimodal Systems Evaluation, Conference on Language Resources And Evaluation (LREC'2002)*, Las Palmas, Canary Islands, Spain.

Dybkjær, L. and N. O. Bernsen
2004 Recommendations for Natural Interactivity and Multimodal Annotation Schemes. In: *Proceedingsof the LREC'2004 Workshop on Multimodal Corpora*, Lisbon, Portugal.

Dybkjær, L. and N. O. Bernsen
2004 Towards General-Purpose Annotation Tools – How far are we today? In: *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'2004)*, Lisbon, Portugal.

Dybkjær, L., Berman, S., Kipp, M., Olsen, M. K., Pirrelli, V., Reithinger, N. and C. Soria
2001 *Survey of Existing Tools, Standards and User Needs for Annotation of Natural Interaction and Multimodal Data*. ISLE Natural Interactivity and Multimodality Working Group Deliverable D11.1.

Dybkjær, L., Bernsen, N. O., Knudsen, M. W., Llisterri, J., Machuca, M., Martin, J.-C., Pelachaud, C., Riera, M. and P. Wittenburg
2003 *Natural Interactivity and Multimodality Working Group Report D9.2. Guidelines for the Creation of NIMM Annotation Schemes*.

Ekenel, H.K., Fischer, M., Jin, Q., Stiefelhagen, R.
2007 Multi-modal person identification in a smart environment. In: *IEEE Conference on Computer Vision and Pattern Recognition 2007 (CVPR '07).*

Ekman, P.
1999 Basic emotions. In: T. Dalgleish and M. J. Power (eds.), *Handbook of Cognition & Emotion,* 301–320. New York: John Wiley.

Enos, F. and J. Hirschberg
2006 A Framework for Eliciting Emotional Speech: Capitalizing on the Actor's Process. In: *Proceedings of the Workshop on Corpora for research on emotion and affect. 5th International Conference on Language Resources and Evaluation (LREC'2006)*, Genova, Italy.

Erzin, E., Yemez, Y., Murat Tekalp, A., Ercil, A., Erdogan, H. and H. Abut
2006 Multimodal person recognition for human-vehicle interaction. *IEEE MultiMedia* 13(2): 18–31.

Fagan, M. J., Ell, S. R., Gilbert, J. M., Sarrazin, E. and P. M. Chapman
2008 Development of a (silent) speech recognition system for patients following laryngectomy. *Medical Engineering and Physics* 30(4): 419–425.

Frigo, S.
2006 The relationship between acted and naturalistic emotional corpora. In: *Proceedings of the Workshop on Corpora for research on emotion and affect. 5th International Conference on Language Resources and Evaluation (LREC'2006)*, Genova, Italy.

Gibbon, D.
2009 Gesture theory is linguistics: modelling multimodality as prosody. In: *Proceedings of PACLIC 23*, Hong Kong.

Gibbon, D., Mertins, I. and Moore, R.
2000 *Handbook of Multimodal and Spoken Dialogue Systems. Resources, Terminology and Product Evaluation*. Berlin/New York: Springer.

Gibbon, D., Trippel, T., Thies, A., Milde, J., Looks, K., Hell, B. and U. Gut
2004 CoGesT: a formal transcription system for conversational gesture. In: *Proceedings of LREC 2004*, Lisbon.

Gibbon, Dafydd, Roger Moore and Richard Winski (eds.)
1997 *Handbook of Standards and Resources for Spoken Language Systems*. Berlin: Mouton de Gruyter.

Gnjatović, M. and D. Rösner
2008 Emotion Adaptive Dialogue Management in Human-Machine Interaction: Adaptive Dialogue Management in the NIMITEK Prototype System. In: *Proceedings of the 19th European Meetings on Cybernetics and Systems Research*.

Golan, O. and Baron-cohen, S.
2006 Systemizing empathy: Teaching adults with Asperger syndrome or high-functioning autism to recognize complex emotions using interactive multimedia. *Development and Psychopathology* 18: 591–617.

Goldberg, L.
1990 An alternative description of personality: The Big-Five factor structure. *Journal of Personality and Social Psychology* 59: 1216–1229.

Goodwin, C.
1981 Conversational organization. Interaction between speakers and hearers. New York: Academic press.

Goodwin, C.
2000 Action and Embodiment Within Situated Human Interaction. *Journal of Pragmatics* 32: 1489–1522.

Gordon, R. G.
2005 *Ethnologue: Languages of the World*. Dallas: SIL International.

Gorin, A. L., Riccardi, G. and J. H. Wright
1997 How may I help you? *Speech Communication* 23: 113–127.

Grammer, K., Filova, V. and M. Fieder
1997 The communication paradox and possible solutions. Towards a radical empiricism. In: A. Schmitt, K. Atzwanger, K. Grammer and K. Schafer (eds.), *New aspects of human ethology,* 91–120. London/New York: Springer.

Gratch, J., Rickel, J., André, E., Badler, N., Cassell, J. and E. Petajan
2002 Creating interactive virtual humans: Some assembly required. *IEEE Intelligent Systems* 4: 54–63.

Gunes, H. and M. Piccardi
2005 Fusing Face and Body Display for Bi-modal Emotion Recognition: Single Frame Analysis and Multi-Frame Post Integration. In: *Proceedings of 1st Int. Conf. Affective Computing and Intelligent Interaction (ACII'2005)*, Beijing, China. Berlin/New York: Springer.

Haag, A., Goronzy, S., Schaich, P. and J. Williams
2004 Emotion recognition using bio-sensors: First steps towards an automatic system. In: Lecture Notes in Computer Science, 3068, 33–48. Berlin/New York: Springer.

Hartmann, B., Mancini, M. and C. Pelachaud
2005 Implementing Expressive Gesture Synthesis for Embodied Conversational Agents. In: *Proceedings of Gesture Workshop (GW'2005)*, Vannes, France.

Hassel, L.,and E. Hagen
2005 Adaptation of an automotive dialogue system to users' expertise and evaluation of the system. In: *Proceedings of the 6th SIGdial Workshop on Discourse and Dialogue*, 222–226.

Healey, J.A. and R. W. Picard
2005 Detecting stress during real-world driving tasks using physiological sensors. *IEEE Transactions on Intelligent Transportation Systems* 6(2): 156–166.

Heracleous, P., Kaino, T., Saruwatari, H. and K. Shikano
2007 Unvoiced speech recognition using tissue-conductive acoustic sensor. *EURASIP Journal on Advances in Signal Processing* 2007(1): 1–11.

Hewett, Thomas T., Baecker, Ronald, Card, Stuart, Carey, Tom, Gasen, Jean, Mantei, Marilyn, Perlman, Gary, Strong, Gary and William Verplank
1996 ACM SIGCHI Curricula for Human-Computer Interaction. New York, ACM.

Hochberg, L. R., Simeral, J. D., Kim, S., Stein, J., Friehs, G. M., Black, M. J. and J. P. Donoghue,
2008 More than two years of intracortically-based cursor control via a neural interface system. In: *Neurosicence Meeting Planner 2008, Program No. 673.15*, Washington, DC.

Holzapfel, H.
2008 A dialogue manager for multimodal human-robot interaction and learning of a humanoid robot. *Industrial Robots Journal* 35(6): 57–67.

Honal, M. and Schultz, T.
2005 Identifying User State using Electroencephalographic Data. In: *Proceedings of the International Conference on Multimodal Input (ICMI)*, Trento, Italy, Oct 2005.

Honal, M. and T. Schultz
2008 Determine Task demand from brain activity. In: *Biosignals* 2008, vol. 1, Pages 100–107, Funchal Madeira, 2008.

Horvitz, E., Breese, J., Heckerman, D., Hovel, D. and K. Rommelse
1998 The lumiere project: Bayesian user modeling for inferring the goals and needs of software users. In: *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*, 256–265.

Hsiao, R., Fuhs, M., Tam, Y.-C., Jin, Q. and T. Schultz
2008 The CMU-Interact 2008 mandarin transcription system. In: *Proceedings of InterSpeech 2008*, Brisbane, Australia, September 2008.

Hueber, T., Aversano, G., Chollet, G., Denby, B., Dreyfus, G., Oussar, Y., Roussel, P. and M. Stone
2007 Eigentongue feature extraction for an ultrasound-based silent speech interface. In: *Proceedings of IEEE ICASSP*, Honolulu, vol. 1, 1245–1248.

Hueber, T., Benaroya, E.-L., Chollet, G., Denby, B., Dreyfus, G. and M. Stone
2010 Development of a silent speech interface driven by ultrasound and optical images of the tongue and lips. *Speech Communication Volume 52, Issue 4,* April 2010, Pages 288–300.

Hummel, J.,Figl, M., Birkfellner,W., Bax, M. R., Shahidi, R., Maurer, C. R. and H. Bergmann
2006 Evaluation of a new electromagnetic tracking system using a standardized assessment protocol. *Physics in Medicine and Biology* 51: N205-N210.

Iqbal, S., Zheng, X. and B. Bailey
2004 Task evoked pupillary response to mental workload in human computer interaction. In: *Proceedings of Conference of Human Factors in Computer Systems (CHI)*, 1480.

Istrate, D., Scheffler, N., Fredouille, C. and J.-F. Bonastre
2005 Broadcast news speaker tracking for ester 2005 campaign. In: *Proceedings of European Conference on Speech Communication and Technology (Interspeech)*, Lisbon, Portugal, September 2005.

Itahashi, Shuichi and Chiu-yu Tseng (eds.)
2010 *Computer Processing of Asian Spoken Languages*. Los Angeles: Consideration Books, Japan, c/o The Americas Group.

Izzetoglu, K., Bunce, S., Onaral, B., Pourrezaei, K. and B. Chance
2004 Functional Optical Brain Imaging Using Near-Infrared During Cognitive Tasks. *International Journal of HCI* 17(2): 211–227.

Jain, A.K., Hong, L., Kulkarni, Y.
1999 A multimodal biometric system using fingerprint, face and speech. In: *Second International Conference on AVBPA*, Washington D.C., USA, 182–187.

Janarthanam, S. and O. Lemon
2008 User simulations for online adaptation and knowledge-alignment in Troubleshooting dialogue systems. In: *Proceedings of 12th SEMDIAL workshop*.

Jin, Q. and T. Schultz
2004 Speaker segmentation and clustering in meetings. In: *Proceedings of ICSLP*, 2004.

Jin, Q., Schultz, T. and A. Waibel
2007 Far-field speaker recognition. *IEEE Transactions on Audio, Speech, and Language Processing* 15:2023–2032.

Johansson, H.
2001 *Integrating Multimodal Information*. MA thesis, Linköping University, Sweden.

Johnson, W. L., Rickel, J. W. and J. C. Lester
2000 Animated Pedagogical Agents: Face-to-Face Interaction in Interactive Learning Environments. *International Journal of Artificial Intelligence in Education* 11: 47–78.

Jorgensen, C. and K. Binsted
2005 Web browser control using EMG based sub vocal speech recognition. In: *Proceedings of the 38th Annual Hawaii International Conference on System Sciences, IEEE*, 294c.1–294c.8.

Jorgensen, C. and S. Dusan
2010 Speech interfaces based upon surface electromyography, *Speech Communication Volume 52, Issue 4*, April 2010, Pages 354–366.

Jorgensen, C., Lee, D. D. and S. Agabon
2003 Sub auditory speech recognition based on EMG signals. In: *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*, 3128–3133.
Jou, S., Schultz, T. and A. Waibel
2007 Multi-stream articulatory feature classifiers for surface electromyographic continuous speech recognition. In: *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, IEEE*, Honolulu, Hawaii.
Jou, S., Schultz, T., Walliczek, M. and F. Kraft
2006 Towards continuous speech recognition using surface electromyography. In: *Proceedings of INTERSPEECH 2006 and 9th International Conference on Spoken Language Processing*, vol. 2, 573–576.
Jung, T.P., Makeig, S., Stensmo, M. and T. J. Sejnowski
1997 Estimating Alertness from the EEG Power Spectrum. *IEEE Transactions on Biomedical Engineering* 4(1): 60–69.
Jurafsky, Daniel and James H. Martin
2009 *Speech and Language Processing: An Introduction to Natural Language Processing, Speech Recognition, and Computational Linguistics*. 2nd edition. New Jersey: Prentice-Hall.
Kaiser, E., Olwal, A., McGee, D. and Benko, H., Corradini, A., Li, X., Cohen, P. and S. Feiner
2003 Mutual disambiguation of 3D multimodal interaction in augmented and virtual reality. In: *Proceedings of the Fith International Conference on Multimodal Interfaces (ICMI'03)*, Vancouver, British Columbia, Canada, ACM Press.
Kapur, A., Kapur, A., Virji-Babul, N., Tzanetakis, G. and P. F. Driessen
2005 Gesture-Based Affective Computing on Motion Capture Data. In: *Proceedings of the 1st Int. Conf. Affective Computing and Intelligent Interaction (ACII'2005)*, Beijing, China, Springer.
Kehler, A.
2000 Cognitive Status and Form of Reference in Multimodal Human-Computer Interaction. In: *Proceedings of the 17th National Conference on Artificial Intelligence (AAAI-2000)*, Austin, Texas.
Kendon, A.
2004 *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press.
Kennedy, P. R., Bakay, R. A. E., Moore, M. M., Adams, K. and J. Goldwaithe
2000 Direct control of a computer from the human central nervous system. *IEEE Transactions on Rehabilitation Engineering* 8(2): 198–202.
Kettebekov, S., Yeasin, M., Krahnstoever, N. and R. Sharma
2002 Prosody Based Co-analysis of Deictic Gestures and Speech in Weather Narration Broadcast. In: *Proceedings of the Workshop on Multimodal Resources and Multimodal Systems Evaluation, Conference On Language Resources And Evaluation (LREC'2002)*, Las Palmas, Canary Islands, Spain.
Kim, K. H., Bang, S. W. and S. R. Kim
2004 Emotion recognition system using short-term monitoring of physiological signals. *Medical and Biological Engineering and Computing* 42: 419–427.

Kipp, M.
2001 Anvil – A Generic Annotation Tool for Multimodal Dialogue. In: *Proceedings of the 7th European Conference on Speech Communication and Technology (Eurospeech'2001)*, Aalborg, Danemark.

Kipp, M.
2004 *Gesture Generation by Imitation. From Human Behavior to Computer Character Animation*. Florida, Boca Raton, Dissertation.com.

Kipp, M., Neff, M. and I. Albrecht
2006 An Annotation Scheme for Conversational Gestures: How to economically capture timing and form. In: *Proceedings of the Workshop Multimodal Corpora: from Multimodal Behaviour Theories to Usable Models in Association with the 5th International Conference on Language Resources and Evaluation (LREC2006)*, Genova, Italy.

Klien, G., Woods, D., Bradshaw, J., Hoffman, R., Feltovich, P. and K. Associates
2004 Ten challenges for making automation a team player. *IEEE Intelligent Systems*: 91–95.

Koons, D. B., C. J. Sparrell and K. R. Thorisson
1993 Integrating simultaneous input from speech, gaze, and hand gestures. In: M. T. Maybury (ed.), *Intelligent Multimedia Interfaces,* 243–261. Menlo Park, CA: AAAI/MIT Press.

Kopp, S., Krenn, B., Marsella, S., Marshall, A., Pelachaud, C., Pirker, H., Thórisson, K. and H. Vilhjalmsson
2006 Towards a Common Framework for Multimodal Generation: The Behavior Markup Language. In: *Proceedings of 6th International Conference on Intelligent Virtual Agents (IVA'06)*, Marina del Rey, CA: Springer.

Kopp, Stefan and Ipke Wachsmuth
2012 Artificial interactivity. In: Alexander Mehler, Laurent Romary and Dafydd Gibbon (eds.): *Handbook of Technical Communication*. Berlin/Boston: De Gruyter.

Krahnstoever, N., Schapira, E., Kettebekov, S. and Sharma, R.
2002 Multimodal human-computer interaction for crisis management systems. In: *Proceedings of IEEE Workshop on Applications of Computer Vision*, 203–207.

Kranstedt, A., P. Kühnlein and I. Wachsmuth
2004 Deixis in Multimodal Human-Computer Interaction. In: *Proceedings of Gesture-Based Communication in Human-Computer Interaction. 5th International Gesture Workshop, GW 2003*, Genova, Italy. A. Camurri and G. Volpe, Springer. LNAI 2915.

Krenn, B., Lamolle, M., Pelachaud, C., Pirker, H. and M. Schröder
2006 *D6e: Report on Representation Languages*. Workpackage 6 Deliverable.

Larsson, S. and D. Traum
2000 Information state and dialogue management in the TRINDI Dialogue Move Engine Toolkit. In: *Natural Language Engineering Special Issue on Best Practice in Spoken Language Dialogue Systems Engineering*, 323–340. Cambridge: Cambridge University Press.

Lausberg, H. and H. Sloetjes
2009 Coding gestural behavior with the NEUROGES-ELAN system. *Behavior Research Methods, Instruments, & Computers* 41(3): 841–849.

Lauth, Codrina, Bettina Berendt, Bastian Pfleging and Albrecht Schmidt
2012 Ubiquitous computing. In: Alexander Mehler, Laurent Romary and Dafydd Gibbon (eds.): *Handbook of Technical Communication*. Berlin/Boston: De Gruyter.

Li, R., Schultz, T., Jin, Q.
2009 Improving speaker segmentation via speaker identification and text segmentation. In: *Proceedings of Interspeech 2009*.

Li, X. and Ji, Q.
2005 Active affective state detection and user assistance with dynamic bayesian networks. *IEEE Transactions on Systems, Man and Cybernetics* 35: 93–105.

Litman, D. and S. Silliman
2004 Itspoke: An intelligent tutoring spoken dialogue system. In: *4th Meeting of the North American Chapter of the Association for Computational Linguistics (HLT/NAACL)*, Boston, USA, 233–236.

Loehr, D., Rohlfing, K. and S. Duncan
2005 Comparison of multimodal annotation tools. In: *Proceedings of Interacting Bodies, 2nd ISGS Conference*. (available via http://vislab.cs.vt.edu/~gesture/multimodal/workshop/index.htm.)

Magno Caldognetto, E., Poggi, I., Cosi, P., Cavicchio, F. and G. Merola
2004 Multimodal Score: an Anvil Based Annotation Scheme for Multimodal Audio-Video Analysis. In: *Proceedings of Workshop Multimodal Corpora: Models of Human Behaviour for the Specification and Evaluation of Multimodal Input and Output Interfaces in Association with the 4th International Conference on Language Resources And Evaluation (LREC2004)*, LISBON, Portugal.

Maier-Hein, L., Metze, F., Schultz, T. and A. Waibel
2005 Session independent non-audible speech recognition using surface electromyography. In: *Proceedings of IEEE Workshop on Automatic Speech Recognition and Understanding*, San Juan, Puerto Rico, 331–336.

Mairesse, F. and M. Walker
2005 Learning to personalize spoken generation for dialogue systems. In: *Proceedings of Interspeech 2005 – Eurospeech*, 1881–1884.

Mairesse, F. and M. Walker
2006 Words mark the nerds: Computational models of personality recognition through language. In: *Proceedings of the 28th Annual Conference of the Cognitive Science Society (CogSci)*, 543–548.

Mana, N., Lepri, B., Chippendale, P., Cappelletti, A., Pianesi, F., Svaizer, P. and M. Zancanaro
2007 Multimodal Corpus of multi-party meetings for automatic social behavior analysis and personality traits detection. In: *Proceedings of the ACM workshop on Tagging, mining and retrieval of human related activity information*, 9–14.

Manabe, H., Hiraiwa, A. and T. Sugimura
2003 Unvoiced speech recognition using EMG-mime speech recognition. In: *Proceedings of CHI, Human Factors in Computing Systems*, Ft. Lauderdale, Florida, 794–795.

Manabe, H. and Z. Zhang
2004 Multi-stream HMM for EMG-based speech recognition. In: *Proceedings of 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 1–5 Sept. 2004, San Francisco, California, volume 2, 4389–4392.

Martell, C. and J. Kroll
2006 Using FORM Gesture Data to Predict Phase Labels. In: *Proceedings of Workshop Multimodal Corpora: from Multimodal Behaviour Theories to Usable Models in Association with the 5th International Conference on Language Resources and Evaluation (LREC2006)*, Genova, Italy.

Martell, C., Osborn, C., Friedman, J. and P. Howard
2002 The FORM Gesture Annotation System. In: *Proceedings of Workshop on Multimodal Resources and Multimodal Systems Evaluation, Conference On Language Resources And Evaluation (LREC'2002)*, Las Palmas, Canary Islands, Spain.

Martin, J.-C., den Os, E., Kühnlein, P., Boves, L., Paggio, P. and R. Catizone
2004 *Workshop Multimodal Corpora: Models of Human Behaviour for the Specification and Evaluation of Multimodal Input and Output Interfaces. In Association with the 4th International Conference on Language Resources and Evaluation LREC2004*, Centro Cultural de Belem, LISBON, Portugal.

Martin, J.-C., Kühnlein, P., Paggio, P., Stiefelhagen, R. and F. Pianesi
2006 *Workshop Multimodal Corpora: from Multimodal Behaviour Theories to Usable Models in Association with the 5th International Conference on Language Resources and Evaluation (LREC2006)*, Genova, Italy.

Maybury, M.
1993 *Intelligent Multimedia Interfaces*. Menlo Park, CA: MIT Press.

Maybury, M. and J.-C. Martin
2002 *Workshop on Multimodal Resources and Multimodal Systems Evaluation, Conference On Language Resources And Evaluation (LREC'2002)*, Las Palmas, Canary Islands, Spain.

McNeill, D.
2005 Gesture and thought: codification and social interaction. In: *Proceedings of Embodied Communication I. Opening Conference of the ZiF: Research Group 2005/2006 Embodied Communication in Humans and Machines*. Scientific organization: Ipke Wachsmuth (Bielefeld), Günther Knoblich (Newark).

Mehl, M., Gosling, S. and J. Pennebaker
2006 Personality in its natural habitat: Manifestations and implicit folk theories of personality in daily life. *Journal of Personality and Social Psychology* 90(5): 862.

Merola, G.
2006 Gestures while reporting emotions. In: *Proceedings of the Workshop on subsystem synchronization and multimodal behavioral organization held during Humaine summer school*. Genova, Italy.

Merola, G. and I. Poggi
2003 Multimodality and gestures in the Teacher's communication. In: *Proceedings of the Gesture Workshop 2003*, Genova, Italy.

MIT Media Lab Human Dynamics Group
2004 *Speech feature extraction toolbox*, http://groupmedia.media.mit.edu/data.php.

Montada, L.
2005 Video analysis and the interactional relevance of bodily details. In: *Panel presentation. Interacting Bodies, 2nd ISGS Conference*, Lyon, France, ENS LSH.

Moraru, D., Ben, M. and G. Gravier
2005 Experiments on speaker tracking and segmentation in radio broadcast news. In: *Proceedings of Interspeech 2005 – Eurospeech*, Lisbon, Portugal, September 2005.

Nakajima, Y., Kashioka, H., Campbell, N. and K. Shikano
2006 Non-audible murmur (NAM) recognition. *IEICE Transactions on Information and Systems* E89-D(1): 1–8.

Nakajima, Y., Kashioka, H., Shikano, K. and N. Campbell
2003 Non-audible murmur recognition input interface using stethoscopic microphone attached to the skin. In: *Proceedings of IEEE ICASSP*, 708–711.

Nasoz, F. and C. L. Lisetti
2007 Affective user modeling for adaptive intelligent user interfaces. In: *12th HCI International*, Beijing, China.

Nass C. and K. M. Lee
2000 Does computer-generated speech manifest personality? An experimental test of similarity-attraction. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*, The Hague, The Netherlands, 329–336.

Nass, C., Jonsson, I.-M., Harris, H., Reaves, B., Endo, J., Brave, S. and L. Takayama
2005 Improving automotive safety by pairing driver emotion and car voice emotion. In: *CHI '05 extended abstracts on Human factors in computing systems*, Portland, USA, 1973–1976.

Nass, C., Steuer, J. and E. Tauber
1994 Computers are social actors. In: *Proceedings of the ACM SIGCHI conference on Human factors in computing systems: celebrating interdependence*, 72–78.

Ndiaye, A., Gebhard, P., Kipp, M., Klesen, M., Schneider, M. and W. Wahlster
2005 Ambient Intelligence in Edutainment: Tangible Interaction with Life-Like Exhibit Guides. In: Proceedings of INTETAIN 2005. Berlin/New York: Springer.

Neal, J. G., C. Y. Thielman, Z. Dobes, S. M. Haller and S. C. Shapiro
1989 Natural language with integrated deictic and graphic gestures. In: *DARPA Workshop on Speech and Natural Language*.

Newlove, J.
1993 *Laban for actors and dancers*. New York: Routledge.

Ng, L., Burnett, G., Holzrichter, J. and T. Gable
2000 Denoising of human speech using combined acoustic and EM sensor signal processing. In: *Proceedings of the Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, Istanbul, Turkey, 5–9 June 2000, vol. 1, 229–232.

Nigay, L. and J. Coutaz
1993 A design space for multimodal systems: concurent processing and data fusion. In: *Proceedings of the ACM INTERCHI'93*, Amsterdam.

Oviatt, S. L.
2002 Breaking the Robustness Barrier: Recent Progress on the Design of Robust Multimodal Systems. In: M. Zelkowitz (ed.): *Advances in Computers,* 56. New York/London: Academic Press.

Oviatt, S., Coulston, R. and R. Lunsford
2004 When do we interact multimodally? Cognitive load and multimodal communication patterns. In: *Proceedings of the 6th International Conference on Multimodal Interfaces*, 129–136.

Paaß, Gerhard
2012 Document classification, information retrieval, text and web mining. In: Alexander Mehler, Laurent Romary and Dafydd Gibbon (eds.): *Handbook of Technical Communication*. Berlin/Boston: De Gruyter.

Paek, T.
2006 Reinforcement learning for spoken dialogue systems: Comparing strengths and weaknesses for practical deployment. In: *Proceedings of the Dialog-on-Dialog Workshop, Interspeech*, 683–696.

Paek, T., Thiesson, B., Ju, Y. and B. Lee
2008 Search Vox: Leveraging Multimodal Refinement and Partial Knowledge for Mobile Voice Search. In: *Proceedings of ACM UIST 2008*, 141–150.

Patil, S. A. and J. H. L. Hansen
2010 A Competitive Alternative for Speaker Assessment: Physiological Microphone (PMIC). *Speech Communication, Special Issue on Silent Speech Interfaces*, April 2010, in press.

Pelachaud, C.
2005 Multimodal expressive embodied conversational agent. In: *Proceedings of ACM Multimedia, Brave New Topics session*, Singapore.

Pervin, L. and O. John
2001 Personality: Theory and Research, 8th ed. New York: Wiley.

Pianesi, F., Mana, N., Cappelletti, A., Lepri, B. and M. Zancanaro
2008 Multimodal recognition of personality traits in social interactions. In: *Proceedings of the ACM 10th international conference on Multimodal interfaces*, 53–60.

Pianesi, F., Zancanaro, M. and C. Leonardi
2006 Multimodal Annotated Corpora of Consensus Decision Making Meetings. In: *Workshop Multimodal Corpora: from Multimodal Behaviour Theories to Usable Models in Association with the 5th International Conference on Language Resources and Evaluation (LREC2006)*, Genova, Italy.

Picard, R.W., Vyzas, E. and J. Healey
2001 Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23: 1175–1191.

Pleydell-Pearce, C. W., Whitecross, S. E. and B. T. Dickson
2003 Multivariate Analysis of EEG: Predicting Cognition on the basis of Frequency Decomposition, Inter-electrode Correlation, Coherence, Cross Phase and Cross Power. In: *Proceedings of 38th HICCS*.

Poggi, I.
2006 Social influence through face, hands, and body. In: *Proceediongs of the Second Nordic Conference on Multimodality*, Goteborg, Sweden.

Porbadnigk, A., Wester, M., Calliess, J. and T. Schultz
2009 EEG-based speech recognition – impact of temporal effects. In: *Proceedings of Biosignals 2009*, Porto, Portugal, January 2009, 376–381.

Precoda, K. Zheng, J., Vergyri, D., Franco, H., Richey, C., Kathol, A. and S. Kajarekar
2007 IraqComm: A next generation translation system. In: *Proceedings of Interspeech 2007*, 2841–2844.

Preuss, R.D., Fabbri, D.R., Cruthirds, D. R.
2006 Noise robust vocoding at 2400 bps. In: *Proceedings of the 8th International Conference on Signal Processing, ICSP 2006*, Guilin, China, November 16–20, 2006, vol. 1, 16–20.

Prommer, T.; Holzapfel, H. and A. Waibel
2006 Rapid Simulation-Driven Reinforcement Learning of Multimodal Dialog Strategies in Human-Robot Interaction. In: *Proceedings of Interspeech 2006*.

Putze, F. and T. Schultz
2009a Cognitive Dialog Systems for Dynamic Environments: Progress and Challenges. In: *Proceedings on the 4th Biennial Workshop on DSP for In-Vehicle Systems and Safety*, Dallas, USA, TX.

Putze, F. and T. Schultz
2009b Cognitive Memory Modeling for Interactive Systems in Dynamic Environments. In: *Proceedings of the 1st International Workshop on Spoken Dialog Systems*, Germany.

Quatieri, T.F., Messing, D., Brady, K, Campbell, W. B., Campbell, J. P., Brandstein, M., Weinstein, C. J., Tardelli, J. D. and P. D. Gatewood
2006 *Exploiting nonacoustic sensors for speech enhancement*. In: *IEEE Transactions on Audio, Speech, and Language Processing* 14(2): 533–544.

Quek, F., McNeill, D., Bryll, R., Duncan, S., Ma, X.-F., Kirbas, C., Mccullough, K. E. and R. Ansari
2002 Multimodal human discourse: gesture and speech. *ACM Transactions on Computer-Human Interaction* 9(3): 171–193.

Rauschert, I., Agrawal, P., Sharma, R., Fuhrmann, S., Brewer, I. and A. MacEachren
2002 Designing a human-centered, multimodal GIS interface to support emergency management. In: *Proceedings of the 10th ACM international symposium on Advances in geographic information systems*, 119–124.

Raux, A., Langner, B., Black, A. and M. Eskenazi
2005 Let's go public! Taking a spoken dialog system to the real world. In: *Interspeech 2005 (Eurospeech)*, Lisbon, Portugal.

Reeves, B. and C. Nass
1996 *The media equation: How people treat computers, television, and new media like real people and places*. New York: Cambridge University Press.

Rehm, M. and E. André
2005 Informing the Design of Embodied Conversational Agents by Analyzing Multimodal Politeness Behaviors in Human-Human Communication. In: *AISB Symposium for Conversational Informatics*.

Rehm, M. and E. André
2008 From Annotated Multimodal Corpora to Simulated Human-Like Behaviors. In: Wachsmuth, Ipke andGünther Knoblich (eds.), *Modeling Communication with Robots and Virtual Humans,* 1–17. Berlin: Springer.

Reynolds, D. A., Quatieri, T. F. and R. B. Dunn
2000 Speaker verification using adapted gaussian mixture models. In: *Digital Signal Processing*, 2000.

Rieser, V. and O. Lemon
2008 Automatic Learning and Evaluation of User-Centered Objective Functions for Dialogue System Optimisation. In: *Proceedings of the Sixth International Language Resources and Evaluation*, 2536–2361.

Rugelbak, J. and K. Hamnes
2003 Multimodal Interaction – Will Users Tap and Speak Simultaneously? *Telektronikk* 2: 118–124.

Ruttkay, Z. and C. Pelachaud
2004 *From Brows to Trust – Evaluating Embodied Conversational Agents*. Kluwer: Boston.

Sanderson, C. and K. K. Paliwal
2004 Identity verification using speech and face information. *Digital Signal Processing* 14:449–480.

Sasaki, Felix and Arle Lommel
2012 Multilingual computing. In: Alexander Mehler, Laurent Romary and Dafydd Gibbon (eds.): *Handbook of Technical Communication*. Berlin/Boston: De Gruyter.

Schäfer, R.; Weis, T.; Weyrath, T. and A. Jameson
1997 Wie können Ressourcenbeschränkungen eines Dialogpartners erkannt und berücksichtigt werden? In: *Kognitionswissenschaft* 6: 151–164.

Schiel, F., Steininger, S., Beringer, N., Türk, U. and S. Rabold
2002 Integration of Multi-modal Data and Annotations into a Simple Extendable Form: the Extension of the BAS Partitur Format. In: *Proceedings of the Workshop on Multimodal Resources and Multimodal Systems Evaluation, Conference On Language Resources And Evaluation (LREC'2002)*, Las Palmas, Canary Islands, Spain.

Schönle, P. W., Gräbe, K., Wenig, P., Höhne, J., Schrader, J. and B. Conrad
1987 Electromagnetic articulography: Use of alternating magnetic fields for tracking movements of multiple points inside and outside the vocal tract. *Brain and Language* 31: 26–35.

Schröder, M.
2003 Experimental study of affect burst. *Speech Communication. Special Issue following the ISCA Workshop on Speech and Emotion* 40(1–2): 99–116.

Schultz, T.
2002 GlobalPhone: A Multilingual Speech and Text Database developed at Karlsruhe University. In: *Proceedings of ICSLP*, Denver, CO.

Schultz, T.
2004 Towards Rapid Language Portability of Speech Processing Systems. In: *Conference on Speech and Language Systems for Human Communication*, New Delhi, India, 2004.

Schultz, T. and K. Kirchhoff (eds.)
2006 *Multilingual Speech Processing*. New York: Academic Press.

Schultz, T. and M. Wand
2010 Modeling coarticulation in large vocabulary EMG-based speech recognition. *Speech Communication Volume 52, Issue 4*, April 2010, Pages 341–353.

Schultz, T. and Schaaff, K.
2009 Towards Emotion Recognition from Electroencephalographic Signals. In: *3rd International Conference on Affective Computing and Intelligent Interaction* (ACII 2009), 1–6.

Schultz, T., Black, A., Badaskar, S., Hornyak, M. and J. Kominek
2007 SPICE: Web-based tools for rapid language adaptation in speech processing systems. In: *Proceedings of Interspeech*, Antwerp, Belgium, Aug. 2007, 2125–2128.

Shi, Y., Rose, T. and F. Quek
2004 A System for Situated Temporal Analysis of Multimodal Communication. In: *Workshop Multimodal Corpora: Models of Human Behaviour for the Specification and Evaluation of Multimodal Input and Output Interfaces in Association with the 4th International Conference on Language Resources and Evaluation LREC2004*, Centro Cultural de Belem, LISBON, Portugal.

Singh, S., Litman, D., Kearns, M. and M. Walker
2002 Optimizing dialogue managment with reinforcement learning: Experiments with the NJFun system. *Journal of Artificial Intelligence Research* 16: 105–133.

Smith, M., Gevins, A., Brown, H., Karnik, A. and R. Du
2001 Monitoring Task Loading with Multivariate EEG Measures during Complex Forms of Human-Computer Interaction. *Human Factors* 43(3): 366–380.

Stiefelhagen, R., Fügen, C., Gieselmann, P., Holzapfel, H., Nickel, K. and A. Waibel
2004 Natural human-robot interaction using speech, head pose and gestures. In: *Proceedings of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*.

Stührenberg, Maik
2012 Foundations of markup languages. In: Alexander Mehler, Laurent Romary and Dafydd Gibbon (eds.): *Handbook of Technical Communication*. Berlin/Boston: De Gruyter.

Suppes, P., Lu, Z.-L. and B. Han
1997 Brain wave recognition of words. *Proceedings of the National Academy of Scientists of the USA* 94: 14965–14969.

Sutton, R. S. and A. G. Barto
1998 *Reinforcement Learning: An Introduction*. Cambridge: MIT Press.

Takahashi, K.
2004 Remarks on SVM-based emotion recognition from multi-modal bio-potential signals. In: *Robot and Human Interactive Communication ROMAN (2004), 13th IEEE International Workshop on Robot and Human Interactive Communication*, 95–100.

Tapus, A., Tapus, C. and M. J. Matari
2007 User-robot personality matching and assistive robot behavior adaptation for post-stroke rehabilitation therapy. *Journal of Intelligent Service Robotics, Special Issue on Multidisciplinary Collaboration for Socially Assistive Robotics*: 169–183.

Tardelli, J. D. (ed.)
2004 *MIT Lincoln Labs report ESC-TR-2004–084, Pilot Corpus for Multisensor Speech Processing*.

Titze I. R., Story, B. H., Burnett, G. C., Holzrichter, J. F., NG, L. C. and W. A. Lea
2000 Comparison between electroglottography and electromagnetic glottography. *Journal of the Acoustical Society of America* 107(1): 581–588.

Tran, V.-A., Bailly, G., Loevenbruck, H. and T. Toda
2010 Improvement to a NAM-captured whisper-to-speech system. *Speech Communication Volume 52, Issue 4*, April 2010, Pages 314–326.

Traum, D. and S. Larsson
2003 The Information State Approach to Dialogue Management. In: Jan C. J. van Kuppevelt and R. W. Smith (eds.), *Current and New Directions in Discourse & Dialogue*, 325–354. Dordrecht: Springer.

Trippel, Thorsten
2012 Controlled language structures in technical communication. In: Alexander Mehler, Laurent Romary and Dafydd Gibbon (eds.): *Handbook of Technical Communication*. Berlin/Boston: De Gruyter.

Van Hooff, J. A. R. A. M.
1982 Categories and sequences of behavior: Methods of description and analysis. In: K. R. Scherer and P. Ekman (eds.), *Handbook of methods in nonverbal behavior research,* 362–439. Cambridge: Cambridge University Press.

Vinayagamoorthy, V., Gillies, M., Steed, A., Tanguy, E., Pan, X., Loscos, C. and M. Slater
2006 Building Expression into Virtual Characters. In: *Eurographics Conference State of the Art Reports*.

Vuurpijl, L., L. ten Bosch, S. Rossignol, A. Neumann, N. Pfleger and R. Engel
2004 Evaluation of multimodal dialog systems. In: *Workshop Multimodal Corpora: Models of Human Behaviour for the Specification and Evaluation of Multimodal Input and Output Interfaces in Association with the 4th International Conference on Language Resources and Evaluation (LREC2004)*, Centro Cultural de Belem, LISBON, Portugal.

W3C Consortium
2007 *Voice Extensible Markup Language 1.2 (VoiceXML)*, http://www.w3.org/TR/voicexml21. December 2009.

Wachsmuth, I., M. Lenzen and G. Knoblich
2008 *Embodied Communication in Humans and Machines*. Oxford: Oxford University Press.

Wahlster, W.
1991 User and discourse models for multimodal communication. In: Sullivan J. W. and S. W. Tyler (eds.): *Intelligent User Interfaces*. New York: ACM Press.

Wahlster, W. (ed.)
2006 *SmartKom: Foundations of Multimodal Dialogue Systems*. Berlin: Springer Science+Business Media.

Walker, M. A.
2005 Can We Talk? Methods for Evaluation and Training of Spoken Dialogue Systems. In: *Language Resources and Evaluation* 39(1): 65–75.

Walker, M. A., Fromer, J. C. and S. Narayanan
1998 Learning optimal dialogue strategies: A case study of a spoken dialogue agent for email. In: *Proceedings of the 17th international conference on Computational linguistics*, Montreal, Quebec, Canada, 1345–1351.

Walker, M. A., Litman, D. J., Kamm, C. A. and A. Abella
1997 Paradise: a framework for evaluating spoken dialogue agents. In: *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics*.

Wallbott, H. G.
1998 Bodily expression of emotion. *European Journal of Social Psychology* 28: 879–896.

Walliczek, M., Kraft, F., Jou, S.-C., Schultz, T. and A. Waibel
2006 Sub-word unit based non-audible speech recognition using surface electromyography. *Proceedings of Interspeech 2006*, Pittsburgh, USA, 1487–1490.

Wand, M. and T. Schultz
2009 Speaker-adaptive speech recognition based on surface electromyography. In: *Biomedical Engineering Systems and Technologies, Communications in Computer and Information Science,* Springer, 2010, Volume 52, Part 3, Pages 271–285.

Wegener Knudsen, M., Martin, J.-C., Dybkjær, L., Berman, S., Bernsen, N. O., Choukri, K., Heid, U., Kita, S., Mapelli, V., Pelachaud, C., Poggi, I., van Elswijk, G. and P. Wittenburg
2002 *Survey of NIMM Data Resources, Current and Future User Profiles, Markets and User Needs for NIMM Resources*. ISLE Natural Interactivity and Multimodality. Working Group Deliverable D8.1.

Wegener Knudsen, M., Martin, J.-C., Dybkjær, L., Machuca Ayuso, M.-J., Bernsen, N. O., Carletta, J., Heid, U., Kita, S., Llisterri, J., Pelachaud, C., Poggi, I., Reithinger, N., van Elswijk, G. and P. Wittenburg
2002 *Survey of Multimodal Annotation Schemes and Best Practice*. ISLE Natural Interactivity and Multimodality. Working Group Deliverable D9.1.

Wester, M. and T. Schultz
2006 *Unspoken speech – speech recognition based on electroencephalography.* Master's thesis, Universität Karlsruhe (TH), Karlsruhe, Germany.

Wittenburg, P., Broeder, D., Berck, P., Sloetjes, H. and A. Klassmann
2006 Language Resource Archiving supporting Multimodality Research. In: *Workshop Multimodal Corpora: from Multimodal Behaviour Theories to Usable Models in Association with the 5th International Conference on Language Resources and Evaluation (LREC2006)*, Genova, Italy.

Wittenburg, P., Broeder, D., Offenga, F. and D. Willems
2002 Metadata Set and Tools for Multimedia/Multimodal Language Resources. In: *Workshop on Multimodal Resources and Multimodal Systems Evaluation, Conference on Language Resources and Evaluation (LREC'2002)*, Las Palmas, Canary Islands, Spain.

Wittenburg, P., Kita, S. and H. Brugman
2002 Cross-Linguistic Studies of Multimodal Communication. In: *Workshop on Multimodal Resources and Multimodal Systems Evaluation, Conference on Language Resources And Evaluation (LREC'2002)*, Las Palmas, Canary Islands, Spain.

Wolpaw, J. R., Birbaumer, N., McFarland, D., Pfurtscheller, G., Vaughan, T.
2002 Brain-computer interfaces for communication and control. *Clinical Neurophysiology* 113(6): 767–791.

Zibert, J., Vesnicer, B. and F. Mihelic
2008 Development of a speaker diarization system for speaker tracking in audio broadcast news: a case study. *Journal of Computing and Information Technology* 16: 183–195.

7. Resources for technical communication systems

Dafydd Gibbon

1. System resource requirements

1.1. Overview of topics

Technical communication systems are defined here as devices or device networks which intervene in the communication channel between speaker and addressee. Technical communication contrasts with face-to-face communication. The devices or device networks may be audio alone or audio-visual (multimodal), and standalone devices (such as computers with software for word processing, lexicon databases, dictation or satellite navigation) or complex systems such as telephone networks and chat or voice-over-internet-protocol (VoIP) communication on the internet (see also Allwood and Ahlsén 2012; Martin and Schultz 2012; Lücking and Pfeiffer 2012; all in this volume).

The topic of resources for technical communication systems is extensive and complex; the present article in the context of a general handbook therefore does not aim at providing detailed recipes for resource specification, compilation and use, but concentrates on generic considerations, and focuses on two specific cases: for text-based communication systems on lexicographic resources, and for speech-based communication systems on speech synthesis systems. References to specialised handbooks and other relevant literature are made at the appropriate places.

There are many general considerations in connection with resource oriented topics in the technical communication system area which are common to a wide range of system development areas. Among these are issues of reusability, interoperability over different platforms, cost-effectiveness, use-case and scenario dependency, as well as data collection paradigms such as the ‘crowd-sourcing’ of data on the internet from arbitrary or selected internet users, and ‘cloud sourcing’, the out-sourcing of resources, tools and other systems to internet-based resources, tools and systems.

A major issue is standardisation of categories and formats of resources for information exchange, either as a local set of consistent conventions, or as conformance to *de facto* standards set by influential institutions and companies (such as formats for media and word processor files), or to internationally agreed standards defined by the *International Standards Organisation* (ISO[1]) (see also Trippel 2012 in this volume). A little known standard which is relevant for technical communication systems is, for example, the language code and name standard ISO 639–3, in which information such as the following (for English and its names in other major languages) is recorded: “eng English English anglais inglés 英语 английский Englisch”. In many practical applications, country names have been used instead of language names, which can lead to confusion.

Many generic requirements such as these are currently in a state of rapid fluctuation, and will probably remain so. Consequently they are only referred to but not expounded in detail in the present article.

Similarly, legal issues concerning intellectual property rights, patents and trademark registration play a central role in creating and providing resources; specific issues in speech and multimodal communication concern data protection issues associated with the ease of identification of voices, and the even greater ease of identification of faces.

Ethical issues are also involved, not only in the deployment of systems, but also in the compilation of data resources, the extremes being data collected without consent of the recorded parties, and data recorded with explicit, informed and signed consent.

Issues such as these must be addressed in practice, but can only be mentioned and not handled here because of their variety, complexity and task-dependence, and also because of their often national and culture-specific character.

1.2. Systems and resources

Like any other scientific and engineering enterprise, the development of speech systems, language systems and multimodal systems (referred to here in brief as Human Language Technology systems, HLT systems) is dependent on the availability of adequate empirical, technical and human resources for their development. This three-way distinction between empirical resources (for instance texts, recordings in different media), technical resources (tools for processing empirical resources) and human resources will be maintained in the present contribution wherever necessary.

In the present context, the discussion of resources for the technical transmission components of such systems (e.g. resources for the encoding, transmission and decoding of acoustic signals, hearing aid technology, optical character recognition, font implementation) are largely excluded. The focus is on resources for components which are specifically within the linguistic and phonetic sub-domains of HLT systems.

The terms ‘system’ and ‘resource’ in the context of technical communication are illustrated informally in this section, and then treated selectively but in more detail later. Where more detail is required, specialised literature with comprehensive information is recommended. Some of these technical communication systems can be realised as standalone systems, others are embedded in larger systems which are not *per se* communication systems. Several contributions to the present volume discuss different kinds of system and resources for these systems. System types and specific resource types are also discussed in most other contributions to the present handbook.

Technical communication systems as understood here include but are not limited to the following, not all of which can be discussed in the present context:

1. Speech input systems: speech-to-text recognition systems, dictation applications, machine and user interface control systems, including prosthetic systems such as voice command systems for motor-impaired users.
2. Speech output systems: geographical information system output, dictation readback components, prosthetic systems such as screen readers for the blind, visualisers for the deaf.
3. Speech dialogue systems: human-machine interaction in tutorial systems, information systems, scheduling and booking systems.
4. Natural language systems: information retrieval (search and parsing) components, database-to-text generation components.
5. Multimodal systems: tutorial systems with avatars, robotic systems with embodied agents, video games, assistive systems with Braille and other tactile output, map-based geographical information systems, systems with other sensors (e.g. airflow, skin resistance, gesture).

Resources for systems such as these include the following:

1. Raw data:
   a. Audio, visual and synchronised audio-visual recordings of interactions in standardised audio and visual formats.
   b. Handwriting, print, keyboard, stylus and finger touch screen input streams in a variety of formats.
2. Annotated data:
   a. Transcriptions of raw data in symbolic notations, produced either manually or automatically.
   b. Annotations of raw data, in which individual segments (tokens) in transcriptions are associated with time-stamps, i.e. temporal pointers (and/or space-stamps, with print) in the raw data.
3. Generalised data:
   a. *Lexicons* (alternative plural: 'lexica') or dictionaries, i.e. inventories of basic language unit (word, idiom) types, for each of which multiple tokens can occur in the data. Each unit type is associated with further kinds of lexical information (typically: phonemic, morphological, syntactic, semantic, pragmatic). A principled distinction between dictionary and lexicon is not made here. A useful informal distinction is often made between *semasiological lexicons* or readers' dictionaries (the most fam-

iliar kind, in which a wordform is known and the meaning is to be found) and *onomasiological lexicons* or writers' dictionaries (in which a known concept, represented by a known word, is known, and a wordform is to be found), for example a *thesaurus* (plural: thesauri) based on a concept hierarchy, or a *terminological dictionary*. Other organisational forms of lexicon are pronunciation dictionaries and multilingual dictionaries, which are not easily classified as semasiological or onomasiological in the usual senses of the terms. Lexicons are dealt with in more detail in Subsection 2.2.

b. *Grammars*, i.e. rule sets which determine the co-occurrence of segment types (not only words but also syllables, speech sounds) with each other, either in sequence (as with words in a sentence) or in parallel (as with sentences and intonation). Some grammatical information about local constraints on word sequences, parts of speech (POS), etc. is typically also encoded in the lexicon. In practical systems, straightforward finite state automata (or regular grammars) formalisms are often used (Beesley & Karttunen 2003); for some theoretical linguistic purposes such formalisms are too restricted.

c. Statistical *language models*, i.e. pairs of segments or segment sequences with probabilities or sets of probabilities, as in *Hidden Markov Models* (HMM) or *Artificial Neural Nets* (ANN). From a linguistic point of view, a language model is a special case of a grammar, either an extremely simple Finite State Automaton or Regular Grammar with probability-weighted transitions (as in statistical diphone and triphone models), or in more sophisticated probabilistic Context Free Grammar, in which nodes in hierarchical structures containing word and sentence constituents have probabilistic weights (for further detail consult Jurafsky & Martin 2000 and Carstensen et al. 2010; see also Martin and Schultz 2012 as well as Paaß 2012 in this volume).

1.3. Intellectual resources

1.3.1. Notations

Intellectual resources are notations, symbolisms, formalisms, interfaces, i.e. means of representing data, facts, figures, models and theories. Some of these resources are standardised in order to facilitate exchange of information, some are introduced *ad hoc* for specific, often temporary purposes, or in order to support competitive development of proprietary systems, and others, e.g. text document formats such as PDF (Adobe) and RTF (Microsoft Corporation) and audio formats such as WAV (Microsoft Corporation) are proprietary notations respectively which have become *de facto* standards (see also Rahtz 2012 in this volume).

Well-known open intellectual resources range from the *International Phonetic Alphabet* (IPA[2]; easily accessible on the internet) for representing speech sounds, through tagsets for parts of speech (van Halteren 1999), to notations for predicate logics, attribute logics, and to the hierarchy of formal languages and formal grammars which underlies syntax formalisms as well as parsers and generators for these. Some of these are regulated by the community, some are regulated by specific bodies, for example as HTML (by W3C[3]) and IPA (by the *International Phonetic Association*); on standard notations for speech, see Gibbon et al. (1997).

In discussion of resources, it is convenient to make a distinction between specific and generic data representation notations.

1.3.2. Specific data representation notations

By 'data' is meant observable or in general physical objects and their properties. A data representation notation provides a model for describing data in a systematic way, and may or may not be related to an explicit and coherent formal theory.

Classic cases of data representation notations are phonetic alphabets such as the IPA, which provide an exhaustive and consensually standardised vocabulary (see also Trippel 2012 in this volume) for representing speech sounds and their properties. The IPA is not based on an explicit and coherent theory of speech sound production, transmission or perception, but has been developed pragmatically since the late 19th century in terms of its empirically demonstrated usefulness. Ostensibly, the IPA is based on the physiological constraints on speech sound production, and the consonant chart comes very close to reaching this goal. However, the vowel diagramme is better explained by acoustic theory. Tone and intonation, on the other hand, are represented by icons for percepts. Terms for phonation types such as 'breathy' and 'creaky' are auditory metaphors.

In a somewhat more general sense, the HTML tree graph structured text representation language is domain-specific data because it is designed for text data, not other data types. HTML has a 'semantic interpretation' in terms of actual formats in terms of the CSS (*Cascading Style Sheet*[4]) language, which defines more specific features of the physical appearance of texts. Other data types, such as complex media objects like videos and graphics are not included in the HTML formalism, but HTML includes a pointer ('anchor') concept for linking to these. The pointer concept enables the construction of arbitrary, not necessarily tree-structured texts, i.e. hypertexts. HTML and CSS are formally defined: both express tree structures with sets of attributes attached to the tree nodes, each attribute associated with a numerical or textual value. The semantics of HTML expressions is given a relatively general definition in terms of text

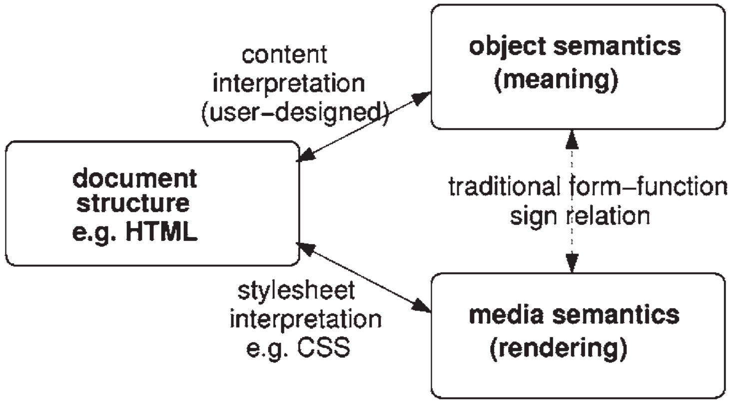


Figure 1. Text Content – Structure – Rendering (CSR) model.

rendering, i.e. appearance, but not in terms of conventional object semantics, i.e. 'meaning' in the usual sense of the term, which has a different kind of semantic interpretation function outside the text and remains the concern of the user. CSS expressions provide a more detailed rendering interpretation. The relation between conventional semantic interpretation of meaning and 'semantic' interpretation of rendering properties is illustrated in Figure 1.

At the most detailed text level of the character or letter, the internationally standardised data representation notation is Unicode, which provides a uniform numerical encoding system for a wide range of standardised alphabets for languages, data description notations, and logics. The media semantics of Unicode entities are, as with HTML and CSS, given in terms of glyphs, i.e. renderings of characters as defined by specific font categories. The object semantics of Unicode entities is left to the intuition of the user, as with HTML (Figure 1), and is frequently inconsistent. For example, the IPA 'semantics' of phonetic properties is not coherently expressed: characters are not kept within a well-defined coherent code-page, but, where they are similar to other characters, e.g. Latin characters, they are defined elsewhere in the Unicode character set. There is no layer of abstraction which includes the object semantics of the characters.

1.3.3. Generic data representation notations

The term 'generic' in 'generic data representation notation' indicates that this notion of 'data' is not restricted to observable or physical objects, but may capture any kind of entity. While HTML and its parent formalism SGML, for example, were specifically related to texts and text-like objects when they were introduced, its successor XML is generic (see Stührenberg 2012 in this volume): XML can be and is used to represent any kind of object and its properties, from observable data through texts, archives, and computer programmes (however, very many domain-specific versions of XML have been developed for special purposes, such as VoiceML for speech synthesis objects).

Like HTML, XML expresses tree structures with sets of attributes attached to the tree nodes, each attribute associated with a numerical or textual value. Where other, more complex data structures are required, additional implicit or explicit notational conventions are required. Where an XML document is linked to another entity (such as another XML document), far more complex graph structures can be created, for example. These structures are extrinsic to XML and need other means for monitoring, consistency checking, parsing, etc., than the context-free (or even finite state) parsers which are appropriate for tree analysis.

The following example shows the structure of an automatically generated XML archive document based on an interview (abbreviated with "…"; names, attributes and values modified for publication):

```
<?xml version="1.0" encoding="UTF-8"?>
<accident COST="NEGLIGIBLE" PERSONNUMBER="3">
<title>Walking incident</title>
<report> Yesterday I … </report>
<participant gender="Mr" firstName="An" initialName=""
   lastName="Other" institution="Home News Ltd."
   contact="none" involvement="Victim" role="Witness"/>
<episode-collection>
<episode-item type="ITEM">Yesterday I …</episode-item>
<episode-item type="ITEM">and …</episode-item>
…
</episode-collection>
</accident>
```

An XML 'element' or 'object' has a body consisting of a string of text whose start and end are delimited by tags; the tags are delimited by angle brackets, and the start tag may contain attribute-value pairs representing properties of the element. A second type of element consists of a single tag; an example of this is the first line of the example is a special tag for the element "xml" with metadata about the XML version. The "accident" element which follows occupies the rest of the example, with the start tag on the second line and the end tag on the last line. In the start tag, there are two attribute-value pairs, first the attribute "COST" with the value "NEGLIGIBLE", second the attribute "PERSONNUMBER" with the value "3". Embedded in the body is a series of other elements, "title" (full element), "report" (full element), "participant" (tag only) and "episode-collection" (full element). The element "episode-collection" contains a series of more deeply embedded elements, all with the name "episode-item". The elements "accident", "participant" and "episode-item" contain attribute-value pairs. The depth of embedding is not restricted (in principle, the tree can contain recursion, which would be required, for example, if sentences with subordinate clauses were to be represented). The overall tree graph in this example, of depth three, can be recognised; each node ('element') is associated with a flat tree of attribute-value pairs, of depth two.

The tree graph data structure represented by XML imposes restrictions; not all data types can be comfortably represented by tree structures. Embedded tables, for example, are not simple tree structures, because there are additional constraints on the agreement of the width of rows in the table; theoretically, embedded tables can be represented by indexed context-sensitive languages (a special kind of Type 1 formal language which is more complex than a Type 2 or context-free formal language (cf. Hopcroft et al. 2006). However, in practice this formal property has to be 'faked' in the processing algorithm or by the human designer: the even-branching fan-out needs to be additionally calculated (or provided manually).

Other generic data representation notations are used in standard database technologies, the most prominent type being the linked relation tables used in *Relational Database Management Systems* (RDBMS).

1.4. Operational resources

Operational resources are the tools for manipulating data, and the underlying algorithms on which the implementations are based. Tools are task specific, and therefore depend not only on the modality choice but also on the scenario choice involved. The focus will be on software tools such as parsers and visualisation software. Hardware tools (such as specific computers, audio and video recording devices, specialised input and output devices such as special keyboards, touch screens, braille pads and printers) will not be dealt with. A rough categorisation of operational resources will be discussed below.

2. Resources for text systems

2.1. Informal overview

The main resources for text-based systems, which include information retrieval services of most kinds, are, in general, large collections of texts (for many purposes harvested from the internet), and the search tools for investigating the composition of these texts, whether standalone non-linked texts or hypertexts. Such tools ranging from the 'find and replace' string search function of text editors and word processors to the keyword oriented search of help systems and the keyword plus heuristics (popularity; advertising; 'Did you mean …?') full text search of internet browsers. These tools, as found for instance in the development and uses of a word processor, utilise resources based on many levels of linguistic and computational linguistic knowledge.

First a prominent example from text linguistics is constituted by the 'styles' or format templates which determine both the text structure and appearance of a document and require parsing tools which can handle the categories involved. In computational linguistic terms, language units such as characters, words, sentences, paragraphs, documents (articles, books etc.) are assigned appropriate attribute-value structures (feature structures) which need form the basis of text parsers which in turn are used to assign the text rendering or 'appearance semantics' in print. Character codes are assigned implementations in fonts (nowadays typically Unicode) with visual font properties (glyphs and highlighting attributes). The parsing of larger units of language such as words implies the recognition of word boundaries and (for hyphenation), the recognition of the internal structure of words in terms of characters, syllables and morphemes, phonological and morphological analyses. Word prediction resources and spelling correctors require dictionaries. The parsing of sentences for capitalisation, selection and grammar checking requires the recognition of sentence constituents and sentence boundaries. The handling of paragraphs demands a facility for handling their properties such as left, right, top, bottom boundaries (i.e. margins). The most ubiquitous text unit for current word processors is the paragraph: any title, heading, caption, etc. is typically handled as a special type of paragraph, distinguished from other paragraph types by differences in paragraph properties such as top and bottom spacing, left and right indenting, line alignment (left, right, centre, justified), as well as by fonts and their attributes.

Second, an independent formatting layer determined not by language properties but by media properties must be handled. The page structure of a book requires the handling of line-breaks, page-breaks. The page structure of a newspaper requires in addition the non-linear handling of article breaks and continuations. The 'megastructure' of a dictionary requires the handling of cross-references. The constraints on a scientific paper to provide supporting evidence for the content may require a word processing system to provide automatic handling of cross-references for the table of contents, figures (and lists of figures), tables (and lists of tables), footnotes, term indices and bibliographical references. The file architecture of a hypertext network requires the handling of link anchors and targets. At character level, the format layer requires the handling of kerning (character spacing and overlap), ligatures (blends of more than one character) and diacritics (accent marks).

A special case of a text system is the lexicon database, as the basic resource either for a printed dictionary or encyclopaedia, or for a hyperlexicon on the internet or as part of a help system. This example will be dealt with in a separate subsection.

2.2. Example of resource creation for text-based systems – lexicography

Lexicography is the scientific and technological discipline concerned with dictionaries, lexicons, and encyclopaedias. Lexicographic methods are taken partly from the humanities (in particular the language sciences – the 'art of lexicography'), and partly from computational linguistics. Many technical communication systems contain lexicons – or, indeed actually are lexicons – and a lexicon is arguably the most complex linguistic component of a technical communication system (cf. van Eynde et al. 2000).

Theoretical lexicography is concerned with the structure of lexicons and with types of lexical information; the study of types of lexical information is also known as lexicology. Applied lexicography is concerned with the analysis, creation and evaluation of dictionaries, lexicons and encyclopaedias. A dictionary itself may be regarded as a system, most clearly when the dictionary is stored as an electronic database, processed with operational lexicographic resources for distribution on CD-ROM or DVD, or on the internet. In the present discussion, no distinction will be made between dictionary, lexicon and encyclopaedia; for discussion of this distinction reference should be made to the extensive lexicographic literature (see the final section).

The media in which a dictionary is implemented, the architecture of the dictionary and the requirements which are to be met by the architecture and the implementation will be determined by the dictionary use cases and, more specifically, by the dictionary market. A non-exhaustive list of typical use cases might (not including embedded lexicon subsystems) include the alphabetic dictionary (organised by wordforms), the thesaurus and the synonym dictionary (organised by meaning), the idiom dictionary, the bilingual dictionary, the pronunciation dictionary, the rhyming dictionary, the concept-based terminological dictionary.

2.2.1. Lexicon resource structure

In generic terms, any dictionary is a set of *lemmas* (singular: 'lemma'; alternative plural: 'lemmata') organised in a specific well-defined *macrostructure* such as a list or a tree hierarchy, the lemmas each being associated with a well-defined *microstructure* of *data categories*. Additionally, lemmas may be interlinked with cross-references and additional explanations; the cross-references constitute the *mesostructure*. The overall structure of the dictionary, together with its published metadata and perhaps also any additional explanatory information is sometimes referred to as the *megastructure*.

Consequently, simplifying the issue, lexicographic resources must first of all contain specifications of the megastructure, macrostructure, microstructure and mesostructure in terms of the desired use cases. For practical applications, each kind of structure requires its own particular combination of empirical, technical and human resources. The classic case of the semasiological alphabetic dictionary will be taken as an example of such specifications, and the structure types of such dictionaries (which in general apply, *mutatis mutandis*, to other dictionary types) will be outlined as follows.

1. *Macrostructure*. The macrostructure of the alphabetic dictionary is a list of *headwords* or *lemmas* sorted alphabetically, each being the first element in a *lexical entry* or *article* which otherwise contains lexical information about the headword. Macrostructures have certain specific features:
   a. Attention must be paid to the sort order; the traditional ASCII sort order is inadequate in the context of international Unicode conventions, and needs to be specified explicitly in each case. While alphabetic sorting is adequate for many languages with alphabetic orthography, it is evidently less adequate for syllabic scripts and inadequate for logographic scripts. For languages with a very small set of lexicalised prefixes (many African languages), simple alphabetic arrangement is also inadequate.
   b. The alphabetic dictionary is a variety of *semasiological* dictionary, in which the headword represents a wordform and the lexical information concerns the meaning of the wordform. The inverse relation is found in *onomasiological* dictionaries like the thesaurus, where the headword represents a known concept and the lexical information concerns the wordform.
2. *Microstructure*. In the simplest case, the microstructure of a lexicon is an ordered list of types of lexical information. The microstructure of a technical dictionary resource is the most complex part, and also the most difficult to standardise, despite cooperative efforts going back many decades (Atkins et al. 2008; van Eynde et al. 2000). These types of lexical information (also known among computational lexicographers as '*data categories*') concern the following main properties of words:
   a. Word *form* (spelling and hyphenation; pronunciation and prosody, e.g. stress or tone).
   b. Word *structure* (internal: prefixes, suffixes, constituent words in compounds; external: part of speech, grammatical restrictions).
   c. Word *meaning* (descriptive components such as abstractness, animacy, pragmatic components such as style, taboo).
   d. Inter-article *cross-references* (to synonyms, antonyms, examples, sources, etc.).
   e. In a *lexical database*, also *metadata* about the lexicographer, date of processing, comments).
   f. *Hierarchical information*: in more complex cases, the microstructure can be hierarchical, organised as a set of related sub-entries, typically words derived or compounded from the same root.

3. *Mesostructure.* The cross-references in the dictionary constitute a more or less explicit network of relationships between words. The following kinds of relation or network structure may be noted:
   a. The main relations are typically between synonyms and antonyms.
   b. A lexical system such as a WordNet (Fellbaum 1998) uses an elaborated and explicit version of this kind of cross-reference structure as its macrostructure.
   c. In addition, implicit cross references are made by the use of category names such as the parts of speech: a term such as ‘noun’, or a pronunciation transcription, is not explained for each entry, but reference must be made to the introductory sketch grammar in the megastructure of the dictionary.
   d. An interesting formal feature of the mesostructure of lexicons is often *cyclicity* in cross-references, usually unintended. The ‘cross-reference depth’ of the mesostructure of a dictionary could contain, for example, the following: the word ‘thing’ is defined in terms of the nearest kind ‘object’, the word ‘object’ is defined in terms of the nearest kind ‘entity’, and the word ‘entity’ is defined in terms of the nearest kind ‘thing’. Without references to external examples, this cyclicity is inevitable.
4. *Megastructure*. The megastructure defines the structure of the overall implementation of the complete dictionary: the actual organisation in a book, a database or on a website. The most straightforward case is the book: the front matter, including the title page, publication metadata page (with date, author and copyright and publisher details), foreword and preface, table of contents; the sketch grammar; the body (lemma-based list of articles); back matter (e.g. publisher’s advertising).

The core of an alphabetic dictionary of this kind is the lexical information contained in the microstructure, and the empirical, technical and human resources for acquiring this lexical information form the largest single kind of resource in a lexicography workbench (see Figure 2).

The entries in Figure 2 are from a traditional alphabetically organised semasiological dictionary. The entries have a microstructure which can be formally modelled as a vector or row in a matrix, with the following elements: headword (simultaneously representing orthographical lexical information), pronunciation (in a variety of IPA transcription), an abbreviation (*adj*, *n*) for the *part of speech* (POS, syntactic category), a definition with a modifying relation in the case of the adjective, an identification of the domain (‘tech’, i.e. technical), and a classical dictionary definition in the case of the nouns. Additionally, structurally (morphologically) related words are given, such as the adverb ‘lexically’, the agentive noun ‘lexicographer’, and, in the case lf ‘lexis’ there is a mesostructural cross-reference, ‘compare VOCABULARY’.

lex·i·cal /ˈleksɪkəl/ *adj tech* of or about words —~**ly** /kli/ *adv*

lex·i·cog·ra·phy /ˌleksɪˈkɒgrəfi‖-ˈkɑː-/ *n* [U] the writing and making of dictionaries —-**pher** *n*

lex·i·col·o·gy /ˌleksɪˈkɒlədʒi‖-ˈkɑː-/ *n* [U] *tech* the study of the meaning and uses of words

lex·i·con /ˈleksɪkən‖-kɑːn, -kən/ *n* **1 a** a dictionary **b** a list of words with their meanings **2** *tech* all the words and phrases used in a particular language

lex·is /ˈleksɨs/ *n* [U] *tech* all the words that belong to a particular subject or language, or that a particular person knows —compare VOCABULARY

Figure 2. Examples of lexicon articles from a traditional dictionary (the Langenscheidt-Longman Dictionary of Contemporary English, 1987).

The classical dictionary definition is also known technically as a *definition by nearest kind and specific differences* (also in Latin: *definitio per genus proximum et differentia specifica*). In the entry 'lexicon', for example, one definition is simply a synonym, but the next has the nearest kind 'list' and the specific differences 'of words' with a further specific difference 'with their meanings'. The technical term for the headword (left-hand side) in a classical dictionary definition is the *definiendum* (Latin for 'to be defined'), and the right-hand side, with the nearest kind and the specific differences, is the *definiens* (Latin for 'defining').

Traditionally, the information for the data categories in the microstructure is acquired from three main resource types: the lexicographer's knowledge of the language; extensive collections of texts in the language; other dictionaries. The ultimate criterion for a practical dictionary will be the lexicographer's knowledge of the language. However, traditional dictionaries are beset with preferences and idiosyncrasies (definitions, spelling variants, alternative plurals) introduced by the lexicographer. Other dictionaries may be a useful source of information, but if used their idiosyncracies will be perpetuated. Basing a dictionary on extensive corpus resources has the advantage of comprehensiveness and facilitates the development of consensual lexical information.

2.2.2. Lexicon creation

The main type of resource for modern dictionaries is the corpus or text collection, which is processed by modern computational and manual lexicographic methods. For a large general purpose alphabetic dictionary, the corpus will contain a selection of word tokens of the order of tens of millions of word tokens (or more), which may well reduce to a set of word types of the order of a hundred thousand (depending on corpus size and the definition of 'word'), yielding a type-token ratio greater than 1:100. For a spoken language system, the lexicon will in general be much smaller, based on a scenario-specific vocabulary.

Resource processing consists of the following main steps, which requiring appropriate software tools (note that the procedures listed after tokenisation are not necessarily conducted in the order given):

1. Tokenisation. Individual word tokens are identified, including abbreviations, numbers, prices, dates, punctuation, identification of complex layout objects such as tables.
2. POS (*part of speech*) tagging. Each token is provided with a label (or set of labels) constituting a hypothesis about its part of speech; the European EAGLES (*Expert Advisory Groups for Language Engineering Systems*[5]) developed a standard POS tagset for European languages, which has been extended and applied to other languages (these sets are in flux; consult the internet for up-to-date details).
3. Word token and word type list creation. A list of (possibly inflected) word types is extracted from the set of tokens, often also in conjunction with the word token frequencies.
4. Lemmatisation. A list of lemmas is created from the list of word types, involving stemming in the simplest case, and morphological analysis in the general case (cf. Jurafsky & Martin 2000 Carstensen & al. 2010).
5. Concordancing. A context dictionary consisting of a list of items (types, lemmas, tags, etc.) and the contexts in which they occur in the texts. The best known kind of concordance is the KWIC (*KeyWord In Context*), a simple list of words and their left and right context strings.
6. Word sketching (Atkins & Rundell 2008). Extraction of a maximum of (grammatical and other kinds of similarity) information about lemmas based on their distribution in the texts.
7. Dictionary database compilation. Semi-automatic (moderated) entry of information into the lexical database.
8. Manual editing of lexicon articles (definitions, etc.).
9. Production. Selection, organisation and formatting of lexical information for the intended dictionary megastructure.

These procedures apply, with suitable modifications, to the compilation of other types of dictionary, including dictionaries for use in multilingual, speech-based and multimodal communication systems.

The following examples of KWIC concordances illustrate one of the important types of lexicographic resource (characters simplified).

The first example is taken from an interactive concordance on the internet, as a lexicographic resource for the Verbmobil speech-to-speech translation project in the early 1990s, for the phrase 'jede Woche' ('every week'):

```
6 × jede Woche
cd1_m004n_FRB003: rzigsten Woche jeweils einmal jede Woche .
cd2_m018n_ANP014: ja , dann k"onnen wir ja in jede Woche zwei
   Termine legen und dann h
cd2_m021n_BEP005: ht besser , das "ahm entweder jede Woche
   zu machen , also ein eine Woc
cd3_m025n_TIS002: wochenweise , ne , weil jeden jede Woche
   einen Termin , und dann m"uss
cd3_m025n_RAL009: also , ich glaube , jede Woche am gleichen
   Tag , das kriegen
cd3_m027n_MPU015: ag festlegen , sondern uns hm jede Woche
   einen anderen Tag aussuchen .
```

The second example is an extract from an automatically created printed concordance entry for texts from the Nigerian language Ibibio (characters simplified; the word ‘abasi’ means, approximately, ‘lord’, ‘ruler’):

abasi:
- ukpe ikpe ke esop idan ye ukwooro iko ke ufok **abasi**
- mme okwooro iko **abasi**
- ukpono **abasi** eyeyin
- **abasi** ukot
- **abasi** imaan
- **abasi** ison ye akwa abasi ibom

2.2.3. Lexicon resource acquisition

Lexicon resource acquisition is a complex procedure, which may, however, be reduced to a sequence based on levels of abstraction. A useful hierarchy of such stages in lexicon acquisition, some of which apply to the acquisition of other linguistic resources such as grammars, is shown in Figure 3.

From the point of view of storage in a lexicographic system, all the objects represented by the inside boxes in Figure 3 are data of different kinds, for which different data structures are required. However, from a linguistic point of view it is convenient to distinguish between *corpus data* and *lexicon data*, as in Figure 3.

The primary corpus of *raw data* consists of the formatted text material to be analysed (and in the case of speech, recordings). The raw data may include bilingual information from parallel or comparable corpora if a bilingual or translation dictionary is being compiled.

The secondary corpus of *processed data* consists of the character streams aligned with the raw data, which are segmented into tokens of the required linguistic units (such as characters and character sequences, affixes, words) in a tokenisation step. In a speech corpus the units may be phonemes and syllables, or prosodic units such as accents and tones; in a video corpus the units may be

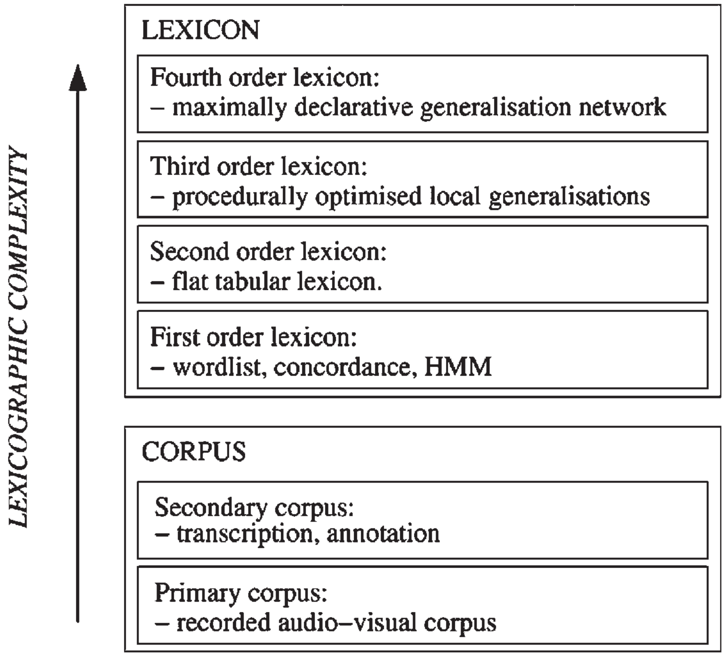


Figure 3. Levels of lexicon data types.

gesture configurations, visemes (facial movements, particularly of the lips, associated with phonemes).

The lexicon data and structures are of many kinds depending on the required lexicon use cases. However, the types shown in the figure have some generic validity for all types.

The *first order lexicon* is the concordance as outlined above, for example a list of word-context pairs. Concordancing is a standard procedure in all lexicon construction.

The *second order lexicon* is rarely formulated explicitly, but represents an intermediate stage between concordances and standard lexicon databases: wordform tokens from the corpus are reduced to wordform types and listed separately if, and only if, they have no distinguishing lexical properties of form or meaning. If they have even one distinguishing lexical property, they are listed separately. In this respect the second order lexicon represents an intermediate stage between the concordance, with no abstraction over sets of entries, and the third order lexicon, with use case specific abstractions.

The *third order lexicon*, the most common type, is lemma based, and brings polysemous and perhaps homonymic items together, provided that they have the same lemma. It is at this stage that distinctions between different use cases become apparent, since the organisation at this level is based on procedural convenience, for instance for semasiological lexicons (with wordform lemmas and semantic lexical information) versus onomasiological lexicons such as thesauri (plural of thesaurus), with concept-based lemmas and form based lexical information. The semasiological-onomasiological distinction is rather simplistic: it is not easy to assign a WordNet, a translation dictionary or a pronunciation dictionary, for example, to the one or the other, though both are clearly third order lexicons.

The *fourth order lexicon* generalises over use cases and models lexical information according to declarative (logical) rather than procedural (use case based) criteria. The fourth order lexicon is mainly of theoretical interest and is less well known than the others, and distinguishes a maximum of linguistic generalisations about pronunciation, grammar and meaning from a minimum of exceptions, based on information extracted from the lexical information available at the lower order levels. The fourth order lexicon uses formalisms such as inheritance graphs (representing implication hierarchies, taxonomies), but has nevertheless been used in practical communication systems (Gibbon & Lüngen 2000). In principle, the fourth order lexicon provides an ideally compact form of storage for lexical information which is intended for re-using in widely different use cases; the generally hierarchical (or even more complex) form is well-suited to the contemporary XML-based data structures used in resource storage.

3. Resources for speech and multimodal systems

3.1. Informal overview

From the perspective of hardware resources, the fields of speech systems and multimodal systems are highly complex, and involve many more components which are dependent on specialised hardware than text based systems. For this reason they cannot be covered comprehensively or in detail in a general handbook article. Fortunately, there exist a number of relatively comprehensive and widely used handbooks on speech resources and, to some extent, on resources for multimodal systems, which should be consulted (see the final section). The present article is restricted to generic considerations and to the specific example of speech synthesis systems.

A general rule is that speech resources are orders of magnitude larger, more complex (and more expensive) to make than text-based resources. A further empirical rule is that multimodal resources are orders of magnitude larger, more complex (and more expensive) than speech resources. Speech systems are often embedded in multimodal systems. Speech and visual communication resources also share properties and techniques at a generic level: both are concerned with the processing of physical symbols and the mapping of segments of these on to symbolic strings and other patterns. This can also apply to text, but only when text is scanned from analogue sources such as handwriting on paper and represents a kind of visual information. *Optical character recognition* (OCR) algorithms used for analysing scanned text share properties with those needed for speech and vision decoding.

The more abstract and language-related levels of speech and multimodal resources are practically identical to those involved in text processing. Therefore it is appropriate to concentrate more on the form-oriented aspects of pronunciation. It is not possible within the confines of the present context to account for the visual aspect of multimodal resources, including gestural communication (signing by deaf communicators, and conversational gesture). In a previous handbook in the field (Gibbon et al. 1997) a distinction was made between three phases of data acquisition for corpus building and processing, each of which requires rather different operational resources in the form of human procedures or software and hardware tools: the *pre-recording*, *post-recording*, and *post-recording* phases. For further details consult also Gibbon et al. (2000).

1. *Phase I: pre-recording phase*. The pre-recording phase is concerned with use case specific scenario and experiment design based on requirements specifications for later processing in the post-recording phase. These requirements and design issues determine the materials (equipment, prompts, texts, scenario layout, participants etc.) and procedures to be used in the recording phase. Many contextual details need to be taken into consideration: for instance, in a noisy application environment the 'Lombard effect' (change of voice characteristics) is to be found, therefore recording and testing under studio conditions may be inappropriate. This preparatory phase is arguably the most complex phase, and the specialised literature should be consulted: if resource design is not right, the implementation will not be right: 'garbage in, garbage out'.
2. *Phase 2: recording phase*. During the recording phase, scenario or experiment-specific recordings are made as raw data (see Figure 3). For specialised purposes, software (or otherwise) controlled randomised or structured prompts (e.g. for systematic testing and experimentation purposes), specialised environments (e.g. sound-proofed rooms; noisy car or airplane settings; telephone; 'Wizard of Oz' or 'man-behind-the-curtain' simulated human-computer communication) may be needed. For more general purposes, less formal environments with across-the-table dialogue involving face-to-face or hidden communicators may be suitable. Here, too, the specialised handbooks should be consulted.

3. *Phase 3: post-recording phase*. The post-recording phase is essentially the implementation phase for the specifications and design developed in the pre-recording phase. The following generic procedure applies to most speech and multimodal data.
   a. In processing speech and multimodal primary data transcriptions (assignment of strings of labels to recordings, without time-stamps) and annotations (assignment of labels to segments of recordings, with time-stamps). Parallel annotations may be assigned to the same data (the layers of parallel annotation streams are known as 'tiers' – rhyming with 'fears', not with 'flyers'). The annotation procedure may be manual (required for initial bootstrapping), semi-automatic (e.g. automatic but post-edited) or automatic (using statistical annotation software with training component). Information for archiving and further processing is extracted from the annotations.
   b. The next steps are generally the extraction of a list of word tokens from the annotations and the creation of a machine-readable pronunciation dictionary using standardised orthographic and phonetic (more usually: phonemic) coding conventions. Although Unicode is generally used for text-based systems, it is very much oriented towards output for printing, rather than convenient input or processing. In the speech technology context, standardised custom alphabets are generally used, the most common of which, in multilingual resources, is still the SAMPA (sometimes 'SAM-PA'), i.e. SAM Phonetic Alphabet, developed in the European Commission funded SAM ('Speech Assessment Methodology') project in the 1980s and extended to cover all languages in the early 1990s.
   c. Several efficient free speech annotation software tools are available, such as Praat[6] (probably the best known), Transcriber[7], WaveSurfer[8] (in view of the rapid development in the field, the internet should be consulted for further information). There is much less agreement about 'alphabets' for annotating video signals, though there are a number of software packages for video annotation, the most widely used being Anvil[9], Elan[10] and EXMARaLDA[11] (see Dipper et al. 2004; Rohlfing et al. 2006).
   d. Further post-recording analysis, e.g. creation of lexicons, word models, grammars etc., is closely related to the analogous levels in text-based system resource development. Finally, evaluation of resource type and quality is an essential part of current best practice in resource creation and deployment.

Beyond these generic aspects of the pre-recording, recording and post-recording phases are very many technical details: specific algorithms for speech stream transformation and visualisation (including waveform, spectrum, pitch track), for speech stream segmentation and in multimodal contexts also for video scene line and object detection (see additional references in the final section of this article).

3.2. Example: resources for speech synthesis systems

3.2.1. Resource types

Speech synthesis systems are generally embedded components of systems with more complex functionality. Their specifications therefore depend on the use cases for their technical environment, and on the technical environment itself. Typical uses for embedded speech synthesis systems are public address systems (e.g. railway stations, airports), geographical information systems (e.g. vehicle satellite navigation systems), information systems for non-literate users, dictation software (for readback), screen readers for the blind, speech-to-speech translation software, multimodal systems including robotic systems. In categorising the types of embedding into more complex systems, a two-way distinction is conventionally made between *Text-To-Speech* (TTS) synthesis, where the input is text and the output is speech, and *Concept-To-Speech* (CTS) synthesis, where the input is a conceptual representation (commonly a database) (see also Martin and Schultz 2012 in this volume). The prevalent type is TTS; in practice, CTS systems may also involve text as an interface.

A TTS system requires resources for developing the following subcomponents:

1. Text parser: the text is pre-processed in order to extract implicit information:
   a. The spelling and ultimately the pronunciation of special text components such as abbreviations and numbers must be extracted.
   b. A pronunciation lexicon, usually with additional pronunciation rules, is required.
   c. A parser is needed for disambiguating the structure by picking the correct word readings from the lexicon and delimiting the phrasing of sentences.
   d. A grapheme-to-phoneme (phonetisation) component is used to derive a transcription of the speech sounds for input to the speech processing component.
   e. A prosody module is needed for deriving intonation and accentuation patterns for input to the speech processing component.
2. Signal processing component: conversion from an interface with parsed and phonetised text with added prosodic information into a synthetic speech signal.

The text parser is a special case of the kind of parser which is used in text processing in general, enhanced with phonetisation and prosodic modelling information, and will not be discussed further here.

For the signal processing component there are several different speech synthesis paradigms, including the following main types, for which paradigm specific resources are required:

1. Pre-recorded ‘canned’ speech. Canned speech is typically used in straightforward information service environments such as satellite navigation systems for vehicles, and for railway station announcements. Systems such as these use a restricted set of utterance templates which permit substitution of station names and times, but also permit a combinatorially large set of new utterances to be synthesised. Canned speech is in principle very comprehensible and very natural, provided that the template units are carefully designed and produced, with close attention paid to the correct prosody (intonation and accentuation), and to appropriate transitions between canned speech units.
2. Unit concatenation speech synthesis. Small units, such as phonemes, diphones, demi-syllables and sometimes larger units, are concatenated to form words and sentences. There are three main approaches, each of which requires different kinds of resource:
   a. Diphone synthesis is one of the first kinds of concatenative speech synthesis, and is still used. In diphone synthesis, pre-recorded speech samples containing all the diphones in the sound system of the language are used, which are concatenated in order to reproduce the patterns of the input syllable and word sequences. A diphone is essentially a pair of phonemes (speech sounds; see below).
   b. Unit selection synthesis, a popular variety of speech synthesis, and in general more natural than diphone synthesis, is based on selecting continuous units from a large recorded corpus. The corpus is designed to contain all the phonemes, generally all the diphones, and perhaps all the triphones (sequences of three phonemes). Units are concatenated after calculating the best possible fit (cost, weight).
   c. *Hidden Markov Model* (HMM) synthesis, a recent development based on stochastic modelling of unit sequences, trained on a suitable corpus.
3. Formant speech synthesis. Formant synthesis is one of the earliest kinds of speech synthesis, and is based on the spectral structure of speech sounds. An acoustic signal is reconstructed from empirical information about vowels, consonants, and the pitch, intensity and duration patterning of the intended synthetic speech signal. In principle, this approach is the most flexible and parametrisable in terms of linguistic and phonetic properties, but is more difficult to use in practical systems than concatenative techniques.

3.2.2. Resource creation

In order to illustrate the resource creation process for a speech system, a traditional diphone concatenation approach is described. For unit concatenation and HMM synthesis, which are more complex and currently under rapid development, the technical literature and internet sources should be consulted (see also the last section for references). However, many of the resources needed for diphone synthesis are also required in similar form for the more complex speech synthesis techniques.

The following resources will be required for a diphone synthesiser:

1. Text processing component:
   a. Text parser as outlined above, which will tokenise words, decode abbreviations, and establish phrasing, and focus points and intervals.
   b. Phonetiser or grapheme-to-phoneme converter, to produce phonemic/phonetic representations of words.
   c. Prosodic model, to utilise the phrasing and focus information to associate pitch, duration and intensity patterning to the word sequence.
2. Signal processing component:
   a. Pre-recording phase:
      i. List of phonemes (and perhaps also major allophones) of the language concerned. The size of the phoneme set varies considerably, depending on the typological properties of the language, from around 15 to several dozen.
      ii. List of diphones based on the list of phonemes. For a phoneme set P the size of the set of diphones is therefore maximally $|P|^2$, the square of the number of phonemes. This set includes both diphones which occur within words and diphones which occur across word boundaries, as well as a pause unit. Since the number of phonemes in the language determines the size of the diphone set, evidently languages vary greatly in the sizes of the diphone sets.
      iii. Prompts containing examples of each diphone. A traditional method of compiling a suitable set of prompts is to find a set of words or longer expressions containing the desired units, and to put these into a standard frame such as “Say X again.” However, a more efficient method of compiling prompts is to create a ‘phonetically rich’ corpus, i.e. to start with a large text corpus and extract the minimum number of sentences which together contain all the diphones in the entire text; this can be automatised, for example by means of a common scripting language such as Unix/Linux shell, Perl, Python or Ruby. If any diphones are missing, further sentences need to be added.

b. Recording phase:
   i. For commercial systems, a professional speaker is usually recorded, in a professional studio. In prototype development, less stringent standards are imposed. Nevertheless, certain technology-dependent conditions need to be met: good voice quality (no ‘creaky’ or ‘breathy’ voice), good control of pitch and volume, pronunciation and intonation patterns which are appropriate for the task concerned. For special applications such as video games, in which emotional and aggressive speech varieties are often encountered, special but still highly controlled conditions apply.
   ii. The recording equipment and environment need to be carefully controlled in order to avoid unwanted noise, echo, delayed acoustic feedback to the speaker (cf. the approximately 20 ms syllable-length delay which leads to the ‘Lee Effect’ of involuntary speech inhibition).
c. Post-recording phase:
   i. Recordings are archived with well-defined file-name conventions, in a suitable data management environment, associated with standardised metadata such as the data proposed by the *Open Language Archive Community* (OLAC[12]; consult the internet for further details).
   ii. Recordings which do not correspond to the requirements are rejected (and re-recorded if diphones are then missing). Audio normalisation (e.g. of intensity) is performed where required, in order to achieve uniform recording properties.
   iii. Annotation of recordings with transcriptions and time-stamps, is carried out using specialised software. The basic kind of annotation, which provides a benchmark for automatic annotation, is manual annotation by phonetically trained skilled annotators. An essential part of the annotation process is the evaluation of the annotations, both by objective means (e.g. monitoring the use of correct annotation labels) and by experimental intersubjective means (e.g. comparing the annotations of independent annotators, which are never 100 % identical but should approach 90 % similarity in use of labels and in time-stamp accuracy as far as possible). An example of annotation will be discussed below. Semi-automatic annotation either uses interactive software for labelling, monitored by a skilled human annotator, or post-editing of automatically generated annotations, Fully automatic annotation, which can be necessary if very large recorded corpora are used, may start with the orthographic prompt text, phonetise the text automatically with a grapheme-to-phoneme converter, and assign time-stamps relating to the recorded signal using statistically trained forced alignment software; post-annotation editing is often

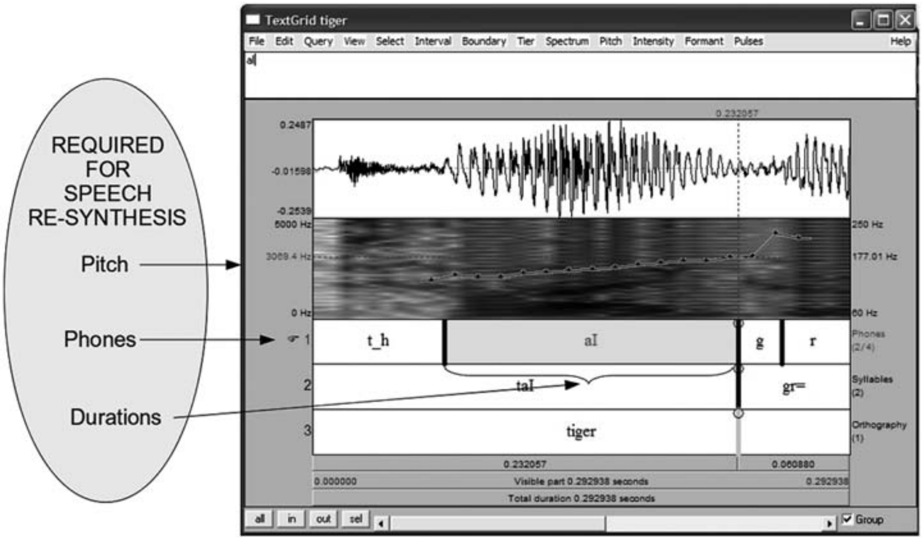


Figure 4. Annotation of speech signal for speech synthesis with the Praat software workbench.

necessary, however, unless tests of samples have shown that inaccuracies are not relevant in practice.

iv. Diphone database creation. Diphones are extracted from the annotated recordings and processed in order to create a diphone database or 'diphone voice' for use in the runtime speech synthesis system. Processing can involve pitch extraction and normalisation, and the normalisation of the intensity (volume) and duration of the diphones in order to facilitate the runtime synthesis procedure. Extensive evaluation of the diphone database during the development process by a skilled phonetically trained evaluator is required in order to establish that the diphones it contains can be used without distortion in many different word and sentence contexts.

v. System evaluation. The evaluation of the diphone voice in the runtime system uses extensive and varied perception tests with subjects who represent potential users, based on criteria of comprehensibility and naturalness.

One of the core resources of any speech technology system is the annotated speech recording. In Figure 4 the main features of a typical annotation using the Praat software workbench (Boersma 2001) is shown.

Figure 4 shows the speech signal in three visualisations: as waveform (top), as spectrogramme (second from top), and as pitch track (pitch trace, fundamental frequency trace), superimposed on the spectrogramme. Below the signal visualisations are three tiers for three annotation streams: phones (phoneme tokens in context), syllables and words. For a diphone synthesiser, the phone tier is the crucial source of information.

For further processing, the stored annotation file is used. The annotation visualised in Figure 4 is stored internally in the Praat system as follows (the ‘Phones’ tier only):

```
File type = "ooTextFile"
Object class = "TextGrid"

xmin = 0
xmax = 0.2929375
tiers? <exists>
size = 1
item []:
  item [1]:
    class = "IntervalTier"
    name = "Phones"
    xmin = 0
    xmax = 0.2929375
    intervals: size = 4
intervals [1]:
      xmin = 0
      xmax = 0.0718
      text = "t_h"
    intervals [2]:
      xmin = 0.0718
      xmax = 0.2323
      text = "aI"
    intervals [3]:
      xmin = 0.2323
      xmax = 0.25579
      text = "g"
intervals [4]:
      xmin = 0.2557
      xmax = 0.2929
      text = "r `="
```

The characteristics of the annotation can be seen clearly: each annotation label is represented by a numbered segment (‘interval’) with three properties, two time stamps and a text label: xmin = 0.2323 (the start of the interval), xmax = 0.25579 (the end of the interval) and text = “g” (the transcription label).

Diphones are constructed from pairs of phones, with the beginning of the diphone starting in the middle, or the most ‘stable’ part of the first phone and continuing to the middle, or the most ‘stable’ part of the second phone. The definition of ‘stable’ varies from phoneme to phoneme, and is generally most easy to identify in vowels and other continuous sounds, and least easy to identify in obstruent consonants.

The more modern techniques of unit selection synthesis and HMM synthesis, as well as others which have not been mentioned, require somewhat different procedures. For example, whereas in standard diphone synthesis the selection of suitable diphones from the recorded corpus is done prior to runtime and stored in a database, in unit selection synthesis the recorded corpus is, essentially, the database, though pre-runtime optimisations and calculation of properties of units in the database are performed. This has consequences for other components of the speech synthesiser and resources for these. In a diphone speech synthesiser such as the well-known MBROLA diphone voice handling system (Dutoit 1997) the pitch and duration values will in general be calculated on the basis of rules of grammar and a prosody description, and included in a well-defined interface between the text parsing and the actual voice handling part. For a unit selection system, the corpus itself in general needs to be annotated with prosodic features, either in a rule-based fashion or derived statistically from the corpus itself.

The development procedure, and the creation of appropriate resources and enabling technologies to facilitate development even of a relatively straightforward traditional diphone based system is evidently highly complex, and the specialist handbooks and their bibliographies should be consulted for further information (see the final section).

4. Recommendations for further reference

In the text reference was made to the need to consult specialist literature for further information, background knowledge and development recipes. For up-to-date information, judicious consultation of the internet is advised, particularly of contributions to major conferences in the fields concerned. However, the following short list of publications will serve as a starting point.

1. Text resources: Abeillé (2003), van Halteren (1999) (corpus processing); Fellbaum (1998), Atkins & Rundell (2008), van Eynde & Gibbon (2000) (Lexicography); Beesley & Karttunen (2003) (finite state modelling);
2. Speech resources: Coleman (2005) (overview); Dutoit (1997) (speech synthesis); Gibbon et al. (1997), Gibbon et al. (2000) (spoken language resources and standards); Wahlster (2000), Wahlster (2006) (speech-to-speech translation; speech in mobile devices).
3. General: Carstensen et al. 2010 (computational linguistics and speech technology, in German); Goldfarb & Prescod (XML technologies); Jurafsky & Martin (2000) (speech and language technologies); Hopcroft et al. (2006) (formal languages); Lobin (2010) (text technology).

Since the field of resources for technical communication systems is developing and expanding rapidly, it is advisable also to consult the proceedings of the most important conferences in the field. The central conference series for resources in both speech and text processing areas is the *Language Resources and Evaluation Conference* (LREC[13]) series, whose events take place every two years; for speech and multimodal communication alone, the Interspeech conference series is the major event.

Notes

1. http://www.iso.org/iso/home.html
2. http://www.langsci.ucl.ac.uk/ipa/
3. http://www.w3.org/
4. http://www.w3.org/Style/CSS/
5. http://www.ilc.cnr.it/EAGLES/home.html
6. http://www.fon.hum.uva.nl/praat/
7. http://xml.coverpages.org/transcriber.html
8. http://www.speech.kth.se/wavesurfer/
9. http://www.anvil-software.de/
10. http://www.lat-mpi.eu/tools/elan/
11. http://www.exmaralda.org/
12. http://www.language-archives.org/
13. http://www.lrec-conf.org/

5. References

Abeillé, A. (ed.)
2003 *Treebanks: Building and Using Parsed Corpora*. Dordrecht: Kluwer Academic Publishers.

Allwood, Jens and Elisabeth Ahlsén
2012 Multimodal communication. In: Alexander Mehler, Laurent Romary and Dafydd Gibbon (eds.), *Handbook of Technical Communication*. Berlin/ Boston: De Gruyter.

Andry, François, Norman M. Fraser, Scott, Simon Thornton, Nick J. Youd
1992 Making DATR work for speech: lexicon compilation in SUNDIAL. *Computational Linguistics* 18:3, 245–267.

Atkins, B. T. S., Rundell, Michael
2008 *The Oxford Guide to Practical Lexicography.* Oxford: OUP.

Beesley, Kenneth R., Lauri Karttunen
2003 *Finite State Morphology.* Stanford: CSLI Publications.

Boersma, Paul
2001 Praat, a system for doing phonetics by computer. *Glot International* 5:9/10, 341–345.

Carstensen, Kai-Uwe, Christian Ebert, Cornelia Ebert, Susanne J. Jekat, Ralf Klabunde, Hagen Langer
2010 *Computerlinguistik und Sprachtechnologie*. Heidelberg: Spektrum Akademischer Verlag.
Coleman, John
2005 *Introducing Speech and Language Processing*. Cambridge: Cambridge University Press.
Dipper, Stefanie, Michael Götze and Manfred Stede
2004 Simple Annotation Tools for Complex Annotation Tasks: an Evaluation. In: Proceedings of the LREC Workshop on XML-based Richly Annotated Corpora, 54–62.
Dutoit, Thierry
1997 *An Introduction to Text-To-Speech Synthesis*. Dordrecht: Kluwer Academic Publishers.
Eynde, Frank van, Dafydd Gibbon
2000 *Lexicon Development for speech and language processing*. Dordrecht: Kluwer Academic Publishers.
Fellbaum, Christiane (ed.)
1998 *WordNet. An Electronic Lexical Database*. Cambridge, MA: MIT Press.
Gibbon, Dafydd, Roger Moore, Richard Winski
1997 *Handbook of Standards and Resources for Spoken Language Systems*. Berlin: Mouton de Gruyter.
Gibbon, Dafydd, Harald Lüngen
2000 Speech Lexica and Consistent Multilingual Vocabularies. In: Wahlster (2000), pp. 296–307.
Gibbon, Dafydd, Inge Mertins, Roger Moore
2000 *Handbook of Multimodal and Spoken Dialogue Systems: Resources, Terminology and Product Evaluation*. New York: Kluwer Academic Publishers.
Goldfarb, Charles F., Paul Prescod
2003 *XML Handbook*. Upper Saddle River, NJ: Prentice Hall.
Hopcroft, John E., Rajeev Motwani, Jeffrey D. Ullman
2006 *Introduction to Automata Theory, Languages, and Computation* (3rd ed.). Addison-Wesley.
Jurafsky, Daniel & James H. Martin
2000 *Speech and Language Processing. An Introduction to Natural Language Processing and Speech Recognition*. Upper Saddle River, NJ: Prentice Hall.
Halteren, Hans van (ed.)
1999 *Syntactic Wordclass Tagging*. Dordrecht: Kluwer Academic Publishers.
Lobin, Henning
2010 *Computerlinguistik und Texttechnologie*. Paderborn: Fink.
Lücking, Andy and Thies Pfeiffer
2012 Framing multi-modal technical communication. In: Alexander Mehler, Laurent Romary and Dafydd Gibbon (eds.), *Handbook of Technical Communication*. Berlin/Boston: De Gruyter.
Martin, Jean-Claude and Tanja Schultz
2012 Multimodal and speech technology. In: Alexander Mehler, Laurent Romary and Dafydd Gibbon (eds.), *Handbook of Technical Communication*. Berlin/Boston: De Gruyter.

Paaß, Gerhard
2012 Document classification, information retrieval, text and web mining. In: Alexander Mehler, Laurent Romary and Dafydd Gibbon (eds.), *Handbook of Technical Communication*. Berlin/Boston: De Gruyter.
Rahtz, Sebastian
2012 Representation of documents in technical communication. In: Alexander Mehler, Laurent Romary and Dafydd Gibbon (eds.), *Handbook of Technical Communication*. Berlin/Boston: De Gruyter.
Rohlfing, Katharina, Daniel Loehr, Susan Duncan, Amanda Brown, Amy Franklin, Irene Kimbara, Jan-Torsten Milde, Fey Parrill, Travis Rose, Thomas Schmidt, Han Sloetjes, Alexandra Thies and Sandra Wellinghoff
2006 Comparison of multimodal annotation tools. *Gesprächsforschung* 7: 99–123.
Stührenberg, Maik
2012 Foundations of markup languages. In: Alexander Mehler, Laurent Romary and Dafydd Gibbon (eds.), *Handbook of Technical Communication*. Berlin/Boston: De Gruyter.
Trippel, Thorsten
2012 Controlled language structures in technical communication. In: Alexander Mehler, Laurent Romary and Dafydd Gibbon (eds.), *Handbook of Technical Communication*. Berlin/Boston: De Gruyter.
Wahlster, Wolfgang (ed.)
2000 *Verbmobil: Foundations of Speech-to-Speech Translation*. Berlin, Heidelberg, New York: Springer.
Wahlster, Wolfgang
2006 *Smartkom: Foundations of Multimodal Dialogue Systems*. Berlin: Springer.

8. Evaluation of Technical Communication

Peter Menke

1. Introduction

This chapter gives an overview of evaluation in the context of technical communication. This is not a trivial task, since throughout the fields and areas of research, the concept of *evaluation* is not defined and interpreted uniformly. In addition, many disciplines restrict the scope of their definitions of the term to their own area of research, hence, to a smaller subset of evaluation objects, methods or contexts. What is common to all these variations of evaluation? Although all definitions and models implicitly do share a common core, it is not easy to find a definitory intersection that matches concepts of evaluation in all areas.

As a consequence, this chapter starts with an attempt at a definition of the concept of evaluation in the following section. It examines evaluation from a scientific point of view: namely as a central part in the typical research cycle that utilizes methods to create models of reality and interprets them, thus assessing a previously defined theory.

In the next section, measures and criteria for the assessment of evaluation quality are described. Subsequently, three sections are dedicated to certain specialized views onto the subject. Each of them deals with one specific aspect or component of evaluation, giving a list of examples from the field of technical communication.

1. The first of these sections thematizes possible *objects of evaluation* in the study area. It answers the questions: *What* is to be evaluated, and what are suitable *indicators* to be used in evaluation? This section iterates through the main topics of this handbook and will summarize possible objects of evaluation for each of them, along with examples of typical corresponding indicators.
2. The second section deals with *evaluation methods*, with the central question: *How* and *by which means* can the pieces of information collected in the previous section be processed or analyzed in order to produce evaluation results?
3. In the third section, several components of evaluation processes will be shown in interaction on the basis of selected real scenarios.

Finally, a summary on the subject concludes this chapter.

2. What is evaluation?

Several implicit and explicit definitions of evaluation have been proposed, many of which (1) are restricting and suiting the term to the use in a specific field or discipline only (e.g., Hirschman and Mani 2003 in the area of computational linguistics, or Widmer 1996 and Widmer et al. 2009 in the area of social science), or (2) give a definition that is too loose and general. For instance, Michael Scriven's profound and comprehensive "Evaluation Thesaurus" begins with the following definition:

> Evaluation is the process of determining the merit, worth and value of things, and evaluations are the products of that process. *Scriven (1991), p. 1*

This, however, also includes the assessment of things that have not been created or designed on purpose. It is difficult to decide whether such processes should also be considered evaluations without weakening the notion of evaluation to that of a mere assessment or examination.[1] A definition of a common concept of evaluation that is more suitable for the application within the scope of this volume could read as follows:

> Evaluation is the process of investigating whether (or to what extent) an object (abstract or concrete) meets certain expectations, requirements, or specifications. Evaluation can also refer to the results of this process.

In fact, virtually any human artifact can be evaluated, because it has been designed with a purpose, according to a plan, a draft or an idea, and its creators had specific expectations when they created it. Therefore, evaluation means to collect and process relevant attributes and properties of the result and to compare them to plans, expectations and demands that were present when creating the artifact.

However, due to man's creativity, the set of human artifacts is huge and heterogeneous, and its members and subsets diverge fundamentally. Hence, there is not one generic set of means of evaluation that can handle all kinds of entities in the world. Instead, if one intends to evaluate a certain entity, one has to create a custom set of mechanisms and procedures of evaluation that takes into account the nature of the given entity, its design and its creation process. All of these instances, however, follow a scheme depicted in Figure 1 which will be explained in the following paragraphs.

Mechanisms and methods usually do *not* work on objects under observation directly. Instead, they abstract and simplify them according to the underlying theory (see arrow **a** in Figure 1) in order to make them more manageable. Or, in other words, they create a *model* of the objects (Stachowiak 1965; 1989) by means of a mapping mechanism between sets of real objects and model objects (see arrow **b** in Figure 1). The more properties and inter-object-relations such a mapping mechanism preserves, the more reliable it is, because it guarantees that

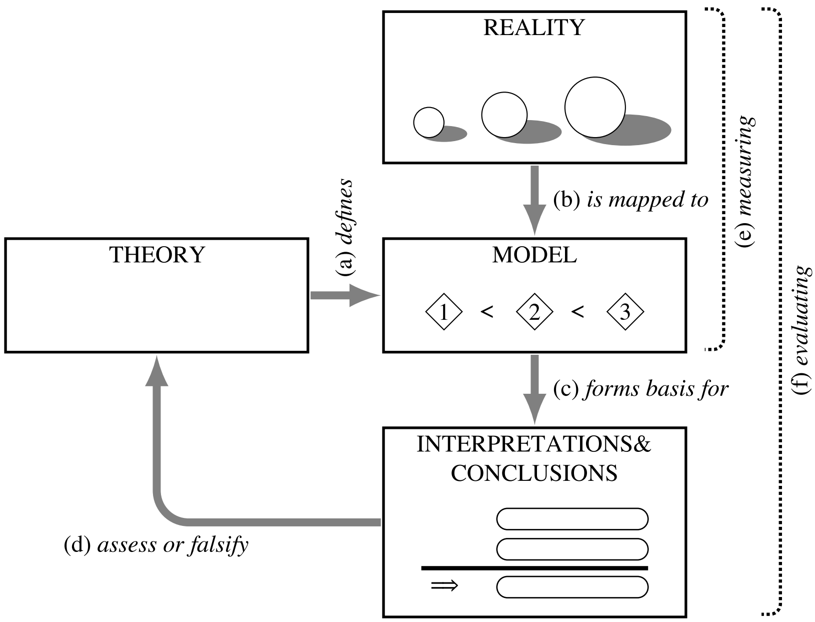


Figure 1. Evaluation as part of a cycle that consists of the proposal of a theory, taking measurements and drawing conclusions on their basis, leading to a falsification of the theory.

analyses on the model may be used for conclusions and attributions of the antecedent objects in reality. Such mappings that preserve all relevant structures when mapping one set or system of objects to another are called *homomorphisms* (Altmann and Grotjahn 1988).[2]

Models do not necessarily have to bear that name explicitly. In several situations, evaluation can also be performed with other structures that *behave* and *appear* like models although they are named differently. For example, an HTML document containing a text might be referred to with other terms than "model". Nevertheless, it is a simplified representation of another object set, and thus, by definition, it is a model of that object set. To avoid unnecessarily bloated sentences, I will refer to all these representations as "models" from now on, regardless whether they explicitly bear that name or not.

The methods and mechanisms then work on the model by performing measurements and subsequent interpreting their results in order to draw conclusions. Arrow **c** in Figure 1 visualizes this process, which is not specific to evaluation but can be part of many other tasks as well. Scientific research is one of the most prominent examples that involves it, starting with a theory that determines and selects both methods and objects from reality.

After drawing the conclusions, they are utilized to assess (and, optionally, falsify) the theory (arrow **d** in Figure 1). This process can (and usually is) iterated in a cycle a→b→c→d→a, where ideally every step modifies, refines and rounds off the theory.

According to this model, actual *measuring* (area **e** in Figure 1) can be interpreted as a subset of an *evaluation*, which also performs concluding interpretations of the results of measurements in a model (see area **f** in Figure 1).[3]

As it was already mentioned in the introduction, this handbook addresses various slightly different topics from the field of technical communication. Many of these topics comprise at least one concept or object for which methods of evaluation have been proposed or established, but these differ drastically depending on their affiliation. As a consequence, the rest of this chapter will act as a *tour d'horizon* through all these areas, with a strong focus on the most interesting aspects of evaluation. Due to the widespread nature of the concept of evaluation it is impossible to provide detailed literature reviews for all areas while keeping this chapter within limits. As a consequence, readers are provided with some elementary publications (for example, introductions or handbooks) for each major area. These should serve as entry points for readers who want to gain deeper insight to some of these areas.

3. Quality criteria for evaluations

As shown in the previous section, evaluation involves measurement operations. Only selected attributes and characteristics of an object of evaluation are taken into account by the chosen method when determining evaluation results. This means that an evaluation must meet a list of so-called *quality criteria* to ensure that its results may be taken serious – and that the assertions and predications about the object of evaluation are true. For such a type of assessment of an evaluation, the term *meta-evaluation* has been coined (Scriven 1969, Stufflebeam 1974, Widmer 1996).

Several institutions and associations have proposed standard collections of such quality criteria. The most prominent of these single criteria are enumerated as follows:

1. *Objectivity* means (in a nutshell) that different (usually human) evaluators yield constant results – or, in other words, that different evaluators base their assessments on an apparatus of methods that are equally available to all of them. We will see that in several methods, objectivity can only be guaranteed to a limited extent, since human abilities (as, for example, judgment, interpretation, opinion and emotion) are required which automatically involve a subjective point of view to the objects under observation. However, in any scientific situation, it is obvious that unnecessary subjectivity should be avoided, and procedures should be executed as objective as possible with respect to given conditions.

2. *Reliability* measures constancy of results in multiple performances of the same type of method or measurement. A procedure is said to be perfectly *reliable* if it always yields equal results for equivalent objects of evaluation, even when procedures are performed repeatedly. In a statistical context, high reliability is equivalent to a low variance of values, distributions or results. From an experimental perspective, reliability is closely related to one reading of the *internal validity* (see below).
3. *Validity*, as the third of three major criteria, originates from the area of statistics where it measures to what extent a result or a measurement represents the measured variables. In the context of evaluation, it is used in quite a similar way: It measures how much the attributions and conclusions from the results of an evaluation are true and applicable to the object of evaluation.

As an example, the well-known intelligence quotient (IQ) is a measure for the still rather elusive concept of human intelligence. Its validity is doubted by many who argue that it is hardly possible to condense such a complex thing as intelligence into a single numeric value. On the basis of such an argumentation, it can hardly be regarded valid as a global intelligence measure.

In the beginning of this chapter it was shown that there is not one general definition of evaluation. Also, the concept of validity does not have one single, established definition, and different authors have used the term and its derivations in different ways. Especially the notions of *internal* and *external validity* can be rather confusing.

In one reading from the area of empirical studies, *internal validity* is given if it can be shown that variation in a dependent variable cannot be caused by anything else than a variation in an independent variable. In this reading, internal validity is closely related to *reliability*: When independent variables are fixed and the system is internally valid, *then* evaluation results must be stable and have a low variance, leading to a high reliability. This is the reading that, among others, Campbell (1957) proposes.

Krippendorff (1980) uses a slightly more general meaning: For him, internal validity

> employs criteria that are internal to an analysis and evaluates whether research findings have anything to do with the data at hand without saying what.
> *Krippendorff (1980), p. 156*

Instead of focusing on specific variables, only evidence for a relation between data and results is required. This reading targets the *consistency* of the underlying theoretical system and the correctness of its application rather than correlations and dependencies between variables.

External validity is closely related to the concept of *generalizability*. An object of evaluation is externally valid if conclusions that have been drawn on the basis of the object may also be applied to a larger sample or population.

4. Objects of evaluation

4.1. Components

If you take one of the more prominent modern models of communication (as, for instance, the model fragment in Figure 2), you probably will be able to identify the following notions and components in it (although they might bear different names):

1. Content, data, messages, or units of information transported, exchanged or delivered during communication (M),
2. Modes (or modalities, depending on the naming convention), coding systems, channels, and media by which communicative processes are established (C),
3. and finally, communicators, participants, interlocutors, and agents involved in communication (either human or artificial; A and B).

In accordance to this handbook's introduction, we can call a communication *technical* in a broader sense if at least one of these three areas is associated with, augmented by, modified or replaced with a technical aspect or substitute. You can find lists of examples for each of these components in Table 1. But that list can only provide a first overview. The following subsections go into detail: they describe various components of evaluation from the field of technical communication according to this schema.

4.2. Content, data, messages, and information

In the following section, areas are described where technical concepts are subject or topic of communication (component M in Figure 2), including the fields of technical support, technical translation and documentation.

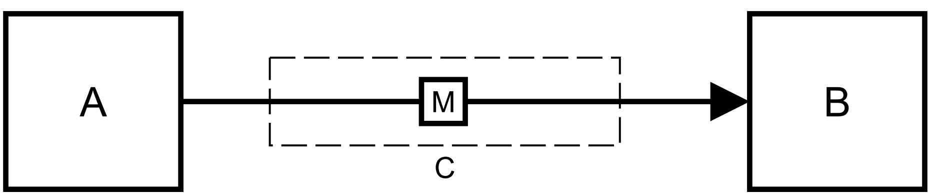


Figure 2. A simplified model of communication that is a subset of many modern complex models, consisting of two participants A and B who communicate by transmitting a piece of information or message M via a medium or channel C. This model can especially be interpreted as a simplified version of the model presented in Shannon and Weaver (1949). It does, however, not claim to be exhaustive for the purpose of actual linguistic research. Instead, it is supposed to show only how the three main components relate to each other.

Table 1. Overview of main components of communication with some examples from the field of technical communication.

COMPONENT	EXAMPLES
content or topic	technical communication, technical documentation, other areas where technical objects serve as *subject* or *topic* of conversation
channel or medium	computer-mediated communication and other novel devices, methods and procedures that facilitate, improve or influence codification and transmission of communicative content
participants	artificial agents, speech processing systems, dialog systems, and other novel devices and objects that take part in interaction as communicators

4.2.1. Technical support, customer dialogs, technical documentation

In a business context, dialogs with customers are evaluated to obtain information about the efficiency of employees as well as customers' needs and demands. Two major approaches are:

1. analyses of conversation records (this could be audio recordings, or the contents of so-called tickets: data structures representing communication between customer and support, usually starting with an initial question or report, and including all subsequent answers), and
2. demanding feedback from customers regarding quality of and satisfaction with the communication.

Technical documentation can also be evaluated. Since most documents from this area already follow rather strict guidelines, evaluation plays a central role during their creation.

See Slem and Cotler (1973) and Gulati and Malcolm (2001) for details about call centers and hotlines, or Muther (1999) and O'Neill et al. (2001) for more information on computer-mediated approaches to customer care. ISO 10002:2004 is an ISO standard dealing with quality management and customer complaints. Finally, Byrne (2006) offers an overview of assessing procedures in the field of technical translation.

4.2.2. Content as object of evaluation

When people make content (blog posts, web pages, tweets, articles, media such as video and audio files, product descriptions, answers with matching solutions, etc.) accessible in the internet, they usually want to get feedback or to initiate conversations about these items. Thus, content in the web can be augmented with evaluation devices or widgets. A common method are rating widgets that usually let users select a numeric value from 1 to 5 stars (Hofmann 2001, Marlin 2004), although currently more sophisticated methods have been explored that do not require explicit input by users (Claypool et al. 2001).

For more detailed information, see the chapter on internet-based communication in this handbook (Waltinger and Breuing 2012).

4.3. Media and channels

Messages and units of information have to be transported or transmitted via a channel or a medium, and in a specific encoding or modality (component C in Figure 2). There are several publications and theories concerning multimodal communication, each defining these terms differently (for instance, see De Ruiter 2004 or Kress and Van Leeuwen 2001 as well as Martin and Schultz and Lücking and Pfeiffer both in this handbook). Therefore, in this section I do not want to take this discussion further. Instead, I set a focus on the interactive web as a novel medium over which people can communicate.

4.3.1. Web analytics

With the advent of the internet, commerce and business have claimed their territory in the web, and they have even more gained in importance and influence. As a consequence, many professions, branches and concepts emerged in only few years. One of them is the notion of *web analytics* (Hassler 2008). This refers to a pool of methods that assist in the optimization of websites, commercial online presences and shops under various aspects, most of which count among evaluative procedures.

In *search engine optimization* (Erlhofer 2010, in German), websites are modified in such a way that search engines assign them a higher rank, which results in an earlier position in search results. The sets of criteria include the quality of metadata, accessibility by search engine robots, linkage degree from and to other websites, but also analyses of website structure and content.

In the area of *website optimization* (King 2008), the structure and layout of complete websites as well as single web pages is evaluated. Several methods have been established, ranging from layout-focused methods like *eye tracking* over to *statistical* tools like Google Analytics (Fang 2007). This system collects large amounts of statistics of visitors, enabling analysts to track their complete visits as sequences of visited pages, along with sophisticated reports built on top of these data. Intended results are called goals, and the process described here is also known as *goal tracking*. One of these goals can be for a normal visitor to become a buying customer through some causes immanent to the online presence. This prominent goal type, and the overall process that caused it is known as a *conversion*.

One prominent measure for the overall importance and relevance of a complete website is the numeric Google PageRank measure (Langville et al. 2008), as described by its inventors in Page et al. (1998). Although Google's own modified variant of the PageRank system cannot be accessed publicly, there are numerous other web services that can retrieve (or at least estimate) the PageRank value for a given internet address, making this important measure available to website evaluators.

4.4. Communicators, participants and interlocutors

In the last decades, great effort has been put into the design of artificial agents and communicators. A large variety of such systems has been presented, ranging from humanoid robots (that also incorporate modalities such as mimics and proxemics), over virtual agents and avatars to sophisticated dialog systems. All these have in common that they can act as communicators and interact with humans (therefore acting as components A and B in Figure 2).

4.4.1. Speech processing systems and artificial agents

When systems are designed that adopt or perform human(-like) behavior (either as speech producers, as in text-to-speech-systems, or as speech recipients, as in voice command software, or both, as in dialog systems), it is usual that their evaluation is an obligatory phase during their development. Designers of such a system usually aim at a typical range of goals, including the optimization of

1. its processing of natural language on various levels (e.g., the phonetic, morphological, lexical, syntactic, semantic or prosodic level);
2. its knowledge (the size of its data pool, or the scope of its competence);
3. its flexibility (for example, its reaction to unexpected input);
4. its look and appearance, and other effects on dialog partners (systems that explicitly attempt to imitate human characteristics are usually evaluated with respect to the success of those imitations);
5. its partner model (for example, its ability to identify people by their voice, or to model the contents of previous conversations or the level of knowledge of the conversational partner);

and various other aspects. Each of them can be in the focus of an evaluation (combinations are also possible and usual), and depending on the aspect, different methods are to be applied.

See Volume III of Gibbon et al. (1997) for a thorough introduction of the assessment of spoken language systems. Paek (2001) deals with the evaluation of language and dialog systems. In addition, Gibbon et al. (1997) in toto, Huang et al. (2001), and Gibbon et al. (2002) offer a general overview of dialog systems and spoken language systems. For more detailed information, see also the chapters on speech technology (Martin and Schultz 2012), on resources for technical communication (Gibbon 2012), and on artificial interactivity by Kopp and Wachsmuth in this handbook.

4.4.2. Information retrieval

Information retrieval is a collection of theories and mechanisms that enables users to find texts, documents, and other pieces of information relevant for a given question, query, or search criterium. Prominent instances of information retrieval systems are search engines in the internet. But most other systems that allow for the performance of searches or queries also class among them.

Information retrieval (at least in the web context) usually has to deal with large amounts of data, so efficient heuristics and methods for acquisition and analysis are required. These, however, cannot be summarized in the scope of this chapter. See Baeza-Yates et al. (1999) and Manning et al. (2008). For more detailed information (especially for the field of technical communication), see also the chapter on document classification, information retrieval, and text mining in this handbook (Paaß 2012).

The performance of an information retrieval system is usually evaluated with quantitative methods. Several measures and related concepts have been established, such as measures of retrieval effectiveness (Salton and McGill 1983, p. 159), like precision, recall, and one of their combinations, the F-measure (see below).

In addition, feedback by users can also be taken into account: Especially in the internet context, users can assess results with answers like “useful” or “irrelevant”, thus delivering additional, authentic input to information retrieval systems which might enhance them in later stages (cf. section 4.2).

4.4.3. Human communicators

Human communicators can also be *objects* of evaluation. Results in this area can be of potential use for fields like web marketing, psychology, or personnel administration. As in many other areas, the internet with its innumerable amounts of data is one of the most promising sources in this area.

Visitors and customers in online shops like Amazon (Linden et al. 2003) leave (e)valuable traces when they browse through articles and lists in the firm’s online presence. Most internet users should be familiar with recommendations (Claypool et al. 2001) based on items that were bought previously, or on the basis of the visit history. In this case, visited product pages and data of previous purchases are evaluated in order to show a customer those products that interest her most – in the assumption that she will purchase the offered products more easily than arbitrary other products found by random browsing or searching.

Similarly, in social networks like Facebook or Twitter and other web applications, posts, entries, and other texts generated by users can be analyzed to find content, users, and other objects that are likely to interest them. This can be done by means of network analysis (Borgatti et al. 2009) as well as content analysis (Neuendorf 2002), and data or text mining (Hearst 2003, Feldman and Sanger 2007).

5. Methods

5.1. Types of methods

The field of methods useful in evaluation is large, and for every particular case, different methods are appropriate. To keep track of such a large pool, methods can be grouped by different criteria, and along multiple dimensions. Methods can be *quantitative*, meaning (in a nutshell) that they use numeric measures (see Figure 1). They map objects of evaluation to numbers or structured data (e.g., matrices or graphs). Typical features of quantitative methods are standardization, automatization and reproducibility. Very often, in quantitative research, software and computer power are used, which enables researchers to analyze large amounts of data if enough time and computer memory is available. One of the most prominent examples of a quantitative method is *statistical hypothesis testing* (see below).

Qualitative methods, on the other hand, focus on content and meaning. They are highly analytic and interpretive, and as a consequence, human intelligence is still required when performing most of them. The measures used in qualitative methods differ from those of quantitative methods. Qualitative methods map objects of study to nominal or, sometimes, ordinal scales. This does not mean that measured values are always unstructured. On the contrary, typical values usually contain a lot of structural information. These can be reports, complex attributions and predications, judgments, and inferences and assumptions on the background and causes of observed data. *Discourse analysis* is a typical example of a qualitative method (see below). Denzin and Lincoln (2000) provide a detailed introduction to qualitative research and its methods.

It must be stated that in almost every distinction, this is not a sharp binary differentiation, but rather a continuum. There are always methods that incorporate properties of both classes. For example, several variants of *expert review* (see below) incorporate both quantitative and qualitative procedures.

Another distinction between methods is their *degree of objectivity*. This, again, correlates with the degree of dependence on human judgments. The more often human experts need to rate data without properly defined measuring methods (thus on the basis of their personal opinion only), the more subjective are the results. In certain areas, these results are vital because they cannot be accessed in any other way. Especially the interpretation of the more elaborate aspects of human language still can only be mastered by human speakers. For instance, the automatic processing and interpretation of metaphors (Ortony 1993 and especially the contribution by Lakoff 1993), of irony or humor (Gibbs et al. 1991), and of speech acts (Searle 1974) still bears problems, although there are promising approaches (Ferrari 1996, Kintsch 2000, Stolcke et al. 2000, Mihalcea and Strapparava 2006). The major down-side of subjective interpretations by humans is a potential lack of validity and reliability. See the section on quality criteria for a more detailed discussion of this issue.

In the following paragraphs, the most prominent methods of evaluation are described along with some references to continuative literature.

5.1.1. Criteria, guidelines, and questionnaires

Checking data against *catalogues of criteria* and *guidelines* is one of the most important and fundamental methods. It serves as a basis for many other methods, but can also be performed as a standalone procedure.

Guidelines can have different structures: They can be present as a fixed sequence of questions to be answered, or they can be designed as a flow diagram, where any given answer has a potential influence on the questions still to come. At the end of such a procedure, there usually is either a single result, or a ranking of some or all possible results, or an unordered set of attributions and characteristics.

Questionnaires are, eventually, a very straightforward presentation of a criteria catalogue. They can contain questions with different means for feedback: Answers can be allowed in single-line text boxes (for numbers, words or short phrases), as free text in multi-line text fields, or as ticks, check marks, or time scale lines (corresponding to values selected from different kinds of sets or scales).

Regularly, questionnaires are filled in by participants who are unaware of experimental design and underlying research questions. It is, therefore not a trivial task to design the layout and wording of such forms, since harmless looking details can have heavy influence on the behavior of participants, and as a consequence, on collected data.

While questionnaires already have a long history in their printed form, they experienced a renaissance upon the arrival of the internet. In electronic form, questionnaires are even more practical, because the phase of manual data acquisition can be omitted (or, at least, reduced), since data is already coded digitally as soon as it is entered into the form.

For more information about guidelines, forms and questionnaires, see Oppenheim (2005) and Bradburn et al. (2004).

5.1.2. Interviewing

Interviewing is another prominent method to obtain data directly from people. An interview is a conversation between an interviewer (who is mostly in charge of order and course of the talk), and one or more interviewees who deliver information.

As with forms and questionnaires, there are many theories and textbooks that give advice on how interviews must be designed to fulfil a certain purpose. One of the major characteristics of an interview is whether it is *guided*: In a guided interview, the interviewer in charge of the choice of topics. She also controls how close interviewees have to stick to the questions asked. In free (or informal) interviews, on the other hand, interviewees are left more authority – they may narrate freely, and sometimes they are allowed to chose subjects of their own accord, or to influence the course of the interview.

It is obvious that guided interviews are preferable if data is supposed to be processed with subsequent quantitative methods, since the degree of data uniformity depends on the interviewer's control over data collection during a guided interview.

A standard work on this subject is Kvale and Brinkmann (2009); see also the corresponding sections in Denzin and Lincoln (2000).

5.1.3. Thinking aloud

A method closely related to interviewing is *thinking aloud*: After an action or an incident, a person is asked to narrate freely about her impressions, thoughts and reflection during that situation (which could be an experiment, the watching of a movie, or a product test). Since there are hardly any controlling influences, individuals can recapitulate freely, which usually accomplishes high insight into that person's thoughts, opinions and impressions related to that situation. Naturally, this kind of data is highly subjective and complex. This method is explained in detail in van Someren et al. (1994).

5.1.4. Expert review

Expert review as a method of evaluation also classes among the subjective methods of evaluation, because a human (thus subjective) observer rates and judges her observations according to her expertise as well as her personal opinions and given criteria. However, experts usually are competent enough to minimize the impact of their personal opinion and focus on more objective features like the occurring patterns or behaviors defined in their guidelines. Exemplary application of expert review is described in Carroll et al. (1972).

5.1.5. Conversation analysis

Conversation analysis is a mostly qualitative approach of the examination of (preferably natural) conversation between two or more participants, with a focus on emergent patterns and structures. Ideal data for conversation analysis are recordings of natural conversations with a setting that reveals the experimental situation as little as possible. This is necessary for the participants to act naturally. In recorded dialogs, experts usually apply qualitative methods. Depending on the theoretical background, they also attempt to find emergent structures, repetitions and correlations of the phenomena that are subject of their research questions.

Although ethnomethodology, an ancestral discipline of conversation analysis, tries to avoid any theoretical assumptions before analyzing data, in conversation analysis there are at least some basal theories and concepts that underpin data analysis. For example, the concepts of turns, especially turn-taking, of sequences of conversational units and of repairs are common, and one of the objects of study is their variations and their distributions in concrete data sets. For more information about conversation analysis, see for example Psathas (1995), Wetherell et al. (2001), ten Have (2007), or McCarthy (2008). Since this is a rather large discipline, there are many other publications that can serve as equally good introductions.

5.1.6. Eye tracking

Eye tracking is the process of observing a person's eye movements in order to reconstruct her gaze directions. During the last years, technology has improved to such an extent that eye tracking devices found their way into scientific research (and into marketing and advertising industry as well).

One approach for visualizing gaze behavior is the so-called heat map. This is an image of the space in front of the participant where the frequency of focussing is coded in one or more special colors. The more often the person looked at an area, the more intense it is colored. These heat maps enable people to get an overview of those areas that got the most attention by the participant. It is also possible to connect subsequent fixations with lines to create a path, thus adding a temporal order of the heat spaces.

But, depending on the circumstances, analysis of raw eye tracking data can also be done, usually in quantitative contexts, where attributes like frequencies or average gaze durations are considered.

See Duchowski (2007) for a general introduction to eye tracking methodology. In Granka et al. (2004), eye tracking is applied to analyze the search behavior of users in the internet, and Jacob and Karn (2003) utilize the method to assess usability in the area of human-computer interaction.

5.1.7. Statistics

Many methods produce nominal or numeric values, in smaller or larger quantities, and with more or less structure attached to them. There are numerous standardized methods in the field of *statistics* that can operate on such data sets in order to produce results, regardless of where these data sets come from. Statistics utilizes probability theory to explain why these data sets (consisting of groups of numeric values) usually are distributed according to one out of several types of distributions, and how and when conclusions may be drawn on the basis of an analysis and comparison of such distributions.

Descriptive statistics is the most basic subdiscipline. It provides methods and measures to describe such distributions with the aid of certain *parameters*. These summarize central properties of a distribution. The most important ones are the *location* and the *dispersion*. Important location parameters are mode, median, and the arithmetic, harmonic and geometric means. Dispersion is usually measured in variance, standard deviation or the interquartile range.

Inferential statistics utilizes these central properties. One possible application is to attest a correlation between two variables. To achieve that, a hypothesis is formulated that postulates some kind of *difference* between two samples or data sets. To be more concise, the hypothesis assumes that this difference occurs in a *dependent* variable if an *independent* variable is actively modified in an experiment.

The inverse of this first hypothesis, the so-called *null hypothesis*, is then formulated, and as soon as this null hypothesis can be rejected or falsified, a strong indication is given that the original hypothesis might be true (although this cannot be proved directly).

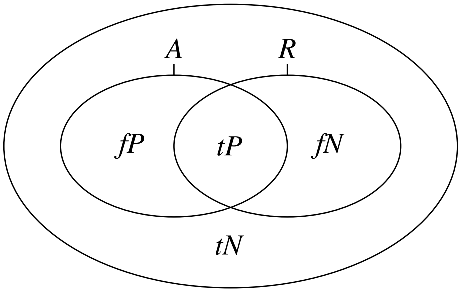
Several *statistical hypothesis tests* for many different constellations of data sets and samples have been developed. Their results can serve as evidence for decisions about the rejection of a hypothesis. In the example above, as soon as a *significant* level of difference is measured, it may be assumed that the null hypothesis can be rejected, which in turn supports the alternative hypothesis. This *significance* can itself be quantified by special formulas and functions that cannot be summarized in this chapter.

These were only a few examples of statistical hypothesis tests. Statistics is a large scientific field, and even elementary publications on the subject can fill several volumes. For introductory texts, see Tabachnick and Fidell (2001) or Box et al. (2005), or any other textbook on the subject.

5.1.8. Measures for effectiveness and quality

In the context of classification tasks, *effectiveness* (Salton and McGill 1983, p. 159) is a term for the average correctness of classification results. There are a number of established measures and concepts especially for binary classifications. Also, information retrieval can be interpreted as a classification task (the classification of documents as either relevant or irrelevant for a given query or search term), enabling users to apply the same mechanisms in the evaluation of information retrieval settings. Both approaches have their own, specific apparatus of terms and concepts, which makes it difficult for novices to understand the interrelations and equivalences between both views. As a consequence, both perspectives are summarized, followed by a comparison of their key concepts.

1. It was already mentioned that in *information retrieval*, the main task is to select a subsets of documents according to given criteria. If a set of documents D is given, and a user wants the system to retrieve all documents that are relevant with respect to a given query (for example, all documents that contain a given search string), then the document set D can be described in two ways:
   (a) It can be seen as the union $D = R \cup I$ of *relevant* and *irrelevant* documents: Relevant documents R have the desired property (and, as a consequence, really are what the user is looking for), whereas irrelevant documents I do not. This separation of D is based on real properties of the documents.
   (b) It can also be seen as a separation of the document set based on the system's opinion: The *answer set* A (the set of documents from D that is returned as the system's answer to the user), along with its complement, the non-answers X. These are documents that have not been collected as answers. The union of answers and non-answers also yields the complete document set: $D = A \cup X$.
2. From a *binary classification* perspective, it is also possible to see these sets as parts of a confusion matrix. Such a matrix groups together all elements according to two factors: Their real class membership (for two classes 1 and 0), and the class membership predicted by the classifier.[4] Since both of these can have exactly two values, the result is a 2×2 matrix, and all elements are distributed along the four fields of this matrix. These four fields are:
   (a) *true positives*: class 1 elements that have been correctly classified as class 1.
   (b) *false positives*: elements that have been classified as class 1, but are in fact class 0 elements.
   (c) *false negatives*: elements that have been classified as class 0, but are in fact class 1 elements.
   (d) *true negatives*: class 0 elements that have been correctly classified as class 0.[5]



		PREDICTION	
		class 1 A	class 0 X
REALITY	class 1 R	$A \cap R$ tP	$X \cap R$ fN
	class 0 I	$A \cap I$ fP	$X \cap I$ tN

Figure 3. Schemas of a Venn diagram (left) and of a confusion matrix for the features, elements, and result subsets of a classification or information retrieval task. See Table 2 for a complete list of all abbreviations. The sets *I* and *X* are not marked in the Venn diagram for the sake of clarity, but they can be obtained by creating the complements of *R* and *A*, respectively.

Table 2. Comparison of key concepts from both the IR and the classification perspectives and their inclusions and interrelations.

IR PERSPECTIVE		CLASSIFICATION PERSPECTIVE	
relevants	R	true positives and false negatives	$tP \cup fN$
irrelevants	I	false positives and true negatives	$fP \cup tN$
answers	A	true positives and false positives = all positives	$tP \cup fP = P$
non-answers	X	true negatives and false negatives = all negatives	$tN \cup fN = \mathrm{N}$
answers ∩ relevants	$A \cap R$	true positives	tP
answers ∩ irrelevants	$A \cap I$	false positives	fP
non-answers ∩ irrelevants	$X \cap I$	true negatives	tN
non-answers ∩ relevants	$X \cap R$	false negatives	fN

See Figure 3 for an overview of all relevant components and subsets, and their connections and relations. In addition, a list of all key concepts from both approaches (along with their abbreviations) can be found in Table 2. On this basis, the quality of an information retrieval system or a classifier is measured by one or more of the three following comparisons:

1. *Focus on positives only*: The size of the true positives set *tP* compared to both the false positives and the false negatives. This type of measure is called the *effectiveness*.

2. *Focus on negatives only*: The size of the true negatives set *tN* compared to both the false negatives and the false positives.
3. *Focus on both positives and negatives*: The size of the set of true or correct results $tP \cup tN$, is compared to the size of the complete document set *D*. This measure is called the *accuracy*.[6]

Effectiveness, to be more concise, is measured by taking three sets and comparing of the proportions of their sizes:

1. The answer set *A*,
2. the set of relevant elements *R*,
3. and their intersection, the set of true positives $tP = A \cap R$.

In order to retrieve a single numeric measurement value, two independent measures are aggregated. Those are called *precision* and *recall* (first defined in Kent et al. 1955, although the terms *precision* and *recall* were not used then).

The *precision p* measures the percentage of correctly classified elements in the answer set:

$$p = \frac{|A \cap R|}{|A|} \left(= \frac{|tP|}{|tP| + |fP|}\right) \tag{1}$$

It can be paraphrased with the question: “How many elements of the elements in the answer set are in fact relevant documents?” As an example: if a system returns ten documents as a result, but two of these documents have been selected in error, then the precision of this retrieval is $p = 8 / 8{+}2 = 0.8$ – in other words, only 80 % of these elements (namely, 8 out of 10 elements) are actually correct.

The *recall r* measures the percentage of retrieved relevant documents in relation to the set of all relevant documents.

$$r = \frac{|A \cap R|}{|R|} \left(= \frac{|tP|}{|tP| + |fN|}\right) \tag{2}$$

In other words: “What percentage of all relevant documents have actually found their way into the answer set?” Another example: If a system puts 4 relevant documents into the result set, but misses another 6 documents that are also relevant, then the recall is $p = 4 / 4{+}6 = 0.4$, because these to sets sum up to 10, and 4 documents are 40 % of all relevant documents.

Precision and recall, as it was already mentioned, measure different properties of a classification. They can be combined into a single value that is a good indicator for the system’s overall effectiveness: The *F*-measure, which is calculated as the weighted harmonic mean of precision and recall:

$$F_w = (1 + w^2) \frac{pr}{w^2 p + r} \tag{3}$$

In this formula, w indicates the weight that is given to recall in relation to precision. For the case $w = 1$, both precision and recall are given the same weight, resulting in the so-called *balanced F*-measure, for which the short notation F and a simplified formula exists:

$$F = 2\,\frac{pr}{p + r} \tag{4}$$

F-values (as well as precision and recall) lie within the range [0,1], where 0 indicates complete failure, while 1 stands for perfect retrieval or classification, where the answer set is equal to the set of relevant documents.

This combined measure makes it possible to rate objects and systems with a single numeric value, regardless of their inner structure, and alone of the basis of their effectiveness. This is one of the major advantages of abstract quantitative methods: A set of numbers is sufficient for them to work, it is totally irrelevant where they originated from, as long as they are correct and adequate.

In other contexts (especially in medical research), the *recall* measure is also called *sensitivity*. In these situations, its negative counterpart, the *specificity*, is equally important. It measures how many of all irrelevant (or class 0) elements have in fact been classified as irrelevant (or class 0). It is calculated as follows:

$$spec = \frac{|X \cap I|}{|I|} \; \left(= \frac{|tN|}{|tN| + |fP|}\right) \tag{5}$$

The *accuracy* is rather similar to the F-measure, but it has the notable difference that it takes both positive and negative classes into account, whereas effectiveness puts a strong focus on positive (or relevant) elements. It is calculated as follows:

$$a = \frac{|A \cap R| + |X \cap I|}{|D|} = \frac{|tP \cup tN|}{|D|} \tag{6}$$

See Salton and McGill (1983) and Baeza-Yates et al. (1999). For more detailed information, see the chapter on document classification and information retrieval in this handbook (Paaß 2012).

5.1.9. Graph and network analysis

Methods in the area of *graph and network analysis* operate on graph structures consisting of nodes and edges. They observe and collect properties of these graphs and draw conclusions on the basis of these properties.

In quantitative network analysis, for example, special quantitative properties of networks are calculated. Multiple networks can then be characterized by a comparison of these properties. Resulting values for similarity or difference can then be the argumentative basis for the confirmation or falsification of given hypotheses.

The set of properties useful as indicators in quantitative network analysis is large. Examples are measures of various lengths inside a network, or measures of compactness, connectedness or density.

The advantage of this method is to obtain additional values and correlations that are observable and emergent only when data is viewed and analysed as a network.

For more detailed information about networks and network analysis, please refer to Newman et al. (2006), Newman (2011) and Mehler (2008).

6. Scenarios

In the previous sections, a variety of objects and methods of evaluation has been presented in isolation. Now some examples will present complete evaluation scenarios that bring together all these components.

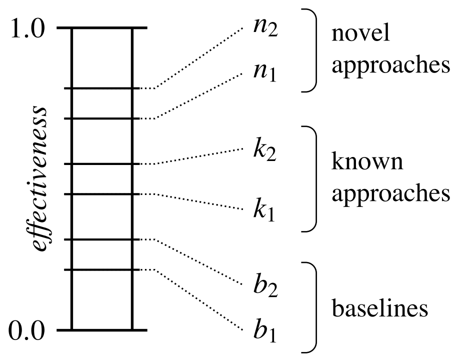
6.1. Evaluation of classification tasks

The evaluation of *classification tasks* (including information retrieval) follows a rather fixed pattern. Some of the phases and procedures have already been described in the previous sections, so in this section, an example of a coherent evaluation will be summarized.

The task of a classifier (see Paaß in this volume) is to extract a correct subset of objects out of a object universe (this is called a binary classification), or to separate such a universe into a correct partition of multiple subsets (or classes) of objects (a multiclass classification). Therefore, it can be evaluated by a sequence of classification instances and by calculating an average value for the effectiveness values of these instances (as it has been described in the previous section).

Although this average effectiveness already indicates a result, at least on a trivial scale between 0 and 1, it is important to compare it to additional contextual values in that interval. These can either be the results of other approaches to the question asked, or so-called baselines.

A *baseline* is a minimal quality that can be achieved by using a trivial, canonical or another relatively effortless approach. For example, in a multiple-choice test with four possible answers for each question (with only one correct answer per question), an average correctness of $\frac{1}{4} = 0.25$ can be reached merely by a random selection of an answer. This method is trivial and does not require significant effort or computational cost, yet its efficiency is considerably higher than 0. Therefore, it can be regarded a baseline for this classification task, and other approaches need to be compared to this baseline rather than to 0. Several other classifications compare their results to a random approach that assigns



Method		*F*-score
novel approach 2	n_2	.800
novel approach 1	n_1	.700
known approach 2	k_2	.550
known approach 1	k_1	.450
baseline 2	b_2	.300
baseline 1	b_1	.200

Figure 4. A fictional, schematic example of classification results in relation to earlier results and baselines. Left: diagram showing the effectiveness of all methods considered. Right: Table view of these values. This is the way results are usually presented. Results of novel methods should outperform both known approaches (if any exist) and baselines to a reasonable degree.

class labels to instances only by chance, without any deep consideration of input data. However, baseline approaches have the option to take various superficial quantities into account, for example the proportion of target classes in relation to the complete data set.

In addition, if other classifiers have already been created for a given task, then it should also be shown that a newly developed system performs significantly better than those systems that are already present. In both cases, classification of selected test data is performed with all classifiers to be compared (for random baselines, there are special algorithms to determine their effectiveness on the basis of the number and sizes of classes). Figure 6.1 shows a ficticious set of results of different classification approaches, along with a comparison to an existing approach and to baselines.

At the other end of the interval, the opposite of a *baseline* is the so-called *gold standard*. It stands for the best classification that can be achieved. This is usually a subset of the object domain that has been annotated manually, or for which class memberships are known through some other process or mechanism. In analogy to the baseline, if the quality of the gold standard itself is lower than 1 (in cases where, for whatever reason, a perfect result of 1 is impossible), then novel methods should be compared to the quality of the gold standard rather than 1.

For example, in the field of syntactic parsing, manually annotated treebanks (sets of sentences that have been annotated syntactically in a tree structure) can serve as gold standard against which the results of an automatic parser can be compared.

6.2. Usability studies

In a *usability study*, research participants interact with the system to be evaluated. The typical scenario is that they use a system without any additional task while experimenters perform measurements which typically involve one or more of the following (depending on the type of system):

1. The duration of the interaction until the task is completed,
2. the number of steps, turns or questions until the task is completed (in dialog systems),
3. the number of clarification questions (in dialog systems),
4. representations or recordings of eye movements and gaze behavior (in systems involving a screen or display),
5. and representations or recordings of mouse movements, typing behavior (in systems with a mouse and/or a keyboard)

A characteristic feature of such a usability study is that participants usually do not have to rate the system themselves – it is their behavior during system usage that implicitly produces evaluation data.

After data has been collected, measurements can be performed. If a single system is evaluated, the resulting values can be used to identify and improve weak points in the system. Subsequent follow-up tests can then be performed and their results can be compared to those of the initial test. This procedure can lead to judgments on whether the system has really been improved by the changes motivated by the initial test.

Another approach is a parallel test of different similar systems that have been preselected because they look promising for the targeted task, but are yet unknown to potential users. In this case, a system with better evaluation results will be given a higher priority in a ranking and will be preferred when it comes to a choice.

6.3. Human evaluation

In *human evaluation*, system users describe and rate their experience with an object of study. This can be achieved in different ways. Normally, users fill out forms and questionnaires with rather constrained questions and rating assignments. These have been designed to rate specific features of the system that have been put in the focus of the current evaluation procedure. However, there are also questions where users may answer in detail, thus enabling them to deliver qualitative feedback rather than quantitative one.

This method is more subjective than usability studies, but it enables researchers to take a much larger number of factors and features into account than those that can be collected by objective observation.

As an example, in Leimeister (2005), a community website is assessed by means of both usability studies and human evaluation.

6.4. Evaluation events

Since conferences and workshops provide opportunity for many researchers to congregate, a special form of parallel evaluation contests evolved, where many developers gather and evaluate their systems on the basis of a fixed collection of prescribed data sets. This bears the advantage of highest possible similarity of input data, leading to relatively high comparability and validity of the results.

During the MUC-7 workshop on message understanding, 18 organizations took part in such an evaluation consisting of six tasks from the field of information extraction (Marsh and Perzanowski 1998, Marsh 1998). Similarly, the SENSEVAL evaluation exercises (Edmonds 2002; see also http://www.senseval.org/) cover the field of word sense disambiguation. The current task (SemEval-2) was held in 2010 (see http://semeval2.fbk.eu/).

7. Reliability assessments

Especially in fields where human judgments play an important role, several methods for the assessment of *reliability* have been proposed. In a nutshell, these methods can measure either the agreement or the difference between two data sets, and on that basis, average data sets with a considerably lower degree of subjectivity can be induced.

A common field is the manual annotation (or labeling) of a set of items. It is common practice to design annotation guidelines that give detailed conditions, hints and other instructions for the usage of the different categories. However, these cannot replace a residual amount of human judgment. This should be obvious, because if the need for human judgment could be eliminated completely, then these guidelines could easily be replaced by an implementation of a heuristic or an algorithm that performs a categorization automatically.

As a consequence, deviating categorizations are still possible (although good guidelines should decrease their frequency). The reliability of annotations on the basis of these guidelines can then be measured by comparing multiple (usually two) sets of categorizations of the same primary data. One popular approach is the *kappa* statistic (Carletta 1996, but see also Stegmann and Lücking 2005 and Artstein and Poesio 2008 for thorough discussions of *kappa* and several other measures).

As with the evaluation of classification tasks, researchers try to yield as high a reliability as possible. This process usually takes place in a cycle where annotation guidelines are modified and improved on the basis of such results of reliability assessments.

8. Conclusion

This chapter gave an introductory but widespread overview of various examples of evaluation in the context of technical communication (with all its varying meanings as they were introduced in the introduction of this handbook). The following things should have become clear:

1. There is no such thing as a complete, generic concept or template of an evaluation. Although different evaluative procedures regularly have many things in common, they can also differ fundamentally.
2. Evaluation can be structured by different aspects: By type and nature of the object of evaluation, by the expected or intended type of results, and by the methods applied (which very often are predetermined by the combination of objects and results). Such an approach was used to structure this chapter, where types of objects and methods have been presented separately.
3. Methods are (in most cases) specialists. They are suitable for specific object-result-constellations, where they can produce highly valuable output – but they are likely to fail in other contexts. Therefore, the choice and combination of methods is vital.
4. Evaluation itself should be evaluated by checking it against a catalogue of quality criteria. Results should then be appraised according to their quality.

Finally, for deeper insights regarding evaluation in general, refer to the aforementioned thesaurus Scriven (1991), and, with a focus on computational linguistics, Hirschman and Mani (2003).

Notes

1. On the other hand, this definition covers the fact that "evaluation" (as many other terms) has both a processual and a result-oriented reading: It can either refer to a process or to the result or final state of this process.
2. This work offers a very detailed introduction to modeling and measuring in the field of linguistics (and beyond). It actually gives a more complex formal definition for such a mapping, but for the scope and the audience of this chapter, this more informal one presented in the text should suffice.
3. For some authors, however, the drawing of conclusions is already part of the definition of a model (for example, see Stachowiak 1989).
4. One can easily switch back and forth between this view and the information retrieval perspective by substituting "class 1" with "relevant" and "class 0" with "irrelevant", respectively.

5. Thus, the terms “positive” and “negative” refer to the answers given by the classifier, whereas “true” and “false” indicate whether the classification was correct or not.
6. Note, however, that some theories (for example, Ferber 2003) use the term “accuracy” as a synonym for the concept of *precision* that will be described in the following paragraphs. It is therefore advisable to determine which reading of accuracy an author uses.

References

Altmann, Gabriel and Rüdiger Grotjahn
1988 Linguistische Meßverfahren. In Ulrich Ammon, Norbert Dittmar, and Klaus J. Mattheier, editors, *Sociolinguistics. An International Handbook of the Science of Language Society*, pages 1026–1039. de Gruyter, Berlin / New York.

Artstein, R. and M. Poesio
2008 Inter-coder agreement for computational linguistics. *Computational Linguistics*, 34(4):555–596.

Baeza-Yates, R., B. Ribeiro-Neto, et al.
1999 *Modern information retrieval*. Addison-Wesley, Boston.

Borgatti, S.P., A. Mehra, D.J. Brass and G. Labianca
2009 Network analysis in the social sciences. *Science*, 323(5916):892.

Box, G.E.P., J.S. Hunter and W.G. Hunter
2005 *Statistics for experimenters: design, innovation, and discovery.* Wiley-Interscience, New York.

Bradburn, Norman, Seymour Sudman and Brian Wansink
2004 *Asking questions*. Jossey-Bass, San Francisco.

Byrne, Jody
2006 *Technical translation. Usability strategies for translating technical documentation*. Springer, Dordrecht.

Campbell, D. T.
1957 Factors relevant to the validity of experiments in social settings. *Psychological Bulletin*, 54(4):297.

Carletta, J.
1996 Assessing agreement on classification tasks: the kappa statistic. *Computational linguistics*, 22(2):249–254.

Carroll, S.J., F.T. Paine and J.J. Ivancevich
1972 The relative effectiveness of training methods. expert opinion and research. *Personnel Psychology*, 25 (3):495–509.

Claypool, M., P. Le, M. Wased and D. Brown
2001 Implicit interest indicators. In *Proceedings of the 6th international conference on Intelligent user interfaces*, pages 33–40. ACM.

De Ruiter, J.P.
2004 On the primacy of language in multimodal communication. In *The Workshop Programme Multimodal Corpora: Models Of Human Behaviour For The Specification And Evaluation Of Multimodal Input And Output Interfaces*, pages 38–41, Lisboa.

Denzin, N.K. and Y.S. Lincoln
2000 *Handbook of Qualitative Research*. Sage Publications, Thousand Oaks.

Duchowski, Andrew T.
2007 *Eye tracking methodology. Theory and practice*. Springer, London.
Edmonds, P.
2002 Introduction to Senseval. *ELRA newsletter*, 7.
Erlhofer, Sebastian
2010 *Suchmaschinen-Optimierung für Webentwickler. Grundlagen, Funktionsweisen und Ranking-Optimierung*. Galileo Press, Bonn.
Fang, W.
2007 Using Google Analytics for improving library website content and design: a case study. *Library Philosophy and Practice*, 9(3):1–17.
Feldman, Ronen and James Sanger
2007 *The Text Mining Handbook. Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press, Cambridge.
Ferber, Reginald
2003 *Information Retrieval: Suchmodelle und Data-Mining-Verfahren für Textsammlungen und das Web*. Dpunkt-Verlag, Heidelberg.
Ferrari, S.
1996 Using textual clues to improve metaphor processing. In *Proceedings of the 34th annual meeting on Association for Computational Linguistics*, pages 351–353. Association for Computational Linguistics.
Gibbon, D., R. Moore and R. Winski (eds.)
1997 *Handbook of standards and resources for spoken language systems*. Mouton de Gruyter, Berlin.
Gibbon, D., I. Mertins, R. Moore and J. Flanagan (eds.)
2002 *Handbook of multimodal and spoken dialogue systems: resources, terminology and product evaluation*. Kluwer, Boston.
Gibbon, Dafydd
2012 Resources for technical communication systems. In Alexander Mehler, Laurent Romary, and Dafydd Gibbon, editors, *Handbook of Technical Communication*. De Gruyter, Berlin/Boston.
Gibbs, R.W. et al.
1991 Psychological aspects of irony understanding. *Journal of Pragmatics*, 16(6):523–530.
Granka, L.A., T. Joachims and G. Gay
2004 Eye-tracking analysis of user behavior in WWW search. In *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 25–29.
Gulati, Sandeep and Scott A. Malcolm
2001 Call center scheduling technology evaluation using simulation. In *WSC '01: Proceedings of the 33nd conference on Winter simulation*, pages 1438–1442, Washington, DC, USA. IEEE Computer Society.
Hassler, M.
2008 Web analytics. *Redline, Heidelberg*.
Hearst, Marti
2003 Text Data Mining. In Ruslan Mitkov, editor, *The Oxford Handbook of Computational Linguistics*, chapter 34, pages 616–628. Oxford University Press, Oxford.

Hirschman, Lynette and Inderjeet Mani
2003 Evaluation. In Ruslan Mitkov, editor, *The Oxford Handbook of Computational Linguistics*, chapter 22, pages 414–429. Oxford University Press, Oxford.

Hofmann, T.
2001 Learning what people (don't) want. *Machine Learning: ECML 2001*, pages 214–225.

Huang, X., A. Acero and H.W. Hon
2001 *Spoken language processing: A guide to theory, algorithm, and system development*. Prentice Hall PTR, Upper Saddle River.

ISO 10002:2004
2004 Qualitätsmanagement – Kundenzufriedenheit – Leitfaden für die Behandlung von Reklamationen in Organisationen.

Jacob, Robert J. K. and Keith S. Karn
2003 Eye Tracking in Human-Computer Interaction and Usability Research: Ready to Deliver the Promises. In *The mind's eye. Cognitive and applied aspects of eye movement research*, pages 573–605. Elsevier, Amsterdam.

Kent, Allen, Madeline M. Berry, Fred U. Luehrs Jr and J. W. Perry
1955 Machine literature searching VIII. Operational criteria for designing information retrieval systems. *American documentation*, 6(2):93–101.

King, A.B.
2008 *Website optimization*. O'Reilly, Sebastopol.

Kintsch, W.
2000 Metaphor comprehension: A computational theory. *Psychonomic Bulletin & Review*, 7(2):257–266.

Kopp, Stefan and Ipke Wachsmuth
2012 Artificial interactivity. In Alexander Mehler, Laurent Romary, and Dafydd Gibbon, editors, *Handbook of Technical Communication*. De Gruyter, Berlin/Boston.

Kress, G. and T.V. Van Leeuwen
2001 *Multimodal discourse: The modes and media of contemporary communication*. Hodder, London.

Krippendorff, Klaus
1980 *Content analysis. An introduction to its methodology.* SAGE.

Kvale, Steinar and Svend Brinkmann
2009 *InterViews: Learning the craft of qualitative research interviewing*. Sage Publications, Thousand Oaks.

Lakoff, G.
1993 The contemporary theory of metaphor. In A. Ortony, editor, *Metaphor and Thought*, volume 2, pages 202–251. Cambridge.

Langville, A.N., C.D. Meyer and P. Fernández
2008 Google's PageRank and beyond: the science of search engine rankings. *The Mathematical Intelligencer*, 30(1):68–69.

Leimeister Jan Marco
2005 Evaluation of a Systematic Design for a Virtual Patient Community. *Journal of Computer-Mediated Communication*, 10(4).

Linden, G., B. Smith and J. York
2003 Amazon.com recommendations: Item-to-item collaborative filtering. *IEEE Internet computing*, 7(1):76–80.
Lücking, Andy and Thies Pfeiffer
2012 Framing multi-modal technical communication. In Alexander Mehler, Laurent Romary, and Dafydd Gibbon, editors, *Handbook of Technical Communication*. De Gruyter, Berlin/Boston.
Manning, C.D., P. Raghavan and H. Schütze
2008 *Introduction to Information Retrieval*. Cambridge University Press, New York.
Marlin, B.
2004 Modeling user rating profiles for collaborative filtering. *Advances in neural information processing systems*, 16:627–634.
Marsh, E.
1998 TIPSTER information extraction evaluation: the MUC-7 workshop. In *Proceedings of a workshop on held at Baltimore, Maryland: October 13–15*, pages 233–234. Association for Computational Linguistics.
Marsh, E. and D. Perzanowski
1998 MUC-7 evaluation of IE technology: Overview of results. In *Proceedings of MUC*, volume 7.
Martin, Jean-Claude and Tanja Schultz
2012 Multimodal and speech technology. In Alexander Mehler, Laurent Romary, and Dafydd Gibbon, editors, *Handbook of Technical Communication*. De Gruyter, Berlin/Boston.
McCarthy, M.
2008 Discourse analysis. *ELT journal*, 62(2):211.
Mehler, Alexander
2008 Structural similarities of complex networks: A computational model by example of wiki graphs. *Applied Artificial Intelligence*, 22 (7&8):619–683.
Mihalcea, R. and C. Strapparava
2006 Learning to laugh (automatically): Computational models for humor recognition. *Computational Intelligence*, 22(2): 126–142.
Muther, Andreas
1999 *Electronic Customer Care. Die Anbieter-Kunden-Beziehung im Informationszeitalter*. Springer, Berlin.
Neuendorf, Kimberly A.
2002 *The Content Analysis Guidebook*. Sage Publications, Thousand Oaks.
Newman, Mark E. J.
2011 *Networks*. Oxford University Press, Oxford.
Newman, M.E.J., A.L. Barabasi and D.J. Watts
2006 *The structure and dynamics of networks*. Princeton University Press, Princeton.
O'Neill, Martin, Christine Wright and Frank Fitz
2001 Quality evaluation in online service environments: an application of the importance-performance measurement technique. *Managing Service Quality*, 11:402–417.

Oppenheim, Abraham N.
2005 *Questionnaire design, interviewing and attitude measurement*. Continuum, London.

Ortony, A.
1993 *Metaphor and thought*. Cambridge University Press.

Paaß, Gerhard
2012 Document classification, information retrieval, text and web mining. In Alexander Mehler, Laurent Romary, and Dafydd Gibbon, editors, *Handbook of Technical Communication*. De Gruyter, Berlin/Boston.

Paek, T.
2001 Empirical methods for evaluating dialog systems. In *Proceedings of the workshop on Evaluation for Language and Dialogue Systems-Volume 9*, pages 1–8. Association for Computational Linguistics.

Page, L., S. Brin, R. Motwani and T. Winograd
1998 The pagerank citation ranking: Bringing order to the web. Technical report, Stanford Digital Library Technologies Project.

Psathas, G.
1995 *Conversation analysis: The study of talkin-interaction*. Sage Publications, Thousand Oaks.

Salton, Gerard and Michael J. McGill
1983 *Introduction to Modern Information Retrieval*. McGraw-Hill, New York.

Scriven, Michael
1969 An introduction to meta-evaluation. *Educational Product Report*, 2(5):36–38.

Scriven, Michael
1991 *Evaluation thesaurus*. Sage, Newbury Park.

Searle, J.R.
1974 *Speech acts: An essay in the philosophy of language*. Cambridge University Press.

Shannon, C.E. and W. Weaver
1949 *The mathematical theory of information*. University of Illinois Press, Urbana.

Slem, Charles M. and Sheldon Cotler
1973 Crisis phone services: Evaluation of a hotline program. *American Journal of Community Psychology*, 1:219–227.

Stachowiak, H.
1965 Gedanken zu einer allgemeinen Modelltheorie. *Studium Generale*, 18(7):432–463.

Stachowiak, H.
1989 Modell. In H. Seiffert and G. Radnitzky, editors, *Handlexikon zur Wissenschaftstheorie*, pages 219–222. Ehrenwirth, München.

Stegmann, Jens and Andy Lücking
2005 *Assessing Reliability on Annotations (1): Theoretical Considerations*. Technical Report No. 2. SFB 360, Bielefeld University.

Stolcke, A., K. Ries, N. Coccaro, E. Shriberg, R. Bates, D. Jurafsky, P. Taylor, R. Martin, C.V. Ess-Dykema and M. Meteer
2000 Dialogue act modeling for automatic tagging and recognition of conversational speech. *Computational linguistics*, 26(3):339–373.

Stufflebeam, Daniel L.
1974 *Meta-evaluation: The Evaluation Center*. Number 3 in Occasional paper series. Western Michigan University Press, College of Education, Kalamazoo.

Tabachnick, B.G. and L.S. Fidell
2001 *Using multivariate statistics*. Allyn and Bacon, Boston.

ten Have, P.
2007 *Doing conversation analysis*. Sage Publications, Thousand Oaks.

van Someren, Maarten W., Yvonne F. Barnard, Jacobijn A. C. Sandberg, et al.
1994 *The think aloud method: A practical guide to modelling cognitive processes*. Academic Press, London.

Waltinger, Ulli and Alexa Breuing
2012 Internet-based communication. In Alexander Mehler, Laurent Romary, and Dafydd Gibbon, editors, *Handbook of Technical Communication*. De Gruyter, Berlin/Boston.

Wetherell, M., S. Taylor and S. Yates
2001 *Discourse as data: A guide for analysis*. Sage Publications, Thousand Oaks.

Widmer, Thomas
1996 *Meta-Evaluation. Kriterien zur Bewertung von Evaluationen*. Haupt, Bern.

Widmer, Thomas, Wolfgang Beywl and Fabian Carlo (eds.)
2009 *Evaluation. Ein systematisches Handbuch*. Verlag für Sozialwissenschaften, Wiesbaden.

II. Technical communication management

9. Lexicography, terminology and ontologies

Serge Sharoff and Anthony Hartley

1. Words and concepts

A text – for example, this chapter – consists of sequences of characters usually separated by punctuation marks and white spaces. Humans reading this text interpret such sequences as words having particular meanings. The relationship between words and objects, both physical and abstract, can be illustrated by the semiotic triangle (Ogden and Richards, 1923), which was first introduced independently by Charles Sanders Pierce and Gottlob Frege at the end of the 19th century and later popularised by Ogden and Richards (see Figure 1). The semiotic triangle has three vertices corresponding to the word, its interpretation (concept or meaning) and the physical or abstract object it refers to (referent). The line between word and object is dotted, since a linguistic expression does not point to an object directly, but only via its interpretation. The word may be a single-word or a multi-word expression. The denotational form of concepts can change over time (e.g., *on line, on-line, online*) and is subject to the rules of word formation of a particular language (e.g., *watershed, ligne de partage des eaux*) without affecting its relationship to the referent. Reference, then, is a function that identifies objects in a domain. This model is clear and useful for many applications within the scope of this chapter. In particular, it helps us distinguish two perspectives for studying the relationship between words and meanings: the *semasiological* perspective starts from the linguistic form of the sign to ask how it is understood, while the *onomasiological* perspective starts from the content of the sign to ask how it is delimited and named.

The realm of ontology construction is the system of concepts itself and its relationship to objects; the emphasis is onomasiological. Terminologists are concerned with what Sager calls 'special languages': "semi-autonomous, complex semiotic systems […] derived from general language [whose] use […] is restricted to specialists for conceptualisation, classification and communication in the same or closely related fields" (Sager, 1993, p. 44). Thus the *terminology* of a discipline is an instrument of 'special reference'. While the traditional approach to terminological activity adopts the onomasiological perspective, this view that concepts exist objectively prior to being named is challenged by advocates of a sociocognitive approach, e.g., (Kageura, 1995; Temmermann, 2000), who argue for equal emphasis on the semasiological perspective to uncover real usage in the 'parole' of a subject community. Lexicographers are concerned not with specialised areas of knowledge but with general knowledge,

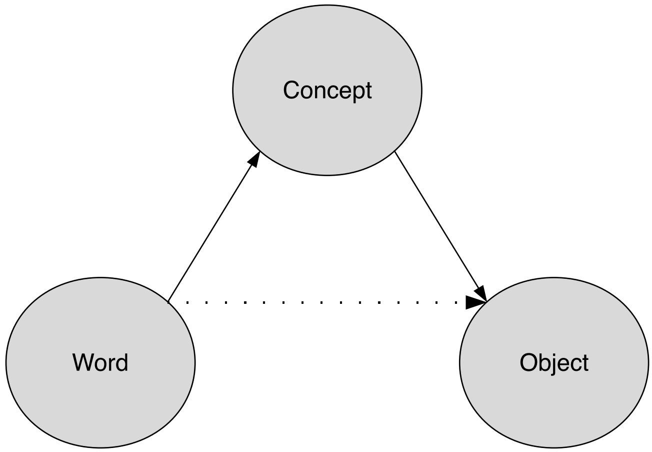


Figure 1. Semiotic triangle.

that is, with the *vocabulary* of 'general reference'; they take the semasiological perspective. In spite of these differences between the goals and methods of ontologists, terminologists and lexicographers, their activities share a common underlying task: to extract meanings from texts and represent them in a form appropriate to their end use.

It is tempting to view the set of interpretations of signs as a finite list, so that we could imagine enumerating the complete set of concepts existing in a given domain and the set of words or terms referring to them. However, when we apply the semiotic triangle to the use of words in the real world, many distinctions become blurred. Let us take an example from computer science of a number of 'mechanisms' with extensive abilities for processing text data. A webpage in HTML is not normally considered to be written in a *programming language*, yet a webpage can contain Javascript, which is normally referred to as a programming language. Again, while XSLT (eXtensible Stylesheet Language Transformations) is not commonly thought of as a programming language, a reference of this sort is possible in some contexts. On the other hand, CSS (Cascading Style Sheets) and LaTeX are not considered as programming languages at all. In other words, in many contexts it is difficult to determine the identity of a concept and the set of words that can be used to refer to it.

An aspect of representing the complex relationship between concepts and words concerns the need to accommodate the expectations and limitations of different audiences. For instance, the entries for *genome* in the Oxford Advanced Learners Dictionary (OALD) and the Concise Oxford English Dictionary (COED) define the concept using different words:

(1) **OALD** *the complete set of GENES in a cell or living thing: the human genome.*
(2) **COED** *1 the haploid set of chromosomes of an organism. 2 the complete set of genetic material of an organism.*

Although both definitions are intended to identify the same referent, the wordings and therefore contents are different, because the OALD is designed for non-native speakers of English while the COED aims to provide more comprehensive definitions.

Different language communities or communities of practice may operate with different systems of concepts, which makes even more difficult the task of determining the reference of a sign and its relationship to signs in another language. For example, Gottlob Frege described the relation between signs and meanings as:

(3) Es liegt nun nahe, mit einem Zeichen (Namen, Wortverbindung, Schriftzeichen) außer dem Bezeichneten, was die *Bedeutung* des Zeichens heißen möge, noch das verbunden zu denken, was ich den *Sinn* des Zeichens nennen möchte, worin die Art des Gegebenseins enthalten ist. (G.Frege, Sinn und Bedeutung)
(4) It is natural, now, to think of there being connected with a sign (name, combination of words, letters), besides that to which the sign refers, which may be called the *reference* of the sign, also what I should like to call the *sense* of the sign, wherein the mode of presentation is contained. (G.Frege, On Sense and Reference, translated by M.Black)

In their general reference occurrences, the two words italicised in the original German text (*Sinn* and *Bedeutung*) are both normally translated into English as *meaning*. However, the distinction drawn by Frege facilitated the development of a special reference in English which had not existed before Frege's text became influential. This illustrates how systems of terminology evolve over time, not only because of the invention of new artefacts and new scientific discoveries, but also because of the need to make distinctions important for a particular theory. Given that such development happens through the medium of language, language has a direct impact on the system of concepts, as well as on the context of their use. (Temmermann, 2000, p. 33) stresses the importance for terminologists of the semasiological perspective: "In a scientific community meaning is negotiated. A concept is not really recognised as such nor taken seriously unless it is named by a term." In linguistic research this is known as logogenesis, i.e., meaning making in the process of development of a discipline, a research area, or an individual text. Halliday and Matthiessen metaphorically refer to this fact as "there are no 'naked ideas' lurking in the background waiting to be clothed; it is language that *creates* meaning" (Halliday and Matthiessen, 1999, p. 602).

One important aspect of logogenesis concerns ways of naming new artefacts and constructs of human cognition, including:

- metaphoric extension of existing terms (*file* on a computer, *rib* of a tyre, *sense* and *reference* in semantics);
- borrowing from other languages (*ketchup, poltergeist, sputnik*);
- neoclassical creation (*vulcanisation, oscillograph, carcinoma, bronchitis, biotechnology*);
- prefixing (*pre-production, supraorbital, thermocouple*);
- compounding (*flywheel, sprocket wheel, fluid power transmission*).

The link between the world and words is further complicated by lexical expression of some grammatical categories, for example, definiteness, which in English or German is expressed by articles (*a, the*), which do not exist in Chinese or Russian. Even if articles are used in a language, their use varies from one language to another. For instance, 'abstract' nouns in French and German take a definite article (e.g., *la paix est …; der Frieden*), in English they do not (*peace is …*).

Each domain has its own terminology (concepts and lexicalisations) and ways for using it. For example, the language of weather forecasts is considerably different from the language of atmospheric physics, even if their subject matter is, broadly speaking, the same. There is a long tradition of terminological studies of specific disciplines, reflected in, e.g., (Wüster, 1955). With respect to computational applications, especially in machine translation, this awareness gave rise to the concept of 'sublanguages', which highlights the importance of circumscribing the domain of terminographic or lexicographic research (Kittredge and Lehrberger, 1982; Grishman and Kittredge, 1986).

In the following sections we will present the three perspectives on the semiotic triangle with respect to ontologies (Section 2.1), terminology (Section 2.2) and lexicography (Section 2.4), followed by subsections devoted to methods for populating ontologies, termbanks and dictionaries (Section 3). After that we will consider their applications (Section 4), emphasising the use of ontologies and lexical resources in the creation and translation of documents.

2. Mapping between words, concepts and reality

2.1. Ontologies

The term 'ontology' originates from classic philosophy, where it refers to theory for describing objects of the universe and their properties (*onto-* comes from the Greek word for *being*). Aristotle was one of the first philosophers to develop a system of universal categories for describing the world in terms of entities and

Unstructured	**c1390** (?c1350) *Joseph of Arimathie* (1871) l. 452, A-non tholomers men woxen ebiggore; sone beeren hem a-bac and brouhten hem to grounde.
Hierarchy	$\mathit{Example}\begin{bmatrix} \mathit{Created}: & \mathit{Date}\begin{bmatrix}\mathit{Year}: 1390 \\ \mathit{Range}: c\end{bmatrix} \\ \mathit{Created?}: & \mathit{Date}\begin{bmatrix}\mathit{Year}: 1350 \\ \mathit{Range}: c\end{bmatrix} \\ \mathit{Biblio}: & \mathit{PubStatement}\begin{bmatrix}\mathit{Source}: \text{Joseph of Arimathie} \\ \mathit{Publication}: \mathit{Date}\,[\mathit{Year}: 1871] \\ \mathit{Line}: 452\end{bmatrix} \\ \mathit{Eg}: & \mathit{Text}\begin{bmatrix}Q: \text{A-non tholomers men woxen} \ldots. \\ \mathit{Language}: \text{Middle English}\end{bmatrix}\end{bmatrix}$
Triples	*IS − A* (*Example*, Arimathie 1871-big) *IS − A* (*Date*, Year-1390c) *Year* (Year-1390c, 1390) *Range* (Year-1390c,‘circa-date’) *Created* (Arimathie1871-big, Year-1390c) . . .

Figure 2. Ontological analysis of an example from OED.

attributes predicated to them. The debate about the nature of the system of categories was one of the main subjects of philosophical debate throughout centuries, which contributed to knowledge representation techniques (Dreyfus, 1982; Sharoff, 2005b), also for information on the history of ontological debates in philosophy check http://www.ontology.co.

The need to input information into the computer and to perform reasoning on its basis gave ontological studies a more technological dimension, resulting in various approaches to knowledge representation. The bulk of communication between humans is not explicit enough for knowledge representation purposes. Figure 2 gives an example of the difference in the level of granularity needed for representing knowledge for the human and for the computer. The Oxford English Dictionary (OED) records one of the examples of uses of the word *big* from the 14^{th} century. The example quoted might be useful for the human reader, but for the computer this representation does not render the properties of the quote explicitly. Formally, the quote can be represented as a member of the class *Example*, which has such attributes as *Created, Biblio* or *Eg* filled with values, which, in turn, can be represented by the classes *Date, PubStatement* and *Text*.

This more explicit ontological specification can be extended even further, e.g., by representing the confidence in knowledge about the two provisional dates for the quote, using greater granularity of information about the language variety (by adding explicit information that Middle English is a subtype of English), as well as representing the actual range of years for each provisional date (in Figure 2 *c* means *circa*; the actual meaning of *circa* has to be defined formally).

Even though the hierarchical representation of concepts is intuitively appealing, it is not sufficient in many cases. For instance, a book can be represented as consisting of pages, paragraphs and chunks of argumentation, each of which can overlap, e.g., a paragraph can be shared between two pages, or the same argument can be mentioned in two separate paragraphs. A more general (but more verbose) way of representing concepts is by using triples containing a relationship and two objects. The same information from the OED example can be represented by triples, specifying the class of each object or concept (an object has to be identified by a unique label, e.g., ‘Arimathie1871-big’ for this quotation), and relations, which join the attributes of labelled concepts with their values. Relations can also connect concepts. Formally, the representation by triples is equivalent to a graph, i.e., a network of labelled nodes connected by directed arcs (relations). Terms like ‘semantic networks’ or ‘conceptual graphs’ are used to refer to representations of this sort (Sowa, 2000).

There are two relations universally used in ontology engineering:

IS-A defining the class-instance relationship (*Bucephalus is a horse*) or the subclass-class relationship (*horses are equestrians; equestrians are mammals*). In some implementations, a separate relation Instance-Of is used in the first case. The subclass-class subtype of this relation is often referred to as hyponymy (the class *horse* is a hyponym of the class *equestrian*, while *mammal* is a hypernym of *equestrian*). Some implementations of the IS-A relationship allow multiple inheritance, when an object can be an instance of several classes defined with different sets of properties, e.g., this book is an information container, which consists of separate chapters, while at the same time, it can be considered as a physical object, which has its size, weight, etc.

HAS-A defining the relationship of containment of one object in another one, e.g., a book can contain chapters (the converse relationship to HAS-A is PART-OF). This relation is often referred to as meronymy (*chapter* is a meronym of *book*).

Any developed ontology defines concepts and relations specific to its domain (e.g., *Created* or *PubStatement* in the example above). More general (Upper) ontologies are designed to provide basic concepts and relations covering a wide range of possible domains, e.g., Cyc (Lenat and Gupta, 1990), Penman Upper

Model (Bateman, 1990), or SUMO, Suggested Upper Merged Ontology (Niles and Pease, 2001). Cyc was a very ambitious project designed to represent the whole body of common-sense knowledge in a machine-tractable form, e.g., “Every tree is a plant” and “Plants die eventually”. The Penman Upper Model was designed as a tool supporting various projects in multilingual generation, so it included such concepts as IdeaReporting (*Many Americans expressed concerns that …*) or Purpose (*Press Ctrl-S to save your changes*) for connecting respective facts. SUMO, on the other hand, was designed with knowledge engineering projects in mind, having sub-ontologies for various domains, e.g., for finances, military devices or viruses.

Often an ontology specification mechanism contains tools for providing information about default values of some attributes (e.g., a horse has 205 bones), facts (an animal consumes food), inference rules (if an animal is deprived of food for a long time, it can die), and events updating the set of facts (e.g., Bucephalus died in 326 BC, thus it was not true that Bucephalus was alive in the model of 327 BC).

The formal language for representing ontologies is sometimes based on the idea of a hierarchy, a network or (more typically) a combination of the two. For instance, RDF (Resource Description Framework), a popular representation framework for Web resources defined by the W3C consortium (Beckett, 2004), uses triples to make statements like those presented in the triples part of Figure 2 (see also (Waltinger and Breuing, 2012) in this volume). RDF is an abstract model whose actual representation is based on *serialisation*, i.e., ways for encoding graphs as a sequence of statements in a file. For example, information about a publication, like the RDF specification (Beckett, 2004), can be encoded in an XML form using the ontology based on the Dublin Core (Hillmann, 2005):

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-
   syntax-ns#"
xmlns:dc="http://purl.org/dc/elements/1.1/">
<rdf:Description
rdf:about="http://www.w3.org/TR/rdf-syntax-grammar/">
<dc:title>RDF specification</dc:title>
<dc:publisher>W3C</dc:publisher>
</rdf:Description>
</rdf:RDF>
```

RDF is a generic framework which does not specify many parameters needed for elaborate ontologies, but it led to development of Web Ontology Language (OWL, note the order of characters in the acronym), specifically designed for representing ontologies (Smith et al., 2004). Objects and relations from Figure 2 can be defined in OWL in the following way:

```
<Date rdf:ID="Year-1390c">
<yearValue  rdf:datatype="&xsd;positiveInteger">1390</year-
Value>
<rangeValue rdf:resource="circa-date"/>
</Date>
<Example rdf:ID="Arimathie1871-big"
<Created rdf:resource="#Year-1390c"/>
<Biblio ... />
</Example>
```

In this notation the date of citation is defined as an identifiable resource consisting of a positive integer value with a range approximation, while the example makes an explicit reference to this resource.

2.2. Terminology

A termbank provides a link between concepts in an ontology and their names (*terms*), typically in several different languages. Thinking within ISO/TC 37 “Terminology and other language and content resources” on the nature of this link has followed the principles of the Vienna school of terminology founded by Eugen Wüster (Wüster, 1979). According to these principles, ideally concepts are defined by being assigned a place in the concept system before they are designated (the onomasiological perspective). Moreover, each concept should be designated by only one term (mononomy) and each term should refer to only one concept (monosemy) – the principle of *univocity*. This ideal case based on a one-to-one mapping is illustrated in the top part of Figure 3.

However, it is quite common that one designation can refer to several different objects even in the same domain, e.g., *terminology* taken as a term of its own can mean a collection of terms (= termbank), a set of practices for creating termbanks (= terminography, akin to lexicography), and also to a theory of the relationship between concepts and terms (akin to lexicology). Alternatively, several terms can be used to refer to the same object in slightly different contexts, e.g., *catalog, directory* and *folder* can all refer to a group of files in a file system.

Temmerman argues very persuasively that polysemy and synonymy of this sort are not only pervasive but also functional, since they enable specialised communities of practice to negotiate and refine their understanding of their shared knowledge space over time (Temmermann, 2000). This supposes a model of knowledge as a multidimensional space in which “the value of a concept with respect to a given axis is generally defined as a range and only exceptionally […] as a point” (Sager, 1990, pp. 15–16). Temmerman proposes that, since many concepts are not clear-cut, they are better considered as categories showing different degrees of prototypicality and possibly overlapping with others.

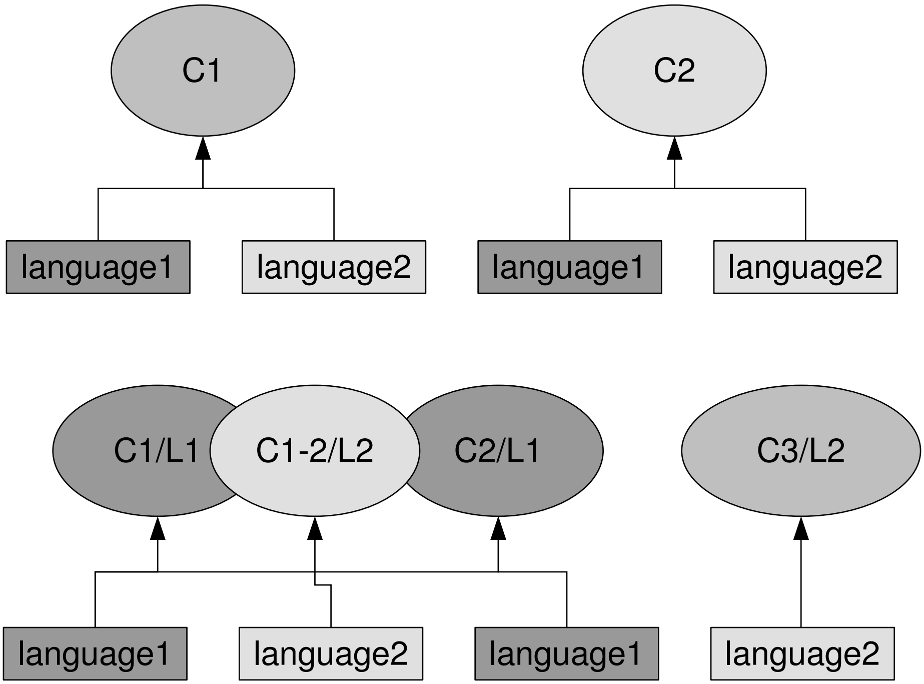


Figure 3. Mapping terms to concepts.

In the multilingual context it is possible that one language (L2) has a word that refers to a object (C3) that is not lexicalised in another language (L1), e.g., *deuxroues* in French refers to the set of cycles, scooters and motorbikes in English (the bottom part of Figure 3). It is also possible that a word in L1 can refer to two separate objects, e.g., *Wortverbindung* in the quote from Frege (3) refers to single-word compounds as well as multiword expressions, while its English translation (4) only to *combination of words*. Similarly *Gegebensein* in (3) is a very precise description to a philosophical concept in German, whereas the word *presentation* in (4) does not function as a philosophical term in English. It is used as a term in special reference in religion and obstetrics, but in (4) only in general reference.

In some cases a term in one language functions as a lexicalised hypernym for a subset of concepts but does not have an equivalent in another language at the same point in the hierarchy. For instance, the French terms *bicyclette, cyclomoteur, véhicule* all have English equivalents, *bicycle, moped* and *vehicle*, respectively. However, while *deuxroues* covers all types of two-wheeled vehicles (bicycles, mopeds, scooters, etc) the English term *two-wheeler* is generally taken as referring to bicycles rather than to *motorised/powered two-wheelers* also.

A considerable part of the activity of a terminologist consists of organising the system of concepts, so in this respect it is similar to the work of an ontologist, except that it also requires assigning names to the concepts in a variety of languages (Sager, 1990; Wright and Budin, 2001). This organisation can often involve standardisation of the delimitation of particular concepts and their designation; this standardisation may be imposed locally, as part of the house style of a particular company, or nationally, regionally or internationally by standardisation bodies such as ANSI, CEN and ISO (see also (Sasaki and Lommel, 2012) and (Trippel, 2012) both in this volume). One common facet of standardisation is the planned creation of neologisms, most often in response to massive borrowing from dominant languages, such as English; L'Office québécois de la langue française has undertaken probably the most concerted effort ever in this field. Such (legitimate) prescriptive activities distinguish terminography from contemporary lexicography, which aims to be descriptive of usage, although historically the dictionaries of language academies also set out to fix use.

However, in addition to terms themselves and their links to concepts, a termbank can contain:

- typical collocations, e.g., *open, save, delete a file* vs. *ouvrir, sauvegarder, supprimer un fichier*, to facilitate their coherent use and translation;
- boilerplate, i.e., fixed expressions, which are expected in a specific domain or specific document type, e.g., the copyright statement attached to Wikipedia articles *All text is available under the terms of the GNU Free Documentation License*;
- common abbreviations of multiword terms, e.g., *CALL* from *computer-assisted language learning*, *ABS* from *antilock brake system*;
- proper names, especially if the tradition of their spelling differs across languages, e.g., *München, Munich* or *Monaco* in German, English and Italian respectively;
- deprecated and preferred terms, e.g., *flame resistant* and *flame retardant*, respectively, in the "ISO Glossary of Fire Terms and Definitions".

2.3. Formats for storing terminology

Individuals and organisations commonly make use of a wide range of more or less sophisticated formats for storing and presenting terminology, from termbanks with rich and highly-structured data to simple flat bilingual lists of equivalent expressions managed in a spreadsheet. If we consider the sophisticated end of the spectrum, instantiated by the EU inter-institutional terminology database IATE[1] and other termbanks accessible via the International Telecommunications Union website[2], we can expect to see represented most if not all the following items:

- links to concepts in an ontology, primarily by means of intensional definitions specifying the characteristics of the concept and the nearest genus, additionally with notes on the scope of usage;
- links to hypernyms, hyponyms and co-hyponyms;
- domain(s) in which a term is used;
- terms themselves and their variant names, abbreviations;
- indicators that the term is standardised, or preferred or deprecated;
- reliability ratings;
- morphological information (gender, number, declension);
- examples of their use;
- links to other terms related by their form, e.g., *poverty, income poverty, Interagency Working Group on Poverty Elimination*;
- housekeeping information.

In the case of multilingual termbanks some information is shared between all terms referring to the same concept, while most information is language-specific.

In designing a termbank it is important to distinguish between the onomasiological and semasiological perspectives (Gibbon et al., 1997). With the first, concept-oriented perspective a given entry, with equivalent terms in multiple languages, corresponds to a single, language-independent concept; the second is able to represent polysemy. The two perspectives are complementary and can be accommodated in a single representation format, but the design of the termbase will differ depending on the preference for either perspective.

The late 1990's and early 2000's saw a flurry of activity, much of it associated with ISO/TC 37, aimed at designing standards for exchanging structured terminological data for the benefits of both human translation and the gamut of machine applications ranging from term extraction and ontology construction to IR and MT. They can be traced back to the Text Encoding Initiative (TEI) (Sperberg-McQueen and Burnard, 2002) and include the SGML-based MARTIF (Machine-Readable Terminology Interchange Format) and OLIF (Open Lexicon Interchange Format) developed for six European languages by a consortium of major translation and publishing companies. Currently, the leading contender is TBX (Term Base eXchange)[3], an XML application whose core structure is based on MARTIF. It is concept-oriented. TBX is sponsored by LISA (Localization Industry Standards Association) and published as ISO 30042.

2.4. Lexicography

Lexicography is one of the oldest human activities since the invention of writing. The need to catalogue a language, analyse obscure sources or communicate with other cultures led to production of word lists accompanied with their definitions, commentaries or translations. One of the earliest dictionaries are available in the form of Sumerian cuneiform tablets listing words of a foreign language (Akkadian) or a glossary of herbs for medicinal use (Hartmann and James, 2001).

In English language lexicography, the rise in the importance of dictionaries coincides with the spread of literacy and the appearance of a large number of printed sources in the seventeenth century: newly educated readers needed an authoritative source to guide them in understanding rare words and resolving their disputes about word meanings. Dictionaries, such as Johnson's 1755 *Dictionary of the English Language*, appeared to satisfy this demand (Kilgarriff, 1997).

A more recent rise in dictionary making in the twentieth century coincides with economic globalisation, which has resulted in the need to use a variety of bilingual dictionaries for the tasks of translation and international communication. Moreover, globalisation has increased the importance of dictionaries for language learners, primarily learners of English, as English had established itself as the world's global language by the second half of the twentieth century.

Traditionally dictionaries were developed using a large collection of slip cards for individual words containing citations, which illustrated their use. At a later stage, these groups of examples were used to define senses. Often the decisions were made on the basis of introspection. For instance, a native English speaker might suggest that the noun *landing* can mean either an aircraft coming to earth or an intermediate platform in a staircase, and can invent examples of their use, e.g., *it was a bumpy landing*.

Development of corpora in the last quarter of the twentieth century has changed the situation considerably: lexicographic evidence in modern dictionaries is mostly based on analysis of concordance lines taken from representative corpora (McEnery and Wilson, 1999). The first dictionary created from scratch on the basis of corpus evidence was COBUILD, developed in a joint project between Collins and Birmingham University (Sinclair, 1987).

The frequency of words in a language follows a 'long-tail' distribution, also known as Zipf's law (Zipf, 1935): very few words are extremely frequent (*the, of, have*), while the frequency of a very large number of words is fairly modest. For instance, in one million words of the Brown Corpus (Kučera and Francis, 1967) there are 23 examples of *landing* and only two occurrences of the verb *to stab*, none of which refers to the metaphorical sense (*to stab somebody in the back*, 'to betray someone'), while in 100 million words of the BNC (Aston and Burnard, 1998) there are 1,049 occurrences of *stab*, with at least 31 examples of its metaphoric use.

This suggests that corpora used for producing dictionaries for general rather than special reference purposes have to be big enough to cover the range of expressions possible with each individual word (the BNC, Bank of English or Oxford English Corpus used for this task measure in hundreds of millions of

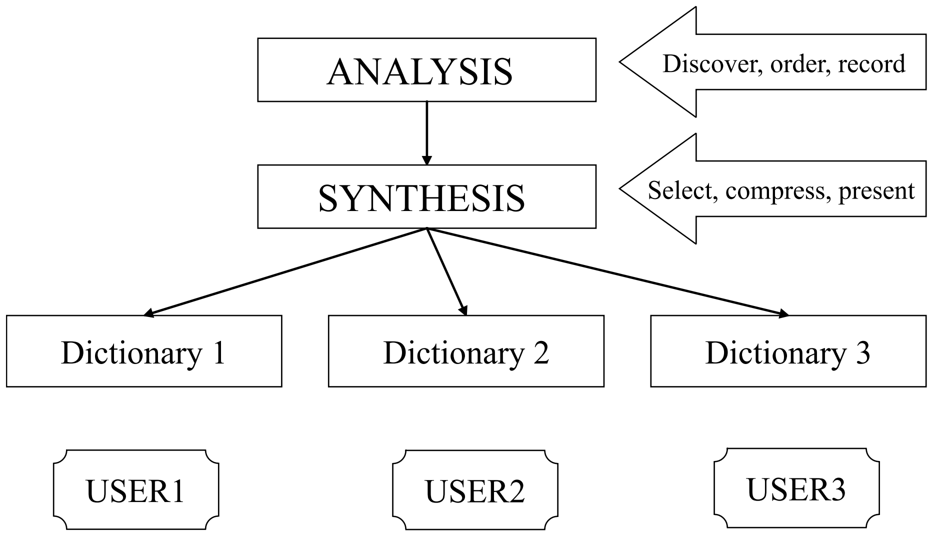


Figure 4. Two stages of dictionary development.

words). However, going through thousands of examples of *landing* or *stab* in a corpus like the BNC is a very time consuming process, which can be facilitated by tools designed for extracting and condensing lexicographic information, such as SketchEngine (Kilgarriff, 2003). Such tools can identify words more commonly modifying the target, e.g., *forced, Normandy, first-floor, emergency, downwind, safe* as modifiers for *landing*, and *gear, strip, craft, stage* as words modified by it. These two lists can suggests adding other senses which might have been missed in introspection, e.g., *landing* in the sense of a military operation.

It is quite common that dictionaries are developed through two stages: through analysis of language data and then its synthesis to present facts to the users (Figure 4). In the first stage lexicographers analyse the behaviour of a word, create sense distinctions, determine its contexts of use and select a range of examples for each context. This results in a systematic overview of facts about the word, which are stored in a database. In the second stage this information is used for producing a number of actual dictionaries. Each dictionary in this case can contain a subset of words recorded in the database. Some senses can be discarded (as less frequent or less relevant to the anticipated users) or collapsed, joining several senses together. Alternatively extra senses can be distinguishes depending on the needs of the users. Definitions can be adapted to the level expected by the user, and translations into a specific language can be defined.

The issue of presentation is important in termbanks, but for dictionaries the way information is delivered to their users is crucial. The same word in the same language can be described in different ways depending on the purpose of a dictionary. A dictionary can be aimed at being a guide to understanding a word in as many contexts of its use as possible. Alternatively, a dictionary can be used as a guide for active production in a foreign language, which implies offering greater guidance to the user about more typical constructions of its use (Rundell, 1999). A dictionary aimed as being a guide for translation can have more senses than reasonable for the same word in a monolingual dictionary, since the target language may make more sense distinctions, so justifying several translations from the source language.

In Figure 5 we list examples of entries for the same word from four dictionaries: LDOCE (LDOCE, 1995), a dictionary for non-native speakers, OED (OED, 1989), a comprehensive dictionary recording historical information about word for researchers and native speakers (only two senses out of 19 are shown), WordNet (Miller, 1990), a concept-oriented thesaurus, and Oxford-Hachette French Dictionary (OFD, 1994), which is designed as a guide for translators. The bilingual dictionary makes distinctions between two senses of landing from a boat (for people and for cargo), which are not made in the two monolingual dictionaries. The definitions used in LDOCE are considerably simpler and it contains more information to guide the users.

A dictionary entry can contain various types of information:

– the headword and its variations (*favour, favor*; *ping pong, table tennis*);
– hyphenation and pronunciation (using different notations to indicate them);
– inflection (especially when it is irregular, e.g., *take, took, taken*);
– sense distinctions (they can be presented as a flat list or as a structured list with several levels of subsenses, like in the OED example; in other dictionaries distinctions between subsenses can be made implicitly, e.g., *crash landing* in LDOCE);
– basic morphological categories (this can affect the structure of the entry, as some dictionaries list all morphological classes as senses in the same entry, while others treat them as different headwords; some grammatical information can be confined to a specific sense, e.g., [C] in the LDOCE example indicates that *landing* used in Senses 2 and 3 is a countable noun);
– definitions or translations of senses (a bilingual dictionary can also indicate the nature of senses by conditions on their use, e.g., *at turn of stairs, by plane*, as well as properties of translations, e.g., their grammatical gender or translation of collocation patterns *landing on*→*atterissage sur*);
– cross-references to other entries (references to synonyms or antonyms, as well as generic ‘see’ links);
– more or less fixed constructions (in some dictionaries they are presented as separate headwords, but most of the time they are incorporated into respective entries, e.g., Sense 8 in the full entry for *landing* in the OED lists 33 compounds);

- examples (they can also contain full bibliographic information, like in the OED example; they are normally associated with senses, but can also act as indications of subsenses, e.g., *Apollo moon landings* in the LDOCE example)
- various constraints on usage, such as the domain, register, region or period of use, e.g., *Aeronaut., slang, British, obsolete*;

The concept of a dictionary is based on a list of headwords, each of which is taken as a separate entity. On the other hand, the goal of a dictionary is to help its users use words or understand them in context. This often leads to introduction of information about word uses into its entry. One way of doing this is based on definitions. For instance, one sense of *leave* in COBUILD is defined as *If you leave someone to do something, you go away from them, so that they do it on their own*, which intends to show the context of its use (e.g., you=a human actor) and possible intentions of the speaker. However, this information is not encoded in any explicit way. One possibility of formalising the context of use is by using systemic networks, which can link meaning-making intentions to lexical realisation options (Sharoff, 2005a). Some researchers have proposed a very elaborate mechanism for representing lexical co-occurrence links between words. For instance, Mel'čuk proposed the notion of lexical functions as generalisations of lexical co-occurrences and designed their typology (Mel'čuk, 1996). Thus, expressions like *heavy rain, strong desire, strict control, sleep firmly* imply the high degree of a quality, so such collocations can be captured by a lexical function *Magn*, such that *Magn*(*smoker*) = *heavy*; *Magn*(*Raucher*) = *stark*; *Magn*(*fumeur*) = *grand*.

3. Populating ontologies, termbanks and dictionaries

3.1. Corpora

As we stated earlier, ontologists deal with concepts, which are manifested in texts. The link between texts and the work of terminologists and lexicographers is even more direct. Hence, they need corpora, large collections of texts representative of a chosen domain and/or genre. Raw texts in corpora are usually accompanied with some kind of annotation, such as text metadata (provenance, text type, intended audiences, etc), linguistic annotation (parts of speech, lemmas, syntactic structures, etc) or alignment with corresponding translated texts (McEnery and Wilson, 1999).

The traditional way of creating corpora (still in use in some situations, especially for lexicography and linguistic research) is based on collecting texts manually on a case-by-case basis. A classic example is the development of the Bank of English (Sinclair, 1987) in the 1980s from scanned and OCR'ed printed texts. The growth of the amount of texts available in electronic form, often over the Internet, has led to a proliferation of corpora in a variety of domains and languages, often resulting in the possibility of using 'disposable corpora' (Varantola, 2003), i.e., corpora created for a particular task and discarded once the task has been fulfilled.

There are several methods for creating disposable corpora from the web. One is based on "focused crawling", which, in its simplest form, involves selecting several websites containing a large number of texts in the target domain and retrieving the entire set of these texts. Given that Wikipedia dumps in a variety of languages are available for download, a version of focused crawling can be based on retrieval of Wikipedias for several language and clustering them into subdomains on the basis of explicit metadata available in Wikipedia (categories, templates and links between pages). One problem with this simple approach to crawling is that only a small subset of relevant data can be retrieved, which might bias the corpus and the results of the ontological, terminological or lexicographic study.

More advanced methods of focused crawling involve starting with a seed set of links and then collecting links to other relevant websites, with the relevance assessed by keywords and/or hypertext links between pages, as more relevant pages tend to have more inter-connections with each other (Chakrabarti et al., 1999).

Yet another method for corpus collection relies on making automated queries to a major search engine, using words defining a domain. Since the initial set of seed queries is likely to be incomplete, corpus collection can include a bootstrapping phase: an initial corpus is created from seed queries, terms are automatically extracted from it and then fed to the search engine to collect a bigger corpus (Baroni and Bernardini, 2004). For instance, the seed query set can include words like *dissociative, epilepsy, pseudo-seizures*. This leads to a corpus yielding new keywords like *amnesia, convulsions, paroxysmal*, which can be used in the second iteration. It is also possible to collect a large (100–200 million words) corpus of general language using this method, if the queries consist of words from the general lexicon (Sharoff, 2006). Finally, keyword-based corpus development can be combined with focused crawling to get a wide variety of seed websites (Stamatakis et al., 2003).

For multilingual applications, parallel corpora containing original texts and their translations can provide useful data for developing bilingual dictionaries and termbanks. Traditionally, again, parallel corpora were created by painstaking collection and alignment of published translations. Nowadays large amount of texts are available on the Web, either in institutional repositories, e.g., United Nations (Eisele and Chen, 2010), European Parliament (Koehn and Knight, 2002), or in multilingual webpages, e.g., newswire stories and their translations in different languages. Resnik and his colleagues developed a method for harvesting webpages that are structurally similar, cross-linked and written in different languages (Resnik and Smith, 2003), while lexical-based methods can be used to detect the degree of similarity between potentially parallel pages (Patry and Langlais, 2011).

Once a corpus has been collected, it has to be ‘cleaned’ in order to become suitable for further processing (Baroni et al., 2008). Usually, the cleaning steps include:

– removal of duplicates and near-duplicates, e.g., normal and ‘print-friendly’ versions of the same page;
– unification of encodings used on webpages, e.g., KOI-8, CP1251 and UTF-8 are frequently used for Russian;
– identification of boilerplates, navigation frames and foreign language fragments (ovals in Figure 6 indicate areas to be cleaned, otherwise words like *Wikipedia* or *Main page* can become disproportionately frequent);
– identification of paragraph breaks (and, possibly, of sentence boundaries).

The next step in corpus processing is tokenisation, i.e., identification of breaks between word-like units. In languages without orthographically marked word boundaries, e.g., Chinese, the procedure is complicated enough to warrant a special competition on automatic text segmentation (Ng and Kwong, 2006). In European languages the procedure is relatively straightforward, but even in this case there can be problems with making tokenisation decisions, e.g., on the apostrophes in *the Zero-X’s course* vs. *Cote d’Azur* in English. A separate problem concerns tokenisation of compounds, which lack explicit word boundaries, e.g., *Agrarstukturverbesserungsmaßnahmen* in German (‘measures to improve the agrarian structure’).

Finally, texts are usually processed by part-of-speech (POS) taggers and lemmatisers. POS tagging helps in identification of basic text chunks, such as noun phrases, while lemmatisation helps in reducing data sparsity. The use of word forms or simple stemming (removal of common endings) is feasible in English, but a word in languages with agglutinative or inflective morphology (e.g., Slavonic or Turkic) can have a large number of different forms. Lack of lemmatisation in such cases considerably reduces the coverage (since many morphological variations in the corpus correspond to one and the same lemma). Some lemmatisers can perform disambiguation, e.g., deciding the lemma of *left* in *on the left* and *he has left*, while others list all possible options for each form.

Finally, for open-ended retrieval tasks corpora have to be encoded to keep and retrieve information concerning metatextual descriptors of documents (at least, their provenance, possibly also their domain and language), results of tokenisation, POS tagging and lemmatisation. Encoding of modern corpora is generally based on XML using two approaches to annotation. One entails the use of tags to “enrich” the text, so that each document has its metatextual tags (e.g., describing the author, title, date of its creation), while each word has information on its POS tag, lemma and other properties. This makes it possible to use standard XML indexing tools (e.g., Berkeley DB XML) or tools designed with corpus queries in mind (e.g., IMS Corpus Workbench or XAIRA). Another approach is to use a "stand-off" annotation, e.g., (Ide et al., 2000), in which each token has an identifier, while its properties are described in separate XML files referring to the identifier. The advantage of the stand-off scheme is that new layers of annotation can be added or removed without disturbing the original corpus. Its disadvantage is related to the assumption that a corpus is a fixed collection of texts with eternal identifiers assigned to each word. However, many modern corpora are frequently updated, e.g., by downloading new versions of the Wikipedia or crawling updated websites, thus putting strains on the need to maintain the link between identifiers in the corpus and its stand-off annotations.

3.2. Term extraction

Large domain-specific corpora offer tremendous advantages for creating and maintaining termbanks that reflect the current state of the art in their domain. However, getting terms out of corpora is not a reliable procedure, which is complicated by the statistical nature of term extraction and differences between languages. The terminology used by researchers in this field is also confusing, as different researchers use different names to refer to slightly different aspects of the procedure, e.g., term extraction, term recognition, keyword identification, keyword recognition, glossary extraction. There are also important differences between methods for extracting single-word and multiword terms. Some term extraction methods can extract terms of both types, while others are restricted to single-word terms only.

The earliest method of keyword identification is the TF*IDF method (Term Frequency × Inverse Document Frequency) first proposed by Karen Spärk Jones in the beginning of 1970s (Spärck Jones, 1972) and modified in many different ways since. The definition used below assumes a collection of $|D|$ documents, $f_{i,\mathrm{j}}$ is the number of occurrences of term t_i in document dj, and $|D_i|$ is the number of documents term t_i appears in. Then TF*IDF is computed for each term in each document (Manning et al., 2008):

$$TF * IDF_{i,j} = TF_{i,j} \times \mathrm{IDF_i}; TF_{i,j} = \frac{f_{i,j}}{\Sigma_k \mathrm{f}_{k,j}}; IDF_i = \log \frac{|D|}{|D_i|}$$

Words with the highest TF*IDF score across the whole collection are treated as terms (the cut-off threshold for selecting the terms depends on the document collection and application, and is usually selected in a trial-and-error manner). The method is fast and does not require any additional data apart from a document collection, thus making it language-independent. As this method can miss terms common to the entire collection (if a term occurs in every document, its IDF will be zero), more successful term extraction methods compare the frequency of words in a document collection against a reference corpus by measuring the importance of seeing t_i in D rather than in the reference corpus. Various methods for computing the statistical significance of this fact are mutual information (MI), χ^2, log-likelihood (LL), etc (Kageura and Umino, 1996), which are all based on the following table for each word:

Table 1. Date used for computing "termhood".

	D collection	Reference corpus	Total
Frequency of term	a	b	a+b
Frequency of other terms	c-a	d-b	c+d-a-b
Corpus size	c	d	c+d

Then the expected values E1 and E2 and the log-likelihood value G2 are calculated by taking into account the relative frequency of terms in the two corpora as well as the absolute number of their occurrences as evidence of its statistical significance (Rayson and Garside, 2000):

$$G2 = 2(a\ln(\frac{a}{E1}) + b\ln(b\frac{b}{E2})); E1 = c\frac{a+b}{c+d}; E2 = d\frac{a+b}{c+d}$$

These methods are successful in extracting single words, but they miss multi-word terms, which take a very large proportion of the termset in any domain. Methods like C-value (Frantzi and Ananiadou, 1999), Glossex (Kozakov et al., 2004), Termex (Sclano and Velardi, 2007) can also identify multi-word units by selecting the most frequent n-grams (sequences of n words) and filtering their list, either using linguistic patterns (e.g., ADJ+NOUN, NOUN+of+NOUN, NOUN+NOUN in English) or the heuristics of containment, i.e., if the frequency of a sequence of words is comparable to the frequency of a longer sequence that contains it, the former n-gram is discarded. For instance, even though *graphical user* is frequent in computer science texts and it follows the pattern of ADJ+NOUN, it is not selected as a potential term, since it is nearly always embedded in a longer expression *graphical user interface*.

Termex utilises a slightly more complex strategy by considering for each term its domain relevance (as the ratio of its frequency within the domain in comparison to a reference corpus), domain consensus (the term is distributed over a large number of documents in the domain), lexical cohesion (similar to the containment filter) and structural relevance (terms that are used in the title or keywords or are italicised/underlined in a web document are more likely to be terms).

Table 2. Qualitative evaluation of term extraction methods, from (Zhang et al., 2008).

Judge	TF-IDF	Weirdness	C-value	Glossex	Termex	Voted
1	0.67	0.80	0.59	0.81	0.93	0.97
2	0.79	0.85	0.69	0.83	0.95	0.97
3	0.77	0.77	0.68	0.83	0.95	0.97

The accuracy of an algorithm extracting a list of terms (T_e) from a domain-specific corpus can be measured by precision (the proportion of terms in the list that are relevant terms in the domain), recall (the proportion of relevant terms in the list in comparison all relevant terms contained in this corpus, T_r) and F-measure, which is a harmonic mean between precision and recall (see also (Paaß, 2012) and (Menke, 2012) in this volume):

$$P = \frac{|T_r| \cap |T_e|}{|T_e|}\ ; R = \frac{|T_r| \cap |T_e|}{|T_r|}$$

$$F1 = \frac{2 \cdot (P \cdot R)}{(P + R)}$$

Usually, there is a trade-off between recall and precision: greater precision (less noise in the list of extracted terms) means fewer terms overall, thus negatively affecting recall, and vice versa. Also computing recall is more difficult as this requires manual extraction of all terms form a potentially large corpus, whereas it is easier to judge precision by assessing the top N terms in the list returned by a term extraction algorithm. A recent study of term extraction methods in (Zhang et al., 2008) shows that precision for the top 100 terms extracted from a corpus of 1.3 million words from the English Wikipedia reaches 97 %, see Table 1. The voting algorithm listed in the last column of Table 1 is based on the weighted voting strategy: the rank of term *t* is based on the ranks $R(t_i)$ of this term produced in all other algorithms weighted by their precision.

3.3. Development of ontologies

The traditional way of developing ontologies is by organising the body of knowledge in a domain into a network of concepts, manual selection of relevant facts and representing the concepts and facts using the ontology language. The Cyc project is a prime example of this method. Upper ontologies are also normally created in this way. However, population of ontologies with facts is much easier to automatise. The CIA Worldfact book, which contains information about countries, their capitals, borders and economic indicators in a somewhat structured form, is relatively easy to convert into a formal ontology. Wikipedia ‘infoboxes’ can be also used for this purpose. Freebase is a community-driven project for populating ontologies with facts (\url {http://www.freebase.com/}).

A diverse corpus with less-structured facts can be also used for semi-automatic population of large ontologies. First of all, a considerable part of the effort involved in constructing ontologies is devoted to extracting the list of names for concepts. The names themselves can be detected by using one of the term extraction mechanisms described above. Some links between names and concepts can be extracted by detecting patterns and applying machine learning techniques. For instance, the contexts in which the names of countries are used share many words, e.g., *to visit/to return to/GDP of Germany/Italy/Spain/Switzerland*; they also occur in lists of two-three names (*in France, Italy and Spain; between Germany and Switzerland*). This information can be used for creating a cluster of similar words, such as names of countries and big cities (Lin, 1998; Sharoff et al., 2006). Once the lists of countries and cities are known, the set of facts about which city is the capital of which country can be retrieved from the same corpus using patterns like *Tirana, capital of Albania; the Hungarian capital of Budapest*, etc. Similar patterns can be used to retrieve IS-A or HAS-A relations from free text definitions like (1) or (2) above. Such methods invariably produce noise, e.g., an example like *Berlin took over from Bonn as the capital of Germany* can produce a false assertion that Bonn *is* the capital of Germany.

More advanced methods for populating ontologies involve also automatic detection of patterns like *is a* or *the capital of* from the seed set of concepts and terms. For instance, if we start with seed facts that *apples, peaches, pears* are kinds of fruit, we can discover automatically frequent ways for expressing the relation between a class Y and its members X in English, including *X, such as Y; Y, including X; X, RB called Y* (*RB* in the last case means that an adverb can be frequently used before the word *called*, e.g., *often, sometimes, popularly, officially*). Then, we can apply the newly learned patterns for the IS-A relation to retrieve more facts. Pantel and *et al*, who presented this method, also evaluated the precision of fact detection using automatically extracted patterns and found it to be in the range of 50–70 % (Pantel et al., 2004). For more information on techniques in ontology extraction, see an overview in (Staab and Studer, 2004).

4. Applications in document authoring and translation

In considering multilingual applications of these activities, it is useful to bear in mind the three major functions of translation, whether by human or machine: dissemination (of outgoing information), assimilation (of incoming information) and communication (dialogue within a group).

Dissemination starts with the authoring of the source text. To be effective and safe, technical documentation must be unambiguous and easy to understand, using words and terms believed to be accessible to the reader (and to the translator). A significant factor in promoting comprehension is consistency – if the same part of a machine is variously referred to as the *cover* and the *flap*, the user is likely to be confused. The provision of consistent and transparent terminology is an asset which offer significant advantages. In larger organisations, efforts to this end may go as far as establishing 'structured' or 'controlled' authoring.

A Controlled Language (CL) is a version of a human language that embodies explicit restrictions on vocabulary, grammar and style for the purpose of authoring technical documentation (Kittredge, 2003; Nyberg et al., 2003). With roots in the Simplified English of the 1930s, the initial objective was to minimize ambiguity and maximize clarity for human readers, including non-native speakers of English, and so to avoid the need for translation altogether. Probably the best-known CL is AECMA Simplified English[4], which is a *de facto* standard in the aerospace industry; the concept has also been widely adopted in the automotive and IT sectors. Even within the same sector CLs vary from one company to another while respecting the same general principles.

Accordingly, at the lexical level (which constitutes the major part of a CL specification), a CL will define the approved technical terms and often explicitly list any deprecated terms which authors tend to use in error. Moreover, the prescriptions extend to non-technical expressions. For example, AECMA restricts the use of *about* to 'concerned with', specifying that *approximately* should be used for the other frequent sense; *support* can only be used as a count noun (*Put a support under the item* but not *Offer support*), and when a verb is required to express this idea it must be *hold* (*Hold the item* but not *Support the item*). Thus, the principle of univocity applies even beyond terminology to what would usually be considered 'vocabulary'.

Tools exist to automate compliance checking at all levels of a given CL[5] and even to propose corrections for recurrent error patterns, some of which will involve misuse of terms. The rules and lexicons of these tools are often user-definable.

Clarity and consistency at source is a major factor in reducing translation costs, and the benefits multiply in direct proportion to the number of target languages. In the dissemination scenario, it is imperative that the terminology be shared across all the applications that contribute to the creation and translation processes – mono- and multi-lingual online glossaries, authoring tools, translation memory tools, machine translation engines – in order to maximise lexical coverage. Hence the importance of exchange formats like TBX, designed for interoperability.

Increasingly, the component parts of technical documents are managed in Content Management Systems (CMS) (see also (Max, 2012) and (Rahtz, 2012) both in this volume). Modularization on a large scale entails significant design and implementation costs, but DITA (Darwin Information Typing Architecture)[6] provides an increasingly widely adopted infrastructure. Designed for developing technical product documentation, it specifies three basic topic types: *concept* (for background information), *reference* and *task*. A task topic, for example, is intended for instructional procedures and is itself modularized into sub-elements containing content for pre-requisites (e.g., preparation of ingredients before cooking begins), steps, options, results and post-requisites (e.g., resetting or cleaning equipment after a process), among others, each instance potentially reusable in many places. This is analogous to ontology building, except that what is represented is not objects but states (e.g., *The stack is empty.*) or events (e.g., *The dialog box appears.*). This has many implications for the work of ontologists, terminologists and 'information architects'.

The assimilation of documents from unrestricted third-party sources is well known to be a much more complex task and the inconsistencies and constant evolution of individual and group usage will continue to trouble the core task of reliably extracting meanings from the words of texts which we identified at the beginning of this chapter.

Much more attention needs to be paid to the shifting nature of terminology within subject communities. The corpora used for constructing ontologies and extracting terminology for both statistical and rule-based MT systems must be constituted to reflect not only a common subject domain and genre but also the strength of networks of communication.

LDOCE: *landing* **land·ing** /'lændin/

1. [C and U] the action of bringing an aircraft down to the ground after being in the air [≠ take-off]: *take-off and landing procedures*; crash/emergency landing (=a sudden landing caused by a problem with the engine etc); *the Apollo moon landings*→**soft landing**
2. [C] the floor at the top of a set of stairs or between two sets of stairs: *the first-floor landing*
3. [C] the action of bringing soldiers onto land that is controlled by the enemy: *the first landings of American Marines at Da Nang*

OED: **I.** The action of the verb **LAND**.

1.a. The action of coming to land or putting ashore; disembarkation.
c1440 *Promp. Parv.* 312/1 Londynge fro schyppe and watur, applicacio. **1577-87** *HOLINSHED Chron.* I. 9/2 They take landing within the dominion of king Goffarus. . . .

1.d. The (or an) action of approaching and alighting on the ground or some other surface after a flight. **happy landings!**: see HAPPY a. 3.
1784 [see LAND v. 8b]. **1909** *Flight* 13 Feb. 93/1 (*heading*) Flight 'landings'. . . .

8. *attrib.* and *Comb.*, as (sense 1) *landing area, fee, . . . tower, vehicle;* (sense 3) *landing-gaff, -hook, -ring*; **landing beam** *Aeronaut.*, a radio beam to guide aircraft when landing; **landing card** . . .

WordNet: *landing* Noun

1. landing (an intermediate platform in a staircase)
2. landing, landing place (structure providing a place where boats can land people or goods)
   - Direct hyponym: dock, dockage, docking facility
   - Part meronym: landing stage
   - Direct hypernym: structure, construction
   - Part holonym: seaport, haven, harbor, harbour
3. landing (the act of coming down to the earth (or other surface)) *the plane made a smooth landing*; *his landing on his feet was catlike . . .*

Pocket Oxford-Hachette French Dictionary: *landing* noun

1. (*at turn of stairs*) palier m; (*storey*) étage m;
2. (*from boat*) (*of people*) débarquement m; (*of cargo*) déchargement m;
3. (*by plane*) atterrissage m (**on** sur).

Figure 5. Examples of entries for *landing* from four dictionaries.

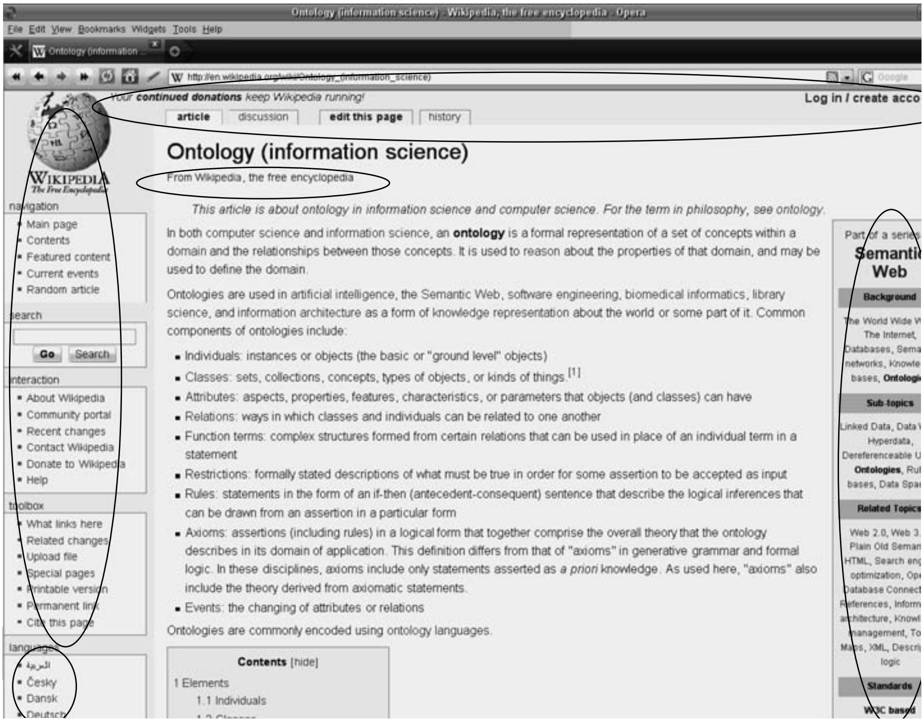


Figure 6. Text cleaning example from Wikipedia.

References

Aston, G. and Burnard, L.

1998 *The BNC Handbook: Exploring the British National Corpus with SARA*. Edinburgh University Press, Edinburgh.

Baroni, M. and Bernardini, S.

2004 Bootcat: Bootstrapping corpora and terms from the web. In *Proc. of LREC2004*, Lisbon.

Baroni, M., Chantree, F., Kilgarriff, A. and Sharoff, S.

2008 Cleaneval: a competition for cleaning web pages. In *Proc. of the Sixth Language Resources and Evaluation Conference, LREC 2008*, Marrakech.

Bateman, J. A.

1990 Upper modeling: Organizing knowledge for natural language processing. In *Proc. of the Fifth International Workshop on Natural Language Generation*, pages 54–61, Dawson, PA.

Beckett, D. (ed.)

2004 *RDF/XML Syntax Specification (Revised)*. W3C Recommendation. http://www.w3.org/TR/rdf-syntax-grammar/.

Chakrabarti, S., van den Berg, M., and Dom, B.
1999 Focused crawling: a new approach to topic-specific web resource discovery. In *Proc. 8th International World Wide Web Conference*, Toronto.
Dreyfus, H. L. (ed.)
1982 *Husserl, Intentionality, and Cognitive Science*. MIT Press, Cambridge, MA.
Eisele, A. and Chen, Y.
2010 MultiUN: A multilingual corpus from United Nations documents. In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*, Valletta, Malta.
Frantzi, K. T. and Ananiadou, S.
1999 The C-value / NC-value domain independent method for multi-word term extraction. *Journal of Natural Language Processing*, 6(3):145–179.
Gibbon, D., Nord, H. U., and Trippel, T.
1997 *EAGLET: EAGLES Termbank for Spoken Language Systems*. EAGLES SLWG. http://coral.lili.uni-bielefeld.de/EAGLES/WP5/termdeliv97/.
Grishman, R. and Kittredge, R. (ed.)
1986 *Analyzing Language in Restricted Domains: Sublanguage Description and Processing*. Lawrence Erlbaum.
Halliday, M. A. K. and Matthiessen, C. M. I. M.
1999 *Construing experience through meaning: a language-based approach to cognition*. Cassell, London.
Hartmann, R. R. K. and James, G. (eds.)
2001 *Dictionary of Lexicography*. Routledge.
Hillmann, D.
2005 *Using Dublin Core*. DCMI. http://dublincore.org/documents/usageguide/.
Ide, N., Bonhomme, P. and Romary, L.
2000 XCES: An XML-based encoding standard for linguistic corpora. In *Proceedings of the Second International Language Resources and Evaluation Conference*, Athens.
Kageura, K.
1995 Towards the theoretical study of terms – a sketch from the linguistic viewpoint. *Terminology*, 1(1):103–119.
Kageura, K. and Umino, B.
1996 Methods of automatic term recognition: A review. *Terminology*, 3(2):259–289.
Kilgarriff, A.
1997 "I don't believe in word senses". *Computers and the Humanities*, 31(2):91–113.
Kilgarriff, A.
2003 Linguistic Search Engine. In *Shallow Processing of Large Corpora: workshop held in association with Corpus Linguistics 2003*, Lancaster.
Kittredge, R.
2003 Sublanguages and controlled language. In Mitkov, R., editor, *The Oxford Handbook of Computational Linguistics*. Oxford University Press.
Kittredge, R. and Lehrberger, J. (eds.)
1982 *Sublanguage: Studies of Language in Restricted Semantic Domains*. Walter de Gruyter.

Koehn, P. and Knight, K.
2002 Learning a translation lexicon from mono-lingual corpora. In *Proc. of the ACL-02 workshop on Unsupervised lexical acquisition*, pages 9–16.

Kozakov, L., Park, Y., Fin, T., Drissi, Y., Doganata, Y. and Cofino, T.
2004 Glossary extraction and utilization in the information search and delivery system for ibm technical support. *IBM System Journal*, 43(3).

Kučera, H. and Francis, W. N.
1967 *Computational analysis of present-day American English*. Brown University Press, Providence.

LDOCE
1995 *Longman Dictionary of Contemporary English*. Longman, Harlow, 3rd edition edition.

Lenat, D. B. and Gupta, R. V.
1990 *Building Large Knowledge-Based Systems: Representation and Inference in the Cyc Project*. Addison-Wesley.

Lin, D.
1998 Automatic retrieval and clustering of similar words. In *Proc. Joint COLING-ACL-98*, pages 768–774, Montreal.

Manning, C. D., Raghavan, P. and Schütze, H.
2008 *Introduction to Information Retrieval*. Cambridge University Press.

Max, A.
2012 Document authoring: from word processing to text generation. In Mehler, A., Romary, L., and Gibbon, D., editors, *Handbook of Technical Communication*. De Gruyter, Berlin/Boston.

McEnery, T. and Wilson, A.
1999 *Corpus Linguistics*. Edinburgh University Press.

Mel'čuk, I. A.
1996 Lexical Functions: a tool for the description of lexical relations in a lexicon. In Wanner, L., editor, *Lexical Functions in Lexicography and Natural Language Processing*, pages 37–102. John Benjamins, Amsterdam.

Menke, P.
2012 Evaluation of technical communication. In Mehler, A., Romary, L., and Gibbon, D., editors, *Handbook of Technical Communication*. De Gruyter, Berlin/Boston.

Miller, G.
1990 WordNet: an online lexical database. *International Journal of Lexicography*, 3(4).

Ng, H. T. and Kwong, O. O. Y. (eds.)
2006 *Proc. Fifth SIGHAN Workshop on Chinese Language Processing at joint COLING-ACL*, Sydney.

Niles, I. and Pease, A.
2001 Towards a standard upper ontology. In *FOIS'01: Proceedings of the international conference on Formal Ontology in Information Systems*, pages 2–9.

Nyberg, E., Mitamura, T., and Huijsen, W.-O.
2003 Controlled language for authoring and translation. In Somers, H., editor, *Computers and Translation. A translator's guide*, pages 245–281. John Benjamins.

OED
1989 *Oxford English Dictionary.* Clarendon Press, Oxford.
OFD
1994 *Oxford-Hachette French Dictionary.* Oxford University Press, Oxford.
Ogden, C. K. and Richards, I. A.
1923 *The Meaning of Meaning*. Routledge, London.
Pantel, P., Ravichandran, D. and Hovy, E.
2004 Towards terascale knowledge acquisition. In *COLING '04: Proceedings of the 20th international conference on Computational Linguistics*, page 771.
Paaß, G.
2012 Document classification, information retrieval, text and web mining. In Mehler, A., Romary, L., and Gibbon, D., editors, *Handbook of Technical Communication*. De Gruyter, Berlin/Boston.
Patry, A. and Langlais, P.
2011 Identifying parallel documents from a large bilingual collection of texts: Application to parallel article extraction in wikipedia. In *Proceedings of the 4th Workshop on Building and Using Comparable Corpora: Comparable Corpora and the Web*, pages 87–95, Portland, Oregon.
Rahtz, S.
2012 Representation of documents in technical communication. In Mehler, A., Romary, L., and Gibbon, D., editors, *Handbook of Technical Communication*. De Gruyter, Berlin/Boston.
Rayson, P. and Garside, R.
2000 Comparing corpora using frequency profiling. In *Proc. of the Comparing Corpora Workshop at ACL 2000*, pages 1–6, Hong Kong.
Resnik, P. and Smith, N.
2003 The web as a parallel corpus. *Computational Linguistics*, 29(3):349–380.
Rundell, M.
1999 Dictionary use in production. *International Journal of Lexicography*, 12(1):35–53.
Sager, J. C.
1990 *A practical course in terminology processing*. Benjamins.
Sager, J. C.
1993 *Language Engineering and Translation: Consequences of automation*. Benjamins, Amsterdam/Philadelphia.
Sasaki, F. and Lommel, A.
2012 Multilingual computing. In Mehler, A., Romary, L., and Gibbon, D., editors, *Handbook of Technical Communication*. De Gruyter, Berlin/Boston.
Sclano, F. and Velardi, P.
2007 TermExtractor: a web application to learn the common terminology of interest groups and research communities. In *Proc. 9th Conf. on Terminology and Artificial Intelligence, TIA 2007*, Sophia Antinopolis.
Sharoff, S.
2005a The communicative potential of verbs of "away-from" motion in English, German and Russian. *Functions of language*, 12(2):205–240.

Sharoff, S.
2005b Phenomenology and cognitive science. In Franchi, S. and Guzeldere, G., editors, *Mechanical Bodies, Computational Minds*, pages 471–487. MIT Press, Cambridge.

Sharoff, S.
2006 Creating general-purpose corpora using automated search engine queries. In Baroni, M. and Bernardini, S., editors, *WaCky! Working papers on the Web as Corpus*. Gedit, Bologna. http://wackybook.sslmit.unibo.it.

Sharoff, S., Babych, B. and Hartley, A.
2006 Using comparable corpora to solve problems difficult for human translators. In *Proc. of International Confenrence on Computational Linguistics and Association of Computational Linguistics, COLING-ACL 2006*, pages 739–746, Sydney.

Sinclair, J. (ed.)
1987 *Looking up: an account of the COBUILD Project in lexical computing*. Collins, London and Glasgow.

Smith, M. K., Welty, C. and McGuinness, D. L.
2004 *OWL Web Ontology Language*. W3C. http://www.w3.org/TR/owl-guide/.

Sowa, J. F.
2000 *Knowledge Representation: Logical, Philosophical, and Computational Foundations*. Brooks/Cole Publishing Co., Pacific Grove, CA.

Spärck Jones, K.
1972 A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation*, 28(0):11–21.

Sperberg-McQueen, C. M. and Burnard, L. (eds.)
2002 *TEI P4: Guidelines for Electronic Text Encoding and Interchange*. Oxford, 4 edition.

Staab, S. and Studer, R. (eds.)
2004 *Handbook on Ontologies*. Birkhäuser.

Stamatakis, K., Karkaletsis, V., Paliouras, G., Horlock, J., Grover, C., Curran, J. R. and Dingare, S.
2003 Domain-specific web site identification: The crossmarc focused web crawler. In *Proc. Second Workshop on Web Document Analysis (WDA)*, Edinburgh.

Temmermann, R.
2000 *Towards New Ways of Terminology Description*. John Benjamins.

Trippel, T.
2012 Controlled language structures in technical communication. In Mehler, A., Romary, L., and Gibbon, D., editors, *Handbook of Technical Communication*. De Gruyter, Berlin/Boston.

Varantola, K.
2003 Translators and disposable corpora. In Zanettin, F., Bernardini, S., and Stewart, D., editors, *Corpora in Translator Education*, pages 55–70. St Jerome, Manchester.

Waltinger, U. and Breuing, A.
2012 Internet-based communication. In Mehler, A., Romary, L., and Gibbon, D., editors, *Handbook of Technical Communication*. De Gruyter, Berlin/Boston.

Wright, S. E. and Budin, G., editors
2001 *Handbook of terminology management*. Benjamins.
Wüster, E.
1955 *Bibliography of Monolingual Scientific and Technical Glossaries*. Unesco.
Wüster, E.
1979 *Introduction to the General Theory of Terminology and Terminological Lexicography*. Springer.
Zhang, Z., Iria, J., Brewster, C., and Ciravegna, F.
2008 A comparative evaluation of term recognition algorithms. In *Proceedings of the Sixth Language Resources and Evaluation Conference, LREC 2008*, Marrakech.
Zipf, G.
1935 *The psychobiology of language*. Houghton Mifflin.

10. Multilingual Computing

Felix Sasaki and Arle Lommel

This article provides a high-level introduction to issues related to computing in multilingual environments. Although much of the early development of computing technology developed in monolingual environments, considerable effort has been devoted over the decades to improving the ways in which computers can deal with text in multiple languages, including those not written with the Latin alphabet, and in enabling computer software to deal with this data. Although working with multilingual data is now often taken for granted by users of mainstream computer operating systems, as recently as the 1990s many users ran into significant issues when trying to deal with text in languages other than a handful of most Western European languages. Starting with widespread adoption of the Unicode encoding standard (Unicode 2011) for text in the late 1990s, the situation with respect to multilingual computing improved rapidly and dramatically (Korpela 2006).

This article approaches the subject through an examination of the processes by which software and content are provided in multiple languages and some of the specific issues that must be accounted for when designing software that will work with multilingual data. In Section 1 we provide definitions of the basic terms *Globalization*, *Internationalization* and *Localization*. In Section 2 we introduce technologies and processes related to globalization like translation memories or terminology management. Section 3 introduces areas in which these technologies and processes are applied, like document translation and software localization. Section 4 provides details of basic technologies underlying multilingual computing, related to character encoding and processing, language and locale identification, and aspects of multilingual computing on the Web. Standardization of these and esp. localization-related technologies is discussed in Section 5. This encompasses various standardization bodies and standards like XML, XLIFF or TBX. The article closes with a conclusion in Section 6.

1. Background

This section describes four related terms (Ishida 2005), all of which have a major linguistic component: globalization, internationalization, translation, and localization. Unfortunately these terms are often used in ambiguous, overlapping, imprecise, or even contradictory ways. Because of this confusion, this section outlines a framework in which to discuss them. It is important to note, however, that not all organizations use these terms in the same ways, and that further clarification is often in order when discussing these topics to ensure that all parties have an equivalent understanding of the processes involved.

1.1. Globalization

Globalization is used in many ways with both positive and negative meanings. Although there is a considerable body of literature on the topics of economic and social globalization, in the context of this chapter, we use the term to mean *the entire process of taking goods and services to international markets, including all business decisions needed to reach those markets* (DePalma 2002). Globalization thus includes business plans, marketing decisions, partnerships, sales plans, after-sales support and anything else that needs to happen for a company to reach new markets, and is not limited to technical tasks. The following items in this section, internationalization and localization, are important *technical* components of this process and often represent a large portion of the direct cost in a globalization project (and it is common for software developers to use *globalization* to refer solely to these technical processes, see e.g. Esselink 2000). It is vital, however, to remember that the technical processes occur in a broader context. As a result, an organization's globalization plans must address questions such as the following:

1. In which markets will a business be active and what are the requirements for those markets?
2. Which products will it sell in which markets?
3. How will it sell those products?
4. How complete will localization be for those products?
5. How will it support those products in the long run?

Technical linguistic tasks are closely tied to this broader business picture because they are undertaken to solve business needs. Only when organizations are able to make informed decisions regarding their market plans can they make appropriate technical decisions to support their business goals.

1.2. Internationalization

Internationalization – often written as "i18n", where "18" represents the number of letters between "i" and "n" – is closely related to localization (see below) and precedes it. Internationalization is the process of designing a product, application, or document to facilitate subsequent adaptation of that product for other markets (Ishida 2005). Internationalization is sometimes called "globalization enablement" as well. The following are typical internationalization activities:

- **Enabling international multilingual data representation and processing** (see Section 4.3). This task can include enabling the use of Unicode (a "universal" character set that contains characters for most of the world's languages – see Section 4.2), ensuring the proper handling of legacy character encodings where appropriate, taking care of issues in string processing[1] (e.g. comparison or ordering), making sure that localizable content can easily be adapted (see Section 4.3 and Section 4.4), etc.
- **Providing means for localization for specific target markets**. For example, if software is to be localized for Arabic- or Hebrew-speaking markets, it must be able to support text that flows primarily from right to left but that contains segments that flow from left to right, or software deployed in Asian markets may need to support vertical text. Unless support for these market-specific linguistic requirements is incorporated in the base architecture of a product it is very hard to "add it in" at a later date.
- **Providing data necessary for localization.** Proper localization requires information (see Davis 2010a) about date, time or calendar formats, formatting information, handling of person names, and address forms etc. Data processing also requires means to identify *locales* (target markets) uniquely, a requirement which is sometimes not easy to implement (see Section 4.6).
- **Separating localizable content from source code or content**. Items that must be translated or otherwise adapted for specific markets must be accessible to translators or other technical staff. For example, if an application includes a graphic with embedded text, the developers must provide a way for this text to be translated rather than using a "flat" JPEG file. See Section 5.3 and Section 5.4 for further examples.

While such tasks are not part of localization itself, they facilitate subsequent localization. However, many of them are also very useful even if no localization takes place because they promote proper management of digital assets and clear design. They should be taken into account at the beginning of application and system development.

If a product has been created only with one market in mind, it will be very time-consuming and expensive to adapt it to other locales. In extreme cases localization may be rendered impossible within reasonable time and cost limitations without prior internationalization. (Ishida 2005) states that failure to internationalize products in advance typically results in costs and time required for localization to double. Even aside from issues of architecture, proper internationalization will result in costs savings by allowing problems to be resolved once (in the source) rather than multiple times (in each of the target locales), thus saving considerable time and effort. As a result internationalization should occur as a fundamental step in the design and development process rather than as an afterthought.

1.3. Localization and translation

The following discussion of localization and translation is based on (Ishida 2005). Localization – often written as “l10n”, where “10” is the number of letters between “l” and “n” – is the adaptation of a product, application or document content to meet all local requirements for a specific locale, including linguistic and cultural needs.

Translation of texts is often the largest component of localization but localization can involve much more than translation. For example, if an automobile made in Japan is to be sold in Germany, it must be localized by having the steering wheel moved from the right side to the left side, along with all of the corresponding changes to its internal mechanisms. An automobile may be an extreme example; localization is best conceived of as an engineering task with a core translation component involving user interfaces and documentation. Beyond these items, however, are other issues like:

– adaptation of numeric, date and time formats
– use of appropriate currency
– keyboard usage
– collation and sorting (see Section 4.4.3)
– symbols, icons and colors
– adaptation or change of text and graphics which may be subject to misinterpretation or viewed as insensitive in a given culture
– taking varying legal requirements into account

In addition to these issues and translation, localization requirements may lead to a substantial redesign of an application or a product, especially if internationalization steps were ignored or incomplete. Many developers have learned about the need for proper internationalization only when they have had to delay or even cancel plans to enter international markets because it was impossible to localize their products without significant reengineering. In some cases even unanticipated issues can require careful localization: In one famous, although perhaps apocryphal, story, a spreadsheet developer in the 1980s found that it could not sell its software in Japan because all errors generated an audible beep, thus notifying to everyone in a common office space that the user had made a mistake; after the software was modified to run silently it was much more acceptable in that market.

2. Globalization-related technologies

A number of sophisticated technologies have been developed to support the internationalization and localization aspects of globalization. In general, most of these technologies focus on the linguistic components of the globalization process because these are the most expensive aspects of the process and the ones that can derive the most benefit from technical solutions (dePalma 2002). This section discusses the general types of tools available and their functions.

2.1. Translation memory

Translation Memory (often abbreviated as *TM*) tools facilitate the reuse of translated text (Reinke 2004). By storing segments (pieces of text, typically sentences) in a database[2], along with their translations, a translation memory tool is able to examine new texts and determine what has been previously translated and what has been changed. It can thus automatically replace text that occurred in previous translations with appropriate translations from the database, eliminating the need for a human translator to retranslate the text. Developed starting in the 1970s, translation memory became popular in the early and mid 1990s as more and more projects were moved to computers and organizations realized that they could save significant amounts of money if they did not need to retranslate text every time they released a new project version.

A typical scenario for translation memory usage is the release of a new version of a product. For example, if a company develops a word processor (e.g. OpenOffice etc.), it will have a number of components, such as user manuals, online help and a user interface, all of which will have been localized for the product release. The next time the product is released (having gone from version 1 to version 2, for instance) much of the content in these items will be the same, while other parts will be completely new or modified. A translation memory tool takes the text in these, divides it into segments, and then compares it to the existing translation, automatically using unchanged portions. New material is collected for human translators, allowing them to focus only on the materials that require their attention. If a segment has been partially changed (for instance one or two words have been changed) it is considered a "fuzzy" match and is returned to the translator along with any similar previous translations that he or she might be able to modify to provide an appropriate new translation.

It is not uncommon for new releases of products to reuse 80 % or more of the text from previous versions, so translation memory technology can save organizations large amounts of time and money during localization. Even within new projects translation memory can provide significant savings since it is not uncommon for the same text to be used in multiple locations (e.g., in user manuals and online help systems). Beyond the simple savings in translation cost and time, translation memory technology also improves consistency within projects by ensuring that the same text is translated the same way each time and that users can expect a similar experience over multiple versions of a product. It thus serves a vital role in helping companies establish a consistent international brand experience.

2.2. Terminology management

Terminology management is the active control and documentation of the terms (specialty words) (Arntz, Picht and Mayer 2009). Specific products and domains (subject matter areas) use specific terms and it is important that translators know these words and use them properly. For instance, in U.S. railway terminology, one portion of the place where two railway tracks join together is known as a “frog”. This term differs from a “general language” understanding of “frog” and translators might not know the term, so it must be documented to assist them in either (a) understanding an English source text or (b) using the term when translating into English. Individual organizations often have their own specific terms that their users expect to see and using the wrong term can lead to confusion or dissatisfaction if it impairs understanding.

A common terminology problem is the use of inconsistent terms for the same concept (Arntz, Picht and Mayer 2009: 228). For example, if an English author uses the terms “start button”, “starter”, “on button”, “start switch”, and “on switch” in a document, how will these be understood by users, who may not understand them to be the same thing and thus waste time looking for multiple items in the product? When the product documentation is localized, how will these terms be translated? In general, most localizers will seek to adequately convey differences in the source text, so the apparent differences between these terms will lead to completely different translations in another language, thus magnifying the problem.

As a result of these problems, most translation tools have a component for managing terminology. (Arntz, Picht and Mayer 2009: 228–256) provide an overview of the related tooling, i.e. computer assisted terminology management. At the simplest level, many translators use spreadsheets or word processors to make simple lists of terms and translations (called *glossaries*). Glossaries serve simple needs but are not suitable for more complex tasks where additional information (metadata) about terms (such as definitions, usage notes, relations to other terms, notes regarding translation, and so forth) may be needed. As a result, full-fledged terminology management systems are capable of storing considerably more information about terms and concepts. The best systems are *concept-oriented*, that is, they are based around units of meaning rather than words. A concept-oriented terminology database, for example, would store the information that “USB” is an acronym for “universal serial bus” and that both refer to the same concept, which might be called something else in other languages. These databases may contain complex metadata needed to facilitate the work of translation. In most cases terminology management systems are also integrated with translation memory systems to provide terminology information to translators while they are working.

2.3. Machine Translation (MT)

Machine Translation, or MT, is one of the oldest translation technologies, with roots going back to military research done in the 1940s. (Koehn 2010:14–19) provides an overview of the history and the approaches to machine translation described below. In machine translation a computer translates the text itself (in contrast to translation memory where the computer recycles previously completed human translations). Although there a number of techniques used in machine translation, the technology typically works best in narrow subject fields with controlled vocabulary (see Ladkin and Trippel both in this volume) and simple grammatical structures. For many years MT was the butt of jokes and not considered a "serious" translation technology, but in recent years is has become more widely used to provide information-only translations (a process called "gisting") on the Internet and by companies that have to translate large amounts of text in a short amount of time. It has also found substantial use by companies with large volumes of support information that customers may need but which would be cost-prohibitive for humans to translate. By some estimates, MT may be the most widely used translation technology, although such usage is typically not carried out by content creators to distribute their content but rather by information consumers who wish to access content that has not been otherwise translated.

Prior to 2000 most MT systems were *rule-based* systems that analyzed texts grammatically and then attempted to interpret their structures and create corresponding texts in the target language. Rule-based MT systems were difficult to develop and maintain, so in more recent years there has been a shift to *example-based* or *statistical* MT systems that rely on access to large bodies of human-translated text to provide translations based on the available examples. A famous example of such a system is the Google Translate service that uses billions of aligned sentence pairs. The Moses system (Koehn et al. 2007) is a widely used open source MT system that allows for automatically training translation models for any language pair. Such statistical MT systems are easier to develop and offer more natural output, but the goal of a translation system that is the equal of a human translator has remained elusive.

2.4. Localization workbenches

It is common for translation memory and terminology management tools to be combined together into integrated *translation workbenches* (Esselink: 359–396). These tools aim to provide a complete translation environment to human translators. They may include access to reference works and often include additional tools to facilitate translation. For example, they may provide the ability to provide "pseudo-translations," texts that mimic the characteristics of real translations (such as increased length and use of accented characters or other writing systems), so that localization projects can be checked for potential problems (such as character corruption, see Savourel 2001: 41, or lack of space in cases where translated texts are longer than the source text) even before a real translation is complete. These workbenches often include other functions for specific tasks, like software localization, so that translators can be more effective for particular tasks.

2.5. Global Translation Management Systems (GTMS)

As the Web became more and more important to organizations, they were faced with the daunting task of keeping multilingual web sites up to date and synchronized. Because web sites are often created by many people working independently, traditional project planning cycles that rely on strict control and sign-off structures were not easy to implement. As a result in the late 1990s a new technology solution to the problem of website translation was developed (Draheim and Toon 2005), the *global translation management system (GTMS)* or *global content management system (GCMS)*. At the heart of these systems is a server-based *change detector* that monitors the site for changes. When changes are made to the web site, the GMTS utilizes a set of business rules to automatically decide what actions need to be taken. In the case of high-priority content, the business rules might dictate that it be sent for translation immediately, while other content might be batched for later action (e.g., it might be translated once a month). These systems generally manage the entire translation process from start to finish with only minimal human intervention and are able to track and manage complex projects with many small components that would be difficult or impossible to manage manually. More recently these systems have been extended to cover other sorts of content besides web pages, and can be used to manage the rapid turnaround demanded in projects where time to market is a critical concern.

3. Globalization scenarios

One of the difficulties in globalization is the diversity of project scenarios. Understanding both general principles and specific concerns and issues is vital when making plans for globalization projects. This section describes both general globalization processes and a variety of issues with which developers must be aware if they are to properly localize products.

3.1. Document translation

Prior to the development of computer systems, translation was the only task needed for most globalization projects and today the largest component in terms of time and investment is translation. Broadly speaking, most globalization projects consist of *technical translation*, rather than *literary translation* (Esselink 2000: 293–309). This sort of translation is characterized by an emphasis on the transfer of information rather than artistic effect. It is therefore treated in a production-oriented fashion with most such translation paid for by the (source) word with discounts for translation memory use. In addition to technical translation, there may also be translation of marketing materials that requires closer attention to cultural and artistic effects.

Most globalization efforts require at least some translation of documents, and as a result the processes outlined in this section apply to some extent to most products, allowing for appropriate differences.

Most translation projects start when a document containing text is complete and has been reviewed and approved for release. At this point the document is sent out to an external service provider to be translated. It will then undergo a process something like the following:

1. **Pretranslation**. The document is processed using a translation workbench's *filter* functions to extract the text in the file and is then analyzed against a translation memory database (Reinke 2004) to identify reusable translations. If no translation memory database exists but an earlier version of the text was translated, the translator may opt to *align* the previous versions to create a translation memory database. The extraction process is important since text is often mixed in with formatting markup or binary data or in a document while translation tools can generally only work with plain or rich text. Similar processes may be applied to any graphics so that they can be translated. After text has been extracted and compared against any translation memory resources, it may be used for *terminology mining*, the process of identifying characteristic terms that may require further attention during the translation process.
2. **Translation**. During this phase the service provider sends the text to the actual translator, who usually works in a translation workbench environment to translate any untranslated text. The translator may also engage in terminology work to identify and document key terms, either found during pretranslation or during his or her work. Before returning the translation the translator should review the text and make his or her corrections to it.
3. **Linguistic review/quality control**. The text is then reviewed for quality issues, generally by someone other than the original translator. Corrections are made in the translation workbench environment so that they are reflected in the translation memory and will be available in the future.

4. **Desktop publishing (DTP)**. Most non-XML formats (XML formats are explained in Section 5 – see also Stührenberg in this volume) require some sort of desk-top-publishing with applications like Adobe InDesign or QuarkXPress to give them the proper format and appearance. In this phase the translation vendor uses the workbench's filter functions to recreate an appropriate document in the application file format from the translated text. In most cases they will then need to manually import translated graphics and make adjustments to the layout of the translated document (translated texts rarely take up the same amount of space as the original document and thus require editing to maintain pagination and to keep text and graphics aligned). In the case of well internationalized documents, the DTP phase may be quite simple, but in other cases it may involve extensive work.
5. **Final review**. After the translated document has been formatted it is reviewed once again. At this point the review is more focused on appearance and layout issues and making sure that the translation job is complete (e.g., that no graphics have remained untranslated and that indexes and contents accurately match the translated content). While linguistic changes may be made at this point, it is better to make them earlier on since they will not be reflected in the translation memory database unless the database is manually corrected to match them.
6. **Return**. After passing through these steps the translated document is returned to the client. Depending on the contract, the vendor may also provide the translation memory database and the results of any terminology research.

This general process may have variations depending on the sort of document being translated, but most processes involve these steps.

3.2. Software localization

The software localization process builds upon the general translation process by adding a number of important steps that are specific to software localization. This section provides a highly simplified view of software localization, focusing on major steps in the process. For further details, the reader is encouraged to consult (Reinke and Schmitz 2005) or Section 3 of (Esselink 2000) on the subject.

1. **Internationalization verification**. Improperly internationalized software cannot be localized without additional engineering steps. As a result it must be carefully tested to ensure that there are no obvious or hidden internationalization problems that will cause problems later on in the localization process. One of the key methods for verifying internationalization is *pseudo-translation* (Esseling 2000: 149). As mentioned earlier, in pseudo-translation translatable texts are replaced with "garbage" text that mimics the

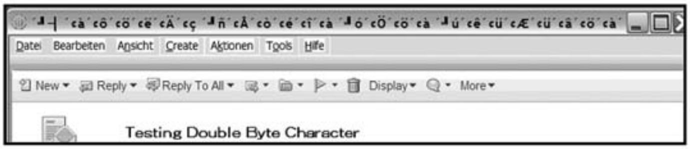


Figure 1. An example of poor internationalization. In this instance a process for rendering window titles did not know how to properly interpret text not in the Roman script and “garbled” it. (Example courtesy of Dale Schultz, IBM.).

characteristics of real text, including word length and character repertoire. For example, if software is being translated into Japanese, English software strings would be replaced by strings of Japanese characters. The software is then compiled to see if any problems arise, such as text that remains in English (because it was embedded in the code rather than properly *externalized* during the internationalization phase), text that is displayed incorrectly (e.g., the Japanese characters “東京都” are transformed into “ï»¿æ ±äº¬éƒ½” because an internal process was not properly enabled to use Unicode) (see Figure 1), or user interface elements that may cut off text that is longer than the English source text. In addition to pseudo-translation testing, the application is tested with data typical of the target locales where it will be released to make sure that it can properly handle that data (e.g., that software won’t convert Japanese names to garbage because it expects only Roman characters or that it won’t crash when encountering data in an unexpected format). Any problems that are found in this phase must be addressed before further localization tasks can take place.

2. **Extraction of localizable resources**. Although software should have localizable resources separated from code, these resources are often stored in formats that require additional extraction before translators can use them. In some instances, translators may also require access to resources not part of the actual software executable as well, such as source files for graphics. All of the appropriate content must be extracted and assembled for easy access by the translators.
3. **Interface translation**. After internationalization verification is complete, all translatable text is extracted and translated, much as it is with simple documentation. Because translatable text may be mixed in with special codes used by computer programs, the translation phase often includes a series of intermediate builds that allow the translator to verify that the text functions as anticipated in context. In addition, software localization typically uses more specialized workbenches that are designed to work with typical software resource types.
4. **Layout adjustment**. Localized software typically requires adjustments in the user interface to allow for linguistic and cultural differences. In simple

cases these will consist of items like changing the size of elements in user interface dialogs to allow for text expansion or changing the order of elements like address or name boxes to match local conventions (e.g., putting family names before given names in localizations for East Asia or Hungary). In more complex cases, layout adjustment can involve the wholesale rearrangement of a user interface or "layout mirroring" for Arabic or Hebrew software. (In layout mirroring, the user interface's direction is reversed from left-to-right to right-to-left, affecting almost every element in the user interface by moving labels to the right of them and reversing the order of grouped items.)

5. **Translation of documentation and help**. After the user interface has been translated and adjusted, any additional content, such as documentation or help, is translated. Because these components require knowledge of the final interface of the software, they are typically translated near the end of the development process.
6. **Functional and usability testing**. After the software has been translated and layout adjustments are complete, the software must be tested thoroughly to make sure that no functional or usability issues have arisen. Although any changes to a program may have the potential to create unforeseen problems, testing generally focuses on those areas that involve text and data handling. At this point the software is tested with locale-appropriate data to verify that it is processed as expected and that there are no problems with items such as currencies, alphabetic sorting, local calendars, time zones, etc.

Software localization requires considerable technical skill and an understanding of the software development process (Section 3 of Esseling 2000). As a result, there is often a division of labor between technical staff with this knowledge, and translators who focus solely on translating content.

4. Multilingual data representation and processing

This section examines some of the fundamental issues in representing multilingual text and data on computers. Software that has been properly internationalized and localized, as described in Section 3, must account for the types of issues addressed in this section. Depending on what locales and languages are addressed, the specific issues may vary, but this section provides an overview of core issues encountered when developing international software.

4.1. Characters, glyphs, fonts and scripts: How “snow” is different from “snow” …

Characters are one of the key topics in multilingual data. But what is a character? Let us introduce some central concepts related to the notion of characters by looking at the ideographic character for “snow”, see Figure 2.

Figure 2. Variations of the ideographic character for “snow”.

The figure shows variations of the character 雪, which means “snow”. This meaning is essential for the definition of the character, and the distinction to others like 雨 (“rain”). The variations in the figure are different *glyphs* for the “snow” character, that is, different abstract visual variants. A *font* assembles glyphs images and is the concrete visualization which we see on a screen or paper. To demonstrate the relation between glyphs and fonts, Figure 3 shows the “snow” character in the Japanese glyph variant in two different fonts.

Figure 3. Font variants of the character for “snow”.

The glyph variants of the “snow” character in the first figure are due to differences in the Chinese, Japanese and Korean *scripts*, also called writing systems (Sasaki and Witt 2003). The character repertoire shared by the Unicode standard and ISO/IEC 10646 encompasses a unification of many characters from these and other scripts. The “snow” character is a prototypical example of the result of this unification: one character in Unicode represents characters from several scripts.

4.2. Unicode

ISO/IEC 10646 is a standard developed and published jointly by ISO and IEC. The Unicode standard is published by the *Unicode Consortium* (Unicode Consortium 2011). Both share the same character repertoire and encoding form. In addition, the Unicode Standard adds information (see below) related e.g. to normalization, handling of bidirectional text and various other implementation information.

Unicode encodes widely used scripts and unifies regional and historic differences. For some purposes this unification is too broad, or some minor scripts are not (yet) part of Unicode. As one solution to the first problem, Unicode provides "variation selectors" which follow a character to identify a (glyph related) restriction on the character. For the second problem, there is no real solution other than to get your character into the Unicode character repertoire. Since Unicode is produced by an industrial consortium, minority scripts and historical scripts have a small lobby. The *Script Encoding Initiative* (Anderson 2003) is participating in the Unicode consortium to give rarely used scripts a voice. Another solution is proposed by the Text Encoding Initiative (see also Rahtz in this volume): the use of Markup to express glyph variants[3].

4.3. Character encoding

A *character encoding* (see Section 4.1 of Dürst et al. 2005) encompasses the information necessary for representing and processing textual data. For example, consider Table 1, which shows how the "snow" character is encoded in Unicode. (Lunde 2009) provides further, detailed information about character encoding of Chinese, Japanese and Korean.

Table 1. Encoding of the "snow" character in Unicode.

Character 雪 "snow". Code point (hexadecimal): U+96EA									
Character encoding	UTF-8			UTF-16		UTF-32			
Code unit string	E9 9B AA			96EA		000096EA			
Byte string	E9	9B	AA	96	EA	00	00	96	EA

In a character encoding, a character receives a unique numeric identifier, a *code point*. In Unicode, the "snow" character has the hexadecimal number 96EA (represented in Unicode with the prefix "U+"). Another part of a character encoding is the *character encoding form* for the code point. Unicode provides three main character encoding forms. All of them share the same code points, that is, the same *character repertoire*. The character encoding forms differ in the basic data type used for representing code points as *code unit strings*. UTF-32 uses as a data type a fixed length of 4 bytes, e.g. 00 00 96 EA. UTF-16 uses 2 or 4 bytes. In the example only two bytes are necessary for representing the "snow" character as UTF-16. Finally, UTF-8 uses a variable sequence of bytes.

Names of character encodings in Unicode are based on the sort of character encoding used. The most widely used Unicode character encoding is UTF-8. If a document contains only Latin based textual data, UTF-8 will lead to small file size, since this data can be represented mostly with sequences of single bytes. If size and bandwidth are not an issue, UTF-32 can be used. Implementing basic routines like character counting becomes very easy (just divide the total number of bytes by four). However, especially for web data, UTF-32 will slow down processing. UTF-16 is for environments which need both efficient access to characters and economical use of storage. In practice the usage of UTF-8 or UTF-16 is recommended, since support for these encodings is mandated by widely used technologies like XML (see Section 5), and UTF-32 is used less frequently.

However, although Unicode provides coverage for most of the world's scripts, many other national character encodings are used as well. Table 2 provides examples, again for the "snow" character in three national character encodings.

Table 2. Examples of three Japanese national character encodings.

Japanese: JIS X 0208	Korean: KS X 1001:1992	Chinese: Big Five (Taiwan)
3267	6468	B3B7

Transcoding (see Section 4.2 of Dürst et al. 2005) is the process of converting between character encodings. Transcoding without ambiguities between Unicode and other character encodings is not always possible. As an example, the report "XML Japanese profile" (Murata 2005), summarizes issues for Japanese specific character encodings.

4.4. Character processing: counting, comparison, and sorting

4.4.1. Character searching

Approaches to searching for characters vary across technologies and the simple question of "how many characters" there are in a piece of text does not have a simple answer (Sasaki 2010). For example, in Java borders of characters are the basic unit for regular expressions (Friedl 2006). Hence, Java takes the beginning of an input sequence into account, even if that sequence is empty. In contrast, XML Schema does not take borders into account, leading to different results in text searches. For example, given the regular expression "a?" would match the empty string "" in Java, but would *not* match it in XML Schema (Walmsley 2001).

4.4.2. Character comparisons

If character sequences are to be counted and compared, they must be in the same character encoding. As mentioned in the discussion of transcoding, this process is sometimes not a trivial one. Another prerequisite for comparison is that strings have to be in the same *normalization form* (Davis 2010b). Normalization is the process of bringing two strings to a canonical form of representation before they are processed. This step is necessary because some character encodings allow multiple representations for the same (surface) string. For example the character ç ("LATIN SMALL LETTER C WITH CEDILLA") can be represented in Unicode as a single character with the code point "U+00E7" or as a sequence of two code points – "U+0063 U+0327" – that represent the "base character" (c) and the "modifier" (¸) as separate characters, even though they are visually one character. In this case it is obvious that even a simple search will provide unforeseen results, if normalization is not assured, since a search for one sequence will not find the other sequence, even though they represent the same surface string.

Unicode defines four normalization forms (see Section 1.2 of Davis 2010b). While the details are not important for this discussion, Table 3 shows the normalization for two sample input and demonstrates the complexity of this topic.

Table 3. The four normalization forms of two Unicode characters.

Normalization form	NFD	NFC	NFKD	NFKC
Example input 1	U+00E7 "ç" LATIN SMALL LETTER C WITH CEDILLA			
Code points 1	U+0063 U+0327	U+00E7	U+0063 U+0327	U+00E7
Example input 2	U+FB03 "ﬃ"LATIN SMALL LIGATURE FFI			
Code points 2	U+FB03	U+FB03	U+0066 U+0066 U+0069	U+0066 U+0066 U+0069

Especially in the case of text on the Internet, where intermediate servers (proxies) may transcode text, what normalization form to expect often cannot be foreseen. Hence, before comparison operations, all text should be transcoded to use one consistent normalization form. Guidance about this topic is given by (Yergeau et al. 2005).

4.4.3. Sorting

Finally, we arrive at sorting characters via a *collation* (Section 3.5 of Dürst et al. 2005). A collation is a sequence that defines how character strings are sorted. The simplest collation is based on the numeric order of codepoints in the character repertoire, but most useful collations require some knowledge of culturally relevant sorting orders. For example, in German the character *Ä* is sorted after *A*, while in Swedish it is sorted after *Z* and in Hungarian *cs* is treated as one character that is sorted after *c*, so that in Hungarian the word *cukor* would sort *before* the word *csend*. Even within one language, multiple sort orders may apply. For example, in German the two strings “Strasse” and “Straße” may be sorted and treated as identical or as different, depending on whether the German “phone book” order or lexicon order is used.

4.5. Language identification: again “snow”

As shown in the German example, collations are related to language and locale identification. Although these concepts are related, they are not identical. First we will discuss language identification and then turn to the concept of locale.

An important use case for language identification was already introduced at the start of section 4.1, where it was shown that proper language identification is needed to select the appropriate glyph for that language. Figure 4 shows the underlying XHTML source code used to visualize the Chinese, Japanese and Korean glyph variants.

```
<span xml:lang="zh-CN">[雪 zh-CN]</span> <span xml:lang="ja">[ 雪
ja]</span> <span xml:lang="ko">[ 雪 ko]</span>
```

Figure 4. XHTML source code for glyph variants.

Modern browsers use the language information in the `xml:lang` and `lang` attributes for selecting the appropriate glyphs from the given font. The attributes contain values following the BCP (“Best Current Practice”) 47 from the IETF (Internet Engineering Taskforce). In addition to `xml:lang` or `lang`, BCP 47 is for example used in the HTTP protocol or other internet and Web technologies (Shklar 2009).

A BCP 47 *language tag* (see examples in the table below) consists of *sub tags* and has the following, here slightly simplified structure:

```
langtag = (language ["-" script] ["-" region] *("-" variant)
*("-" extension) ["-" privateuse])
```

- a language sub tag[4] like “en” (“English”), expressing information about the language.
- an optional script sub tag like “Cyrl” (“Cyrillic script”).
- an optional region sub tag like “AT” (“Austria”).
- zero or more variant sub tags like “1901” (“German spelling of 1901”).
- zero or more extension sub tags.
- an optional privateuse sub tag.

Table 4 provides examples of languages tags.

Table 4. Examples of language tags.

Language tag	sub tag structure	Meaning
`en`	language	English
`de-AT`	language-region	German, Austrian region
`de-CH-1901`	language-region-variant	German, Swiss region, with spelling in 1901 variant
`sr-Cyrl`	language-script	Serbian, in the Cyrillic script

4.6. Locale: how a German might search for pizza in Tokyo

Although we used the term "locale" at the beginning of this article, we take a closer look at its meaning here. Locale is a concept closely related to language. Loosely defined as a "target market" above, it is actually more complex: It means the accumulation of language, region and culture-related user preferences. While language is naturally an important part of locale, the concept of locale goes beyond language. For example, speakers of French might wish to use a variety of currencies, including the euro, the Canadian dollar, and various African and South Pacific currencies (considering only those from French-speaking areas), issues that are not determined by language alone.

The Unicode consortium has created LDML ("Locale data markup language"), which uses BCP 47 values as the basis for locale identifiers. LDML is used for CLDR ("common locale data registry" – http://www.unicode.org/cldr), which provides a large amount of locale-related information. Unfortunately there is no general, agreed-upon standard for locale identification, a problem that is not limited solely to the Internet. Hence, if a service offered on the web wants to take user preferences into account, there is a lot of "just guessing".

As an example for this situation, let us assume a German user who wants to order a pizza while visiting Tokyo. A user preference, conveyed by the browser language setting to for example the Google search engine, might lead to the following query:

```
http://www.google.co.jp/search?hl=de&q=pizza
```

Unfortunately the results for this query, although provided by the Japanese Google domain www.google.co.jp, are pizza stores in Germany. The reason is that the search engine interprets the language setting "hl=de" as locale information and relates it to the region of Germany. The following query uses the setting "hl=ja" and leads to the desired list of Japanese pizza stores.

```
http://www.google.co.jp/search?hl=ja&q=pizza
```

One undesired side-effect of the setting "hl=ja" is that the result page is presented in Japanese. A query whose results are presented in German, but which encompasses pizza stores from Japan, can be achieved by changing the actual query string to contain a Japanese word ピザ (pizza), but keeping the preference for German "hl=de" (although this search requires that the German visitor know the Japanese word for pizza):

```
http://www.google.co.jp/search?hl=de&q=ピザ
```

Formats for locale information and locale specific services on the Web still need some work, and the issues are not only technical, but also relate the understanding and anticipating the needs of human users of the Internet that are not easily anticipated within a simple locale model. The World Wide Web Consortium (W3C) is developing documents for "Language and Locale Identifiers on the Web" (Sasaki 2006) and "Web Services Internationalization (WS-I18N, Phillips, Trumble and Sasaki 2008), which will provide guidance about these topics.

4.7. Further topics and challenges of multilingual processing on the Web

In this section, we have presented some of the basics of multilingual data representation and processing. There are many aspects that we could not touch upon, like internationalization of Web addresses (Ishida 2008), visualization of right-to-left scripts (e.g. Hebrew or Arabic, see Section 3.3 of Dürst et al. 2005), and style-related topics like vertical layout (Anan et al. 2009).

Multilingual data on the Web has many unresolved issues. Information about the character encoding of Web data is frequently missing, which hinders basic text processing and can lead to web pages appearing as garbled text to end users. An unknown normalization form and transcoding or normalization via proxies may lead to unexpected results for string comparison. And finally, formats and the architecture of services for language and locale identification are not yet fully developed. Behind all these issues is the general problem of the availability and reliability of metadata, e.g. metadata about character encoding or language and locale. This is not specific to multilingual data, but to the Web at large.

5. Standards

In the previous section we discussed some of the complex issues that can cause problems for international text processing and globalization projects. While much work remains to be done in terms of improving these issues, the situation is actually considerably better than it was only a few years ago. Largely as the result of concerted industry effort to define standards for data and processes, issues that plagued globalization efforts in previous years have either disappeared or become much less serious.

In this section we will consider the groups responsible for defining standards in the area of multilingual text processing and the standards themselves. Each standard will be explained briefly and examples of its use given.

5.1. Standards bodies

The most important standards bodies currently active in this area are (in alphabetical order), the International Organization for Standardization (ISO), the Localization Industry Standards Association (LISA), the Organization for the Advancement of Structured Information Standards (OASIS), the Unicode Consortium, and the World Wide Web Consortium (W3C):

– **ISO**. Based in Geneva, Switzerland, ISO is active in multilingual content standards through technical committees 37 (“Terminology and other language and content resources”) and 46 (“Information and documentation”). These committees work collaboratively with other organizations to develop standards and are responsible for a number of multilingual standards.
– **LISA**. Based in Romainmôtier, Switzerland, LISA was[5] a member-driven organization with members from a variety of industries that are involved in the localization process. Through its Open Standards for Container/content Allowing Reuse (OSCAR) open standards committee, LISA developed standards for language technology, with an emphasis on data portability and reusability.
– **OASIS**. OASIS is an organization that promotes the development of open standards for XML. OASIS has been active in the development of the XLIFF standard, which will be discussed below, and is currently involved in the development of overarching framework standards for globalization processes.
– **Unicode Consortium**. In addition to its efforts in developing the Unicode standard for character representation, the **Unicode Consortium** has been a leader in the gathering of publicly accessible locale-specific data.
– **W3C**. As the primary developer of standards for the Web, the W3C has been involved with many organizations to ensure that international features are considered in Web technologies and has released the *Internationalization Tag Set* (ITS) standard for XML content.

Together these organizations provide a wide variety of standards (primarily XML oriented in nature) that can be used to simplify globalization processes.

5.2. XML: a ubiquitous framework

eXtensible Markup Language (XML, see Harold and Means 2004 – see also Stührenberg in this volume) is a set of rules for the electronic encoding of documents. It is not a forma *per se* but rather a description of how specific formats can be defined using consistent structures. XML has become widely used for representing any sort of data, not just in technical communication, with XML formats defined for graphics, business directory data, transmission of project management data, representation of musical notation, and many other applications. For multilingual data, XML provides three primary benefits: Unicode support, internationalization of XML-related technologies, and usage of XML for internationalization and localization technologies themselves.

All XML processors have to support at least the UTF-8 and UTF-16 Unicode character encodings (see Section 4.3). "Support" means that nearly any Unicode character can be used as the value of elements or attributes in XML. Although XML element or attribute names or unique identifiers (IDs) were restricted to Unicode (3.2) and its character repertoire for a long time, this restriction has recently been lifted so that most characters in existing or future versions of the Unicode are now allowed.

Another benefit of XML is the internationalization in XML related technologies. Two examples are the transformation language XSLT and the formatting language XSL-FO, which are also introduced in (Harold and Means 2004). XSLT provides among others the `xsl:number` element as a means for language- or locale-specific numbering. The difference between language and locale here becomes a bit blurry (see sec. 4.6). XSL-FO provides properties like script or language which can be used for parameterization of layout characteristics like hyphenation, writing mode (for e.g. Arabic or Hebrew), baseline for glyph alignment etc. Other technologies, which we can only be mentioned here, are XML Schema for document validation (Walmsley 2001), XQuery and XQuery full text for document analysis (Walmsley 2007), and XProc for building XML processing pipelines (Walsh, Milowski and Thompson 2010).

XML is also used for formats specifically dedicated to internationalization or localization. In the following subsections we will introduce some of them.

5.3. XLIFF

XLIFF is the *XML Localization Interchange File Format* (see Savourel 2001). Its purpose is the interchange of localizable data. As we have seen in Section 1.3, localization is a process with many participants. XLIFF allows them to convey localization data and related information in an XML-based, tool independent format.

An XLIFF document contains information like:

- localizable objects, for example text strings extracted from the data in the source language and translated, potentially several target languages.
- additional information like glossaries or about the format in which the source is given.
- localization workflow information.
- extensions, for example tool-specific information.

As (Savourel 2001) describes, there are two basic approaches to use XLIFF: the *minimalist approach* and the *maximalist* approach. The former means that first, localizable information is extracted from the source data and put into an XLIFF document. The non-localizable information is kept in a separate s*keleton*. After translation into a target language, the material from the XLIFF file is merged again with the skeleton. Figure 5 gives an example for a simple HTML file, a related skeleton file with place markers for extracted text, and the XLIFF file.

```
<html><head>
 <title>The document title</title></head>
<body><p>Very simple content here.</p><pre>Some code, not
translatable</pre>
</body></html>
```

```
<html><head>
<title>%1%</title></head>
<body><p>%2%</p></body></html>
```

```
... <header><skl><external-file href='sample.skl'/></skl></header>
<body>
<trans-unit id='%1%'>
<source xml:lang='en'>The document title</source>
</trans-unit>
<trans-unit id='%2% "restype='x-html-p'>
<source xml:lang='en'>Very simple content here.</source>
</trans-unit>
</body> …
```

Figure 5. Simple HTML file, related skeleton, and XLIFF file.

In contrast, the maximalist approach, see Figure 6, means that an XLIFF document contains the complete source data. That means that also structural information, like in the case of HTML the name (e.g. “x-html-title” or “x-html-p”) and position of the element from the source, is kept in the XLIFF document.

```
... <body><group restype='x-html-html'>
 <group restype='x-html-head'>
  <trans-unit id='1' restype='x-html-title'>
   <source xml:lang='en'>The document title</source>
  </trans-unit>
 </group>
 <group restype='x-html-body'>
  <trans-unit id='2' restype='x-html-p'>
   <source xml:lang='en'>Very simple content here.</source>
  </trans-unit>
 </group>
... </group></body>
```

Figure 6. XLIFF maximalist approach.

XLIFF solves several problems: it provides interoperability between localization tools, supports the general localization workflow, reduces the need for proprietary intermediate formats, allows for easier integration of adjunct processing like spell checking etc. (Questions of interoperability in scientific communication are discussed by Romary in this volume). Since it is an XML-based format, it brings all advantages of XML itself like Unicode support, widespread adoption and many related tools for rendering or other processing.

5.4. Internationalization Tag Set (ITS)

To create XLIFF files, it is necessary to know which parts of the source document needs to be translated or not. An example for the latter is the content of the pre element in the HTML document in Section 5.3. To convey this and other information related to localization and internationalization of XML source data, the *Internationalization Tag Set* (ITS 1.0) has been developed. ITS 1.0 provides several so-called *data categories* related to localization and internationalization:

- “Translate” conveys information about what parts of an XML document should be translated or not.
- “Localization Note” is a means to communicate notes to the localizer.
- “Terminology” is used to mark terms and associate information with them, e.g. definitions.
- “Directionality” indicates the directionality of a run of text, to support visualization of right-to-left scripts like Hebrew or Arabic.
- “Ruby” is an annotation to a run of text, providing e.g. pronunciation information for pictographic scripts like Japanese.
- “Language Information” indicates the usage of BCP 47 language identifiers, see sec. 3.3.
- “Elements within Text” describes the influence of the XML document structure on text flows, e.g. nesting of flows or independent flows.

For most data categories, there are two ways of usage: *global* and *local*. Local means that ITS markup is embedded directly into a source document, see the example in Figure 7.

```
<html xmlns:its="http://www.w3.org/2005/11/its"><head>
 <title>The document title</title></head>
<body><p>Very simple content here.</p>
<pre its:translate="no">Some code, not translatable</pre>
</body></html>
```

Figure 7. ITS embedded into a source document.

The default is that element content should be translated, and attribute values are not translated. The `its:translate` attribute at the pre element indicates that the content of that element will not be translated. An ITS processing tool, available e.g. in SDL Trados Studio, can use this information for example to make sure that pre is not taken into account for XLIFF generation.

The other, global usage of ITS conveys the same information as local ITS markup. The difference is that used globally, ITS is independent of the location in the source document and can be also applied to several documents. This is achieved by using XPath (Clark and DeRose 1999) in so-called *ITS global rules* for pointing to the markup (elements, attributes) to which ITS information should be attached. As an example, the ITS rules file in Figure 8 expresses that all pre elements in HTML documents should not be translated.

```
<its:rules version="1.0" xmlns:its="http://www.w3.org/2005/11/its"
  xmlns:h="http://www.w3.org/1999/xhtml">
 <its:translateRule selector="//h:pre" translate="no"/>
</its:rules>
```

Figure 8. Example of ITS global rules.

Within localization workflows, ITS so far mainly played the role of preparing the actual localization and of providing metadata for parts of the process. We have demonstrated that the preparation information is useful for a subsequent step of generating XLIFF documents. ITS 1.0 deserves its name as the Internationalization Tag Set, since many of its data categories like Directionality information or Ruby are specific to internationalization. As time of writing, ITS 2.0 is under development. ITS 2.0 makes data categories available for HTML5 and will add data categories for language technology applications and localization workflows.

5.5. Translation Memory eXchange (TMX)

During the 1990s as translation memory developed, it was realized that the variety of TM tools and formats were not interoperable. Linguistic assets in TM tools can represent a substantial investment, with some of the largest users of TM maintaining hundreds of millions of Euros worth of translations (see the LISA *Translation Memory Survey 2004* – http://www.lisa.org/Translation-Memory-S. 518.0.html for more detail). This data was, however, held "captive" by proprietary tool formats. As a result, in 1998, the first version of Translation Memory eXchange (TMX) was released. TMX provides a mechanism for exchanging TM data between tools. Any two tools that implement TMX can exchange data between them, thus making this data portable between them. Currently at version 1.4b, with version 2.0 under development, TMX has been implemented in tools used by hundreds of thousands of translators.

TMX consists of two parts:

1. A *container* consisting of the higher-level structure of a TM database: aligned segments of text in two or more languages
2. A specification for the format of the *content* of these segments (in TMX called *metamarkup*). This portion specifies how content formatting is represented within segments.

Although the full structure of a TMX file is not presented here (the current standard is available at http://www.lisa.org/tmx), the simplified code snippet in Figure 9 shows the essential structure:

```
[...]
<tu tuid="0002" srclang="hu">
    <tuv xml:lang="hu">
        <note>This segment is a sample</note>
        <seg>
             A hangmagasságot úgy változtatta, hogy egy keskeny
fémlap felett
             kifeszített ellenálláshuzalt különböző pontokon a
sínhez nyomva
              - az ellenállásérték és a rácsfeszültség
módosításával -, a
             gáztöltésű triódacső által keltett rezgések
frekvenciáját növelte
             vagy csökkentette.
        </seg>
   </tuv>
   <tuv xml:lang="en">
        <seg>
             The Trautonium's pitch was changed by pressing a
resistor wire
             that was fastened above a thin metal strip towards a
rail at
             certain points, thereby altering the resistance value
and the
             field voltage, in order to increase or decrease the
frequencies
             generated by gas-filled triode tubes.
       </seg>
   </tuv>
</tu>
[...]
```

Figure 9. TMX code snippet.

The <tu> (translation unit) contains all of the aligned text, while the <tuv> (translation unit variant) tags contains the actual <seg> (segment) elements plus other information associated with a particular language (such as the <note> element).

TMX has been a success in terms of overall implementations and use, but has fallen short of expectations and is subject to ongoing work. In particular, early users found that when they used TMX to move data between TM tools, not all matches from the original database were found by the new tool. Two issues created this result. First, different TM tools segmented text differently (e.g., one might break text into two segments at a semicolon while another tool does not) and as a result the tool would be searching for the "wrong" thing in its database and miss the existing translation. Second, different tools handle internal markup differently (e.g., some store markup in their database while others recreate the markup using the source document when they process the translations). These problems have resulted in a significant decrease in reusability when moving between tools (some informal tests indicated up to 5 % loss of exact matches), even though TMX has largely succeeded in allowing TM assets to be reused. Current development efforts are focusing on ways to reduce such loss, although it is unlikely that all interoperability issues can be eliminated since they are, to some extent, dependent on the architecture of individual TM tools.

5.6. Segmentation Rules eXchange (SRX)

As a result of the loss in matches when using TMX, LISA developed the *Segmentation Rules eXchange* (SRX) standard, which provides a mechanism for representing how text is segmented. The goal is that other tools can use SRX files to emulate how another tool segmented text in order to find matches that might otherwise be missed. In addition, organizations can provide SRX rules that describe how they segment text so that service providers can coordinate their efforts.

SRX rule sets make use of regular expressions (Friedl 2006) to define patterns where text should or should not be broken into segments. For example, the section from an SRX file shown in Figure 10 specifies that segments should be made at new lines (" n") and after common punctuation followed by white space (" s") but should not be broken after one or more numbers followed by a period at the start of a paragraph (e.g., a section number). They go on to over-ride punctuation-based segmentation breaks after certain common abbreviations in English ("etc.", "Mr.", "U.K."):

```
[...]
<languagerule languagerulename="Default">
   <!-- Common rules for most languages -->
   <rule break="no">
      <beforebreak>^\s*[0-9]+\.</beforebreak>
      <afterbreak>\s</afterbreak>
   </rule>
   <rule break="yes">
      <afterbreak>\n</afterbreak>
   </rule>
   <rule break="yes">
      <beforebreak>[\.\?!]+</beforebreak>
      <afterbreak>\s</afterbreak>
   </rule>
</languagerule>
<languagerule languagerulename="English">
   <!-- Some English abbreviations -->
   <rule break="no">
      <beforebreak>\s[Ee][Tt][Cc]\.</beforebreak>
      <afterbreak>\s[a-z]</afterbreak>
   </rule>
   <rule break="no">
      <beforebreak>\sMr\.</beforebreak>
      <afterbreak>\s</afterbreak>
   </rule>
   <rule break="no">
      <beforebreak>\sU\.K\.</beforebreak>
      <afterbreak>\s</afterbreak>
   </rule>
</languagerule>
[...]
```

Figure 10. Section from an SRX file.

Although this example is simple, the rules can be very complex, utilizing the full ICU (International Components for Unicode) set of regular expressions. They are capable of representing any rule-based text segmentation (although, for obvious reasons, they cannot represent arbitrary segmentation methods that do not correlate to regular patterns in text).

5.7. Term Base eXchange (TBX)

Term Base eXchange (TBX) is an XML-based standard for representing structured terminology data. Jointly published by LISA and ISO (as ISO 30042) this standard encapsulates best practices for representing concept-oriented terminology. The roots of TBX go back into the 1980s in work undertaken as part of the Text-Encoding Initiative (TEI). Although relatively few organizations currently actively manage concept-oriented terminology, it is increasing in importance, and TBX fills a need for a mechanism to exchange and share this sort of data between linguistic applications. The sample in Figure 11 shows how TBX can contain substantial *metadata* (information about terms and other elements) that cannot be represented in simple glossary formats.

For example, this sample not only specifies corresponding English and Hungarian terms but also specifies that they function as nouns, that they are “full forms” (rather than abbreviations), an administrative status (“admitted” for the English, “provisional” for the Hungarian), as well as geographical scope for the

```
[...]
<termEntry id="eid-sample-67">
   <descrip type="subjectField">manufacturing</descrip>
   <descrip type="definition">A value between 0 and 1 used in
...</descrip>
   <langSet xml:lang="en">
      <tig>
         <term id="tid-sample-67-en1">alpha smoothing
factor</term>
         <termNote type="partOfSpeech">noun</termNote>
         <termNote type="termType">fullForm</termNote>
         <termNote type="administrativeStatus">admitted</termNote>
         <termNote type="geographicalUsage">Europe</termNote>
      </tig>
   </langSet>
   <langSet xml:lang="hu">
      <tig>
         <term id="tid-sample-67-hu1">Alfa simítísi tényező</term>
         <termNote type="partOfSpeech">noun</termNote>
         <termNote type="geographicalUsage">Hungary</termNote>
         <termNote
type="administrativeStatus">provisional</termNote>
      </tig>
   </langSet>
</termEntry>
[...]
```

Figure 11. TBX example.

usages of terms. These are just a few of the many possible items of metadata that can be stored in TBX. This information becomes especially important in providing guidance to human translators as well as information that can be used in machine processing scenarios (such as integrating human generated terminology data in machine translation).

6. Conclusion

This overview has shown some of the many issues that are faced when preparing programs and data for use in multilingual environments. The technological issues described are grounded in the processes of globalization, localization and internationalization. For proper globalization, a wide range of technologies (translation memories, terminology management, machine translation etc.) needs to be applied, for usage scenarios like document translation or software localization. The basic technical infrastructure for multilingual data representation and processing needs to be set in areas like character encoding, string processing or language identification. Finally, standardized technologies are the fundamental means to achieve interoperability for multilingual content across a chain of content producers, localization providers, translators and – after all – consumers.

While these issues are complex, they are increasingly easy to deal with as more and more developers have adopted the use of standards such as Unicode and CLDR that facilitate the exchange of multilingual/multicultural data and enable users to interact with computers in their own language. While most users will never need to know the details of how this support is enabled, they all benefit from it when the benefits of computer technology are no longer limited to those who happen to speak English.

Currently the world of multilingual computing is changing drastically. Through the Web and other global information spaces, more and more content in many languages is being produced, sometimes with the producers being also the consumers e.g. in the social media space. People realize that for a truly global information society, it is not enough anymore that content is available “in your own language”. Advance in language technology research and application development will lead to e.g. more and more mature machine translation services like described in this paper, fostering a vision of multilingual computing “across languages”, see (Rehm and Uszkoreit 2011) and the efforts of the MultilingualWeb workshop series http://www.multilingualweb.eu.

For realizing this vision it is important that the models and technologies of globalization, localization and translation described in this paper are regarded as a mandatory basis of research and application development. Multilingual computing in any respect will only lead to sustainable results if this basis is taken into account. It is the hope of the authors that this paper will contribute to awareness and build bridges between linguists doing the fundamental, necessary research and the language related industries.

Notes

1. The “Unicode Technical Reports” available at http://unicode.org/reports/ provide background on various aspects of string processing. For details about ordering and comparison see UTS 10 (Davis and Whistler 2010).
2. For more information about data bases in the context of linguistic research, see (Dipper, Götze and Stede 2006).
3. Details are provided by Sasaki (2010).
4. BCP 47 actually calls the language sub tag a “primary language subtag” and allows for inserting an additional, language related subtag after it: the “extended language subtag”. The details of when to use this mechanism are explained at http://www.w3.org/International/questions/qa-choosing-language-tags.
5. At the time of writing this article, the status and the sustainability of the organization LISA was unclear. Nevertheless, the authors assume that the standards developed within LISA are important for the topic of this article and will therefore refer to “LISA” standards.

References

Anan, Yasuhiro, Hiroyuki Chiba, Junsaburo Edamoto, Richard Ishida, Keiichiro Ishino, Tatsuo Kobayashi, Toshi Kobayashi, Kenzo Onazawa and Felix Sasaki
2009 *Requirements for Japanese Text Layout*. W3C Working Group Note 4 June 2009. Available at http://www.w3.org/TR/2009/NOTE-jlreq-20090604/.

Anderson, Deborah
2003 The Script Encoding Initiative. *Multilingual Computing* 14(6): 34.

Arntz, Reiner, Heribert Picht and Felix Mayer
2009 *Einführung in die Terminologiearbeit: Studien zu Sprache und Technik* [Introduction to Terminology Work: Studies about Language and Techniques]. Hildesheim: Olms.

Clark, James and Steve DeRose
1999 *XML Path Language (XPath)*. W3C Recommendation 16 November 1999. Available at http://www.w3.org/TR/1999/REC-xpath-19991116

Davis, Mark
2010a *Unicode Locale Data Markup Language (LDML). Unicode Technical Standard #35*. Available at http://unicode.org/reports/tr35/

Davis, Mark
2010b *Unicode Normalization Forms*. Unicode Standard Annex #15. Available at http://unicode.org/reports/tr15/

Davis, Mark and Ken Whistler
2010 *Unicode Collation Algorithm. Unicode Technical Standard #10*. Available at http://unicode.org/reports/tr10/

dePalma, Don
2002 *Business Without Borders: A Strategic Guide to Global Marketing*. New York: Wiley & Sons.

Dipper, Stefanie, Michael Götze and Manfred Stede
2006 *Heterogeneity in focus: creating and using linguistic databases*. Interdisciplinary Studies on Information Structure, Vol. 2. Potsdam: Universitätsverlag Potsdam.

Draheim, Andrew and Alison Toon
2005 *Global Content Management and Global Translation Management Systems – An Introduction*. In: Lisa Globalization Insider. Available at http://www.translationdirectory.com/article901.htm

Dürst, Martin, François Yergeau, Richard Ishida, Misha Wolf and Tex Texin
2005 *Character Model for the World Wide Web 1.0: Fundamentals*. W3C Recommendation 15 February 2005. Available at http://www.w3.org/TR/2005/REC-charmod-20050215/

Esselink, Bert
2000 *A Practical Guide to Localization*. Amsterdam: John Benjamins.

Friedl, Jeffrey E. F.
2006 *Mastering Regular Expressions*. Sebastopol, California: O'Reilly.

Harold, Elliote Rusty and W. Scott Means
2004 *XML in a Nutshell, Third Edition*. Sebastopol, California: O'Reilly.

Ishida, Richard
2005 *Localization vs. Internationalization*. W3C online-article available at http://www.w3.org/International/questions/qa-i18n

Ishida, Richard
2008 *An Introduction to Multilingual Web Addresses*. W3C online-article available at http://www.w3.org/International/articles/idn-and-iri/

Koehn, Philipp
2010 *Statistical Machine Translation*. Cambridge: Cambridge University Press.

Koehn, Philipp, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondreij Bojar, Alexandra Constantin and Evan Herbst
2007 *Moses: Open Source Toolkit for Statistical Machine Translation*. In: Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL), demonstration session, Prague, Czech Republic, June 2007.

Korpela, Jukka
2006 *Unicode explained.* Sebastopol, California: O'Reilly.

Ladkin, Peter
2012 Verbal Communication Protocols in Safety-Critical System Operations. In: Alexander Mehler, Laurent Romary and Dafydd Gibbon (eds.): *Handbook of Technical Communication*. Berlin/Boston: De Gruyter.

Lunde, Ken
2009 *CJKV Information Processing*. Sebastopol, California: O'Reilly.

Murata, Makoto
2005 *XML Japanese Profile (Second Edition)*. W3C Member Submission 24 March 2005. Available at http://www.w3.org/Submission/2005/SUBM-japanese-xml-20050324/

Phillips, Addison, Mary Trumble and Felix Sasaki
2008 *Web Services Internationalization (WS-I18N)*. W3C Working Draft 15 April 2008. Available at http://www.w3.org/TR/2008/WD-ws-i18n-20080415/.

Rahtz, Sebastian
2012 Representation of Documents in Technical Communication. In: Alexander Mehler, Laurent Romary and Dafydd Gibbon (eds.): *Handbook of Technical Communication*. Berlin/Boston: De Gruyter.

Rehm, Georg, and Hans Uszkoreit
2011 *Multilingual Europe: A Challenge for Language Technology.* In: MultiLingual, 2011: 51–52.

Reinke, Uwe
2004 *Translation Memories: Systeme – Konzepte – Linguistische Optimierung* [Translation Memories: Systems – Concepts – Linguistic Optimization]. Frankfurt a. M.: Peter Lang.

Reinke, Uwe and Klaus Dirk Schmitz (eds.)
2005 *Einführung in die Softwarelokalisierung* [Introduction to Software Localization]. Tübingen: Narr.

Romary, Laurent
2012 Scholarly Communication. In: Alexander Mehler, Laurent Romary and Dafydd Gibbon (eds.), *Handbook of Technical Communication*. Berlin/Boston: De Gruyter.

Sasaki, Felix
2006 *Language Tags and Locale Identifiers for the World Wide Web*. W3C Working Draft 12 June 2006. Available at http://www.w3.org/TR/2006/WD-ltli-20060612/

Sasaki, Felix
2010 Markup Languages and Internationalization. In: Andreas Witt, and Dieter Metzing (eds.), Linguistic Modeling of Information and Markup Languages. *Text, Speech and Language Technology* 1, vol. 40: 67–80. Dordrecht: Springer.

Sasaki, Felix, and Andreas Witt
2003 *Linguistische Korpora* [Linguistic Corpora]. In: Henning Lobin and Lothar Lemnitzer (eds.), Texttechnologie. Perspektiven und Anwendungen, 195–216. Tübingen: Stauffenburg-Verlag.

Savourel, Yves
2001 *XML Internationalization*. Indianapolis: Sams.

Shklar, Leon
2009 *Web Application Architecture: Principles, Protocols and Practices*. New York: Wiley.

Stührenberg, Maik
2012 Foundations of markup languages. In: Alexander Mehler, Laurent Romary and Dafydd Gibbon (eds.): *Handbook of Technical Communication*. Berlin/Boston: De Gruyter.

Trippel, Thorsten
2012 Controlled language structures in technical communication. In: Alexander Mehler, Laurent Romary and Dafydd Gibbon (eds.): *Handbook of Technical Communication*. Berlin/Boston: De Gruyter.

Unicode Consortium
2011 *The Unicode Standard, Version 6.0*, as updated from time to time by the publication of new versions. (See http://www.unicode.org/unicode/standard/versions for the latest version and additional information on versions of the standard and of the Unicode Character Database).

Walsh, Norman, Alex Milowski, and Henry S. Thompson
2010 *XProc: An XML Pipeline Language*. W3C Recommendation 11 May 2010. Available at http://www.w3.org/TR/2010/REC-xproc-20100511/

Walmsley, Priscilla
2001 *Definite XML Schema*. New Jersey: Prentice Hall.

Walmsley, Priscilla
2007 *XQuery*. Sebastopol, California: O'Reilly.

Yergeau, François, Martin Dürst, Richard Ishida, Addison Philipps, Misha Wolf, and Tex Texin
2005 *Character Model for the World Wide Web 1.0: Normalization*. W3C Working Draft 27 October 2005. Available at http://www.w3.org/TR/2005/WD-charmod-norm-20051027/

11. Scholarly Communication[1]

Laurent Romary

The chapter tackles the role of scholarly publication in the research process (quality, preservation) and looks at the consequences of new information technologies in the organization of the scholarly communication ecology. It will then show how new technologies have had an impact on the scholarly communication process and made it depart from the traditional publishing environment. Developments will address new editorial processes, dissemination of new content and services, as well as the development of publication archives. This last aspect will be covered on all levels (open access, scientific, technical and legal aspects). A view on the possible evolutions of the scientific publishing environment will be provided.

1. Introduction

This chapter provides an overview of the various issues related to scientific information seen both from the point of view of the researcher, with his/her need to have access to and disseminate research results, and the research organisation, which has to define means to optimize the efficiency and visibility of the research it performs. Indeed, this chapter will provide some basic guidelines on the design of a scientific information policy that aims at the benefit of research itself.

Scientific information is considered here in the broad sense of the knowledge that a scientist acquires to carry out his/her research as well as the knowledge he/she produces, and then communicates, in the context of his/her research. This definition indeed reflects the extremely individual nature of scientific information, which relates to the capacity that researcher's results will be further used and quoted. From a wider perspective, scientific information can be defined as the knowledge that circulates within a scientific community as part of the research processes. As a consequence, defining a scientific information policy consists in optimising the scientist's information ecology, by providing the best access to existing knowledge, as well as the brightest dissemination to research productions and results.

Whereas the management of scientific information could be seen as mainly targeted to benefit the progress of science, there are some additional factors, which have to be kept in mind when trying to figure out how to organise the corresponding processes. First, managing scientific information is a kind of second-order activity, which comes, from an economical and organisational point of view, in complement, sometimes even in competition, with the actual support (in staff and equipment) that could be directly provided to research. Second, scientific information, being a tangible sign of scientific activity, is the first objective element that is taken into account when assessing the research activity of an individual, a research group, or a research organisation. As a consequence, scientific information management can seldom be considered as a pure technical activity and has often to be taken at a highly political level, whenever it impacts on the strategy of research performing – but also research funding – organisations.

Because of the intrinsic complexity of the field of scientific management, we will focus along the following sections on specific aspects that may help the reader both to identify the current trend, but also to forge himself an idea on the future evolution of the domain. After characterising scientific information at large, we will thus cover successively the domains of scientific information acquisition, the complementary issue of publication repository and open access, digital edition and data management. We will end up this chapter by providing some insights on the infrastructures that are needed for developing further scientific information services and finally by outlining what we think may be the actual information based research environment of the future.

2. Characterizing scientific information

There are various types of scientific information and trying to cover them all in this chapter would amount to describing all scientific fields in details. Indeed, each speciality uses and generates its own types of information and requires, when designing a scientific information policy, distinguished attention. At this stage, we can identify two main classes of scientific information:

– *Scientific publications*, which are mainly written descriptions of ideas, methods and results seen from the point of view of a specific author or group of authors;
– *Data sets*, which are acquired through the use of experimental or observational settings, as well as data production (simulations) or gathering methods such as surveys in the social sciences.

In this last category, we would integrate all the primary sources that are typically used in the humanities, either in the form of ordinary textual documents, or as specific physical artefacts (ethnographical objects, paintings, encryptions, etc.).

It is important to note here that not all types of scientific information are actually digital. In many domains, artefacts such as archaeological objects or natural components such as seeds or bacteriological samples have to be archived in specific repositories for researchers to observe or use them within experiments.

Even if digital surrogates (images, scans, genomic descriptions) are often used to provide as precise a description as possible of them, they barely replace the actual physical source. In this respect, this chapter only refers to digital materials and their management for scientific purposes.

From the point of view of content, scientific information has some peculiarities that makes it differ slightly from other (possibly published) type of information:

- The relation to the author is in itself peculiar, since the content of a scholarly paper (as well as any commentary or even data) is essentially based on a personal creation, which has to be acknowledged, each time the corresponding content is being used by another party;
- Contrary to works of arts, which would be covered by the preceding point, scientific information is also subject to evaluation by the community of research, either through the peer-review process (Suls and Martin 2009), or simply by the way a research work is cited by peer colleagues;
- Scientific results should be, at least theoretically, available in the long-term (Pilat and Fukasaku 2007), since scientific knowledge can be seen as the cumulative contributions of elementary scientific results.

The combination of the three preceding points makes scientific knowledge build a network of *trusted* contributions, where no single piece can indeed be understood without an explicit reference to surrounding works.

Seen from a systemic perspective, scholarly communication processes have a very special characteristic that makes them differ from most of the types of business we may know of, namely in the relation that the business itself has with regards to the corresponding producers and customers. Indeed, researchers play a triple role in their relation to scientific information:

- As producers, since, by definition, scientific information, and in particular, publications, is the dissemination vector of research results;
- As consumers, since in most domains, existing publications and data sets play the main role in the research process, both as testimony of the current state of art or even as direct input for the establishment of new results;
- As quality controller of the scholarly communication process, by intervening at each stages where evaluation is to take place, be it to peer review a submitted paper or to assess the work of a research entity.

This intricate commitment of scholars in the scientific information workflow is actually made sustainable because *in fine* the scholar himself is rewarded from the fact that the process is actually as efficient as possible. One has to make sure that "good" research results are made available to the community, which in turn will use – and quote – them, which will result in the scholar to gain more fame (as well as institutional recognition), and thus more facilities to carry out his own research. As stated in (Edlin and Rubinfeld 2004:130): "Authors are quite different from more traditional production inputs, since part of their compensation derives from reaching readers."

Finally, it is hardly possible to speak about scientific information without tackling, at least, some basic components of the underlying technologies. Even if our purpose here is mainly to ponder upon the workflows associated to scientific information, we will see along the coming pages the strategic orientations that must be taken in order to provide science with trustful environment. This notion of trust (applicable to scientists, institutions as well as the public at large) relies on the capacity of providing reliable digital object management environments, as well as putting these within a network of interoperable components, allowing a researcher to seamlessly access and use scientific information in the various forms that we have identified so far. As we shall see such technological environment should be pragmatically designed, so that they serve a wide community of users, from big research consortia down to individual researchers.

3. Acquiring scientific publication

The acquisition of scholarly content through subscriptions has been for many years the main duty of research libraries, together with the acquisition of scientific books. The journal market, seen from the point of view of the library, can be characterised by a high level of stability, due basically to two main factors: a) the attachment of scholars to their favourite information sources, in a context where conservatism is the main drive for publication and reading, and b) the natural tendency not to break paper collections within the library so that one can characterise the library through its portfolio of "journals". This overall stability has been further reflected in the subscription contracts negotiated between the libraries and the publishers (or subscription agents such as SWETS or EBSCO): from one year to another the evolution pressure on the actual portfolio is essentially based on the explicit request for new titles on the part of the researchers and the identification of discarded, or less consulted, journals on the part of the library.

This intrinsic conservatism of the subscription ecology has been one of the major factors that led to the so-called *serial crisis* in the eighties. This crisis is the result of an unfortunate dynamics[2] by which publishers, taking benefit from their exclusive situation, raised the journal prices at a pace largely exceeding inflation, and libraries, facing their own financial constraints and the impossibility to cut down their core journal portfolios, focused their subscription needs on major publishers. After a few years, this crisis led, on the part of the publishers, to a quick concentration of production means, and, on the part of the libraries, to the impossibility to actually fulfil the needs of their academic communities.

In the following section, we will make an attempt at characterizing the post-serial crisis era, starting from an analysis of the notion of big deal, and exploring some possible trends to facilitate a quick and favourable evolution of the commercial scientific publication landscape.

3.1. A transitional model – Big Deals

In the nineties, the contractual setting for subscriptions changed dramatically, following a joint necessity for publishers to optimize negotiation efforts, and, for academic libraries, to try to compensate for the high increase in price that had characterized the preceding period. This has led to the so-called "Big Deals", which have been characterized by Edlin and Rubinfeld (2004) as follows:

"in a typical Big Deal contract a library enters into a long-term arrangement to get access to a large electronic library of journals at a substantial discount in exchange for a promise not to cut print subscriptions (the prices of which will increase over time)"

As a matter of fact, such Big Deals have quickly evolved in some countries from a rather localized setting, limited to a University for instance, to cover clusters of academic institutions that started to jointly negotiate subscription agreements with publishers. After several years of practices, we would like to provide the reader with a distanced analysis of the lessons to be learnt from such big deals and see how much they may contribute to an evolution of the publishing landscape.

As a whole, Big Deals have had quite a few positive effects on the library landscape. Without analysing the actual financial issues in details, we can analyse the benefits of Big Deals by seeing how much they contributed to provide more maturity and awareness about scientific information within academic institutions:

– Big Deals participated in the creation of real community of practice among research and education libraries, since librarians had a real opportunity to compare their needs, their relations to the scholarly communities, and also their budget capacities;
– By establishing a point of focus on the relations between a given publisher and one or several institutions, they raised the political implication of academic management, but also of scientists themselves, in the decision making process;
– They speeded up the process of moving to electronic content, since most Big Deals include a wide access to the corresponding publisher's catalogue, thus bringing up the necessity of identifying the actual needs related to online access (such as long-term accessibility and archiving, see below), and the underlying infrastructures (access portals) to be put together;

- They forced academic libraries to think in-depth on the value for money that subscription schemes actually bring to institutions, in particular considering such factors as usage (hardly considered in the paper world), or budget consolidation when Big Deals would directly impact on the capacity to establish subscription contracts with smaller publishers not involved in Big Deal negotiations.

Still, these rather organisational benefits should not hide the core problems that Big Deals have brought with them, and which can be analysed along two main lines:

- Big Deals have introduced a highly conservative view on the subscription models between libraries and publishers. Indeed, the journal-based model, coming from the printed world, and the corresponding constraint imposed by publishers that turnovers should be preserved on the basis of the existing print subscription, have locked libraries into a system which give them no margin of manoeuvre in the management of library costs in the future;
- Big Deals have also showed the weaknesses of academics in dealing with strategic negotiations and in particular the difficulty for libraries to be strong at the negotiation table. Recent examples of clashes during a Big Deal negotiation should not hide the fact that many contracts have been renewed without much improvement from one round of negotiation to another.

Considering all factors, and in particular the risk that Big Deals may contribute to fossilize the subscription landscape, it is important to see such models as transient ones and identify possible evolutions that may prevent a long-term dependency to publisher's requirements. First, libraries should not be left alone in negotiating large-scale and long-term agreements with publishers. Whereas they are essential in providing objective information about the existing needs and usage, negotiation teams should always integrate academic managers and scholars with in-depth knowledge about scientific information processes. Second, prior to any negotiation, global objectives should be set, at the benefit of scholarly work, not only in terms of financial benefits, but also in terms of additional benefits that the community may gain from a new contract. We will see in the coming sections how much in particular, the notions of archival and open access may impact on the actual subscription policy of an academic institution.

3.2. Towards new contractual schemes

The difficulties related to the current subscription system could easily be seen as a deadlock for most research organisations since it basically hampers any possibility to define a real strategy in the domain of scientific information at large. Indeed, by freezing part of the budget on fixed expenses, it prevents these organisations from both adapting the corresponding budget to their economical situations or their priorities, but also is a major hindrance to the design of new ways for scientific information to be seamlessly exchanged among the research communities, as would be needed for better and quicker scientific progress.

Still, even if we identify that there is a need for an in-depth evolution of the publishing environment in the scholarly world, we also ascertain that it can only be implemented through the exploration of a variety of new deals between research organisations and publishers. With this orientation in mind, we present in the following sections some possible actions that could be pursued to contribute to such an evolution. All of these have been already implemented and validated within organisations and we will try to draw prospective conclusions from the assessment that we will make for each of them.

3.3. Open access publishing and budget centralisation

The first move we would like to address here actually comes from the publishing sector, which has experimented in the recent years ways of offering open access content (Velterop 2003) on the basis of a payment that would not result from a subscription, but that would take place on the author's side (hence the expression of *author-pays* model). In this domain, one needs to make a clear distinction between *opt-in schemes* and *native open access journals*.

In the opt-in open access schemes, publishers (e.g. the Open Choice scheme by Springer) offer the possibility for an author to finance the full accessibility on the publisher's web site of the final version of his paper at the time of publication. Such schemes have several disadvantages. It creates a burden on the authors, who have to decide on (and finance) the online publication of their papers, one by one. It implements a double payment system in institutions which do have subscriptions to the corresponding journals, and as a consequence, prevent the institutions from getting an accurate overview of the budget they dedicate to publication material. All in all, we strongly recommend not to support such schemes and to clearly inform the scholars about their danger.

True open access journals (such as the PLoS[3] portfolio or most of the titles from BMC) offer the author-pays model whereby authors must systematically pay publication fees to the corresponding publisher, but conversely no subscription is required to get access to the content. As opposed to opt-in schemes, there is here no risk of double payment and the situation is by far clearer for institutions that may centralise a specific budget made widely available to authors (thus reducing the administrative overhead). Still, such schemes (but the reverse arguments exist for the standard subscription scheme) present the dangers that it may become even more difficult to publish for institutions or countries with reduced means, and also that there may be free-riders, i.e. entities (in particular private firms) which benefit from the publication material, without contributing at all to the corresponding cost. Still, the capacity that open access journals offer to centralize the corresponding budget and thus to monitor the actual evolutions within and across research institutions worldwide encourages us to advocate the integration of open access journals within the scientific policies of academic institutions at large.

3.4. Making publishers' offer fit academic needs

Not only has the move from the traditional printed journal model to online delivery changed the capacity of scientists to quickly access information, it has also deeply modified the perception that academic institutions had of the management processes one should deploy for such information. This has resulted in identifying new requirements – and indeed new services – that the publishers should fulfil in conjunction with what would normally be associated with a subscription contract.

This section explores how such requirements can be contemplated, and possibly implemented, from the point of view of the content proper, i.e. how much an institution should claim to receive precise information from publishers, in order to go towards a better and more sovereign management of its own scientific information. We will indeed explore these issues in two stages, related to a) the acquisition and management of reliable meta-data and b) the various archiving levels that one may demand in the context of subscription schemes.

It is quite straightforward to understand why it is of strategic importance for an institution to have a good overview of its scientific production and consequently why managing reliable meta-data about publications is considered as a very sensitive issue. As it provides an insight on the quantity and quality of the actual publication activity of individuals as well as of institutions, such information is at the core of reporting activities, of researchers' assessment, as well as for any strategic planning endeavour. It is also of paramount importance that such metadata be both precise and accurate, since it should allow gaining insights into such a variety of publication features as the actual domains where research is produced, or the collaboration schemes that an organisation may have with other institutions or other countries for instance.

What is meant here by precise and accurate is the capacity for a database, and consequently for an interchange format, to provide metadata information where all elementary and meaningful pieces correspond to a specific field, associated with a precise semantic and mappable onto the most relevant international standards. This has an impact at three main levels of a bibliographical representation:

- In the description of publication information where article information (title, volume, issue, pagination, DOI, publication date) and journal information (title, ISSN) should allow for both a precise identification but also accurate description of the article;
- In the representation of author data, where not only names should be finely represented (all name components should have their own fields), but also affiliations should be dealt with as precisely as possible;
- In the production of basic content related information such as keywords, domain classifications or abstracts.

Whereas the first point is rather straightforward and usually dealt with quite well in various publication platforms, the second one is often underestimated. Still, precise author data and in particular affiliation represent a key aspect for the further exploitation of publication data, in particular for the study of collaboration patterns (Subramanyam 1983). For instance, being able to sort out papers by institutions, or to analyse the geographical patterns of co-publication is part of the patterns that one wants to be able to identify out of a bibliographical database.

In the ideal case, there would be an excellent opportunity to compile such meta-data on the basis of the information available from publishers (at least for journal papers). However, as demonstrated in particular by the technical work within the European PEER project[4], scientific publishers have no coherent framework for such data and, in fact, some have difficulty to provide precise information concerning authors and their affiliation. It occurred also that publication repositories are not that in a better shape and the PEER project is the opportunity to demonstrate how much a standard-based approach (based on the customisation possibilities of the TEI[5]) could lead to a higher degree of interoperability (Ide and Romary 2004) between platforms.

If we were to make a move towards higher interoperability across platforms, the scenario contemplated here would be that any publication repository, whether public or private that contains precise and reliable chunks of information should be in the position to deliver them to complement the information that another repository would require. Whereas such a scenario may be extremely complex to implement across any kind of platform, in particular because of the heterogeneity of formats we have identified above, it is possible to contemplate the maintenance of a clearinghouse[6], which would at least serve the network of public publication repositories.

We should note here that the precision required for exchanging meta-data between publication repositories in such scenarios as the ones addressed here has nothing to do with the very shallow formats required by current harvesting processes, which, for instance, are part of protocol such as OAI-PMH (Lagoze and Van de Sompel 2001). Indeed, we anticipate an essential evolution of publication repositories, which, seen as trusted sources of information, will be able to provide, at any level of granularity, all the description attached to their digital content (see also Groth et al. 2010).

Beyond meta-data, it is also important to consider how much an institution can demand on the availability of the full publication content and identify how much we could globally make progress in this domain. The issue here is quite simple to state: the move from paper content to digital one has changed the relation between the customers and the content in two complementary ways:

– The move to digital content has deprived step-by-step academic libraries from their role of reference archive for scholarly content (Greenstein 2000);
– Publishers have progressively taken up the management of the archival dimension of digital content, claiming that this content is always “accessible” from their own platforms, which would offer better technical facilities as well as all guaranties for a sustainable availability (Kling and McKim 1999).

These changes have created a strong dependency towards the commercial sector, with academic institutions only lately identifying the need to define a strategy in this domain. What indeed would we have if such and such publisher would disappear, or if a major conflict arises between the academic sector and the private one, where the former would lack authority towards the latter for sake of this dependency? Even if not central such a feeling has contributed to the open access movement, when scholars (and their institutions) identified that in the digital age content could, and indeed should, be pooled together in a coordinated manner.

In order to go towards a better management of published material by research institutions themselves, there is a need to act explicitly during the negotiation stage of subscription contracts. There are indeed three levels of requirements that could be identified in the domain of long-term access to publication:

– If the data is to be solely archived by the publisher, one should minimally require a *perpetual access right* to the material subscribed at a given period. Concretely, this is to guaranty that even if a contract with the publisher is interrupted, all journal content that has been paid for (new issues and backfiles) will remain accessible for the population that had benefitted from the initial agreement;
– To guaranty a better independence from the publisher’s services and secure the availability of subscribed material, one can introduce a clause allowing the institution (or the corresponding consortium) to host the content and deliver it through its own servers. Further archival copies should of course be needed to ensure continuous and long-term accessibility;
– Finally, an institution can require that all the subscribed material that also corresponds to content authored by its researchers[7] could also be uploaded in a designated publication repository, in keeping with the institution’s open access policy.

These three levels actually impact on the general copyright scheme that should be associated to scholarly content. While we argue that the private sector has an essential role in providing core services, such as certification, in the scientific publishing workflow, new deals have to be set so that such services are never correlated with any kind of exclusive copyright transfer.

As can be understood, we try to argue here that academic institutions should strive towards the establishment of scientific information repositories where publication meta-data and content are pooled together within a sustainable and reliable environment. Such an orientation bears a quite natural technical side, but it should also be taken very seriously as a component of ongoing discussions with the private sector. At the highest level, and in particular in the context of national licence program, such evolutions have to be discussed with the same level of attention as budget or coverage issues.

3.5. General perspective

As a whole, we see there is a need for a real strategic view in the way we are to handle future commercial deals with the publishing sector at large. On the one hand, we need to identify means to make our budget more depending on the real need of research organisation rather than see it be determined by external interests. On the other hand, we need to identify the actual services that we require from publishers and negotiate with them the best value for money, considering also the services that we do not actually want from them. As we shall see from the following developments on research repositories, research performing organisations may want to achieve some of the functions that were so far externalized to the private sector, when these reflect a strategic interest for them.

4. Scientific information and open access – the role of publication repositories

4.1. Defining open access again?

The open access movement has gained in the recent years enough momentum and fame that one could believe that there is hardly any need to enter too much a discussion about its scope and objectives[8]. Epitomized by the Berlin declaration[9], the open access ideas reflect a tension between two opposite sets of forces:

- The necessity for the publishing business to keep its revenue, thus leading to ever-increasing subscription budgets for the academic institutions;
- The feeling that dissemination of scientific information in the digital age should be straightforward (i.e. without intermediaries) and fundamentally cheap.

With a sense that there is no clear way out of this situation, the open access movement is an echo to the feeling that academics should take action and devise their own means of disseminating their research papers. The so-called *green way to open access* is thus focussing on providing technical platforms[10] and policy settings[11], for the self-archiving and free dissemination of scholarly papers. In a way, it is complementary, or even contributes, to the attempt outlined in the first part of our paper to define new collaboration schemes between the academic institutions and the commercial publishers.

Still, the green open access movement can be characterized by a quite narrow view on scholarly publishing and the angelic idea that researchers may just join the self-deposit movement by conviction or by being mandated. The situation is *de facto* more complex, and instead of making here a theoretical or comparative analysis[12] of the publication repository landscape, we would like to illustrate the various views that may exist upon a publication repository, by presenting the specific case of the *Hyper Archive en Ligne* (HAL) repository, which, across the years has become the reference environment for the French academic landscape.

4.2. A reference case – HAL

In the mid-nineties, *Hyper Archive en Ligne* (HAL) was put together by physicists who wanted to implement a mirror archive to the already long-standing *ArXiv* (Ginsparg 1994). At that time, the main drive was to benefit from an independent platform that could have its own editorial policy, be independent from possible access problems to the United States and also be able to develop its own functionalities. With the support of the CNRS[13], a service unit[14] was put together and a first implementation made operational within several months. The spirit of this first environment was mainly targeted towards pre-prints (stage 1), to favour an early dissemination of research results.

Step by step, the French physicist community, as well as mathematicians, got used to deposit their papers in HAL and in the mid-nineties, the repository gained momentum when the CNRS decided to promote it within other disciplines and adopt it as its main source of information to published material by its researchers[15]. At the same period, several research organisations in France, among them INRIA and INSERM[16], decided to adopt HAL as a way to make concrete their adhesion to the Berlin declaration on Open Access[17]. This extension process led to a national agreement, signed in 2007, of most of the French research institutions, together with the national conference of French Universities to sign an agreement to work jointly on the further development and support of HAL.

As analysed in Armbruster and Romary (2010), the evolution quickly outlined above has set HAL as a publication repository that can hardly be reduced to the usual narrow concept of an institutional repository.

- It offers services to multiple research performing organisations such as CNRS, INSERM or INRIA, as well as universities. Each institution may have its specific portal through which their researchers may deposit and where all corresponding productions are visible. Still, any content is part of a single repository, thus allowing searches to be extended to the whole content;
- It bears a strong multidisciplinary mood, either because of the disciplinary nature of some of the supporting institutions, or because specific communities have set their own portal, but also more generally, because it matches the cultures of various scholarly communities by both considering pre-prints or published material as its core input;
- It has reached a cross-*institutional* recognition which makes it be the default repository for Universities or research funding organisations[18] aiming at indentifying their research production.

All in all, and things being seen from the wider perspective of a national scientific information policy, HAL is providing a high level service, well integrated in the institutional landscape, for a very low budget through its highly centralized technical nature. Still, institutional recognition is, as we know, useless, if researchers themselves are not convinced that using the platform may bring some added value to their own work. It is thus worth observing why in such a community as computer science, and in particular within INRIA, HAL has gained such a fame and become part of the scientists' daily practice.

4.3. Publication repositories – the researcher's view

One immediate feedback that is received from colleagues using the HAL-INRIA portal on a regular basis is the high visibility that the corresponding papers receive immediately from most international search engines (hear "Google"). Indeed the centralized management of the archive has made it feasible for years to maintain good working relationship with the corresponding technical teams and to tune the software interfaces to optimize the visibility. From the point of view of the researcher, this is observed through the immediate high retrievability of a paper, as soon as relevant keywords for the corresponding field are typed in. This visibility can also be traced on the ranking of publication repositories produced by Webometrics (Aguillo et al. 2010; cf. also Björneborn and Peter Ingwersen 2004), which actually shows HAL and HAL INRIA as performing respectively on the platform and the institutional ranking.

The second important aspect, which is often mixed up with the traditional notion of institutional repository (cf. Armbruster and Romary 2010) is the specific institutional setting. Indeed HAL provides specific institutions (or communities) with dedicated portals which allow them to deal with the publication archives in their own way, by adapting the graphical charter, adopting their own editorial policy for content, or building up their additional tools (e.g. connection to in-house reporting mechanisms). Such views allow both the institution and the researchers to actually feel at home within the HAL platform, while benefiting from all generic features and evolutions that are developed globally.

Last but not least, the wide acceptance of the HAL Platform within INRIA can also be a great deal attributed to the important editorial support provided by the library network, which systematically reviews all entries after authors make a deposit, or at times even helps authors to make batch deposits when necessary. By completing missing meta-data, correcting actual descriptions (bibliographical descriptions, spell-checking of abstracts), by directly interacting with authors in border-line cases (software documentation, management of anteriority) and finally ensuring that the most appropriate tools and interface are being put higher-up on the developers' agenda, the librarians convey a feeling of stability that encourages even more the researchers to see HAL-INRIA as a trustful archive.

4.4. Perspective: intelligence in the platform

The experience we can gain from observing such a platform as HAL let us take a real distance with some of the most polemic (and at time conservative views[19]) on open access and anticipate on how much publication repositories can be made a tool at the service of science at large. Without anticipating on the wider vision we will outline at the end of the paper, we can try to make a short term project on the way we see publication archives provide even better services to researchers.

Indeed, and taking up the notion of ecology of publication repositories from (Romary and Armbruster 2010), we see that if we want to better integrate publication archives in the researcher's information ecology, we need to provide him with a variety of seamless services that will facilitate the deposit and further use of research articles. Such services would actually occur at various levels:

– At the deposit stage of a paper, one should relieve the depositor from typing in information that could be retrieved from other sources, or the paper itself. Typically, a good management of authors and associated affiliations, in particular in the context of usual co-authors is essential. In complement, automatic information extraction techniques[20] from textual content should allow to pre-fill most of the meta-data information required for a proper management of content in publication repositories;
– We should work towards providing researchers with more capacities at managing their *workspaces* within publication repositories. They should be allowed to keep drafts unpublished, add collections of graphics or images, or even additional documents such as slides, posters or videos, which may come as natural complements to the paper itself;

– At the dissemination level, it is essential to provide efficient tools for researchers to compile information from the repository, either as publication lists when preparing a paper or an institutional report, or to generate his web page or that of his research group.

In the current publication repository picture (cf. Armbruster and Romary 2010), such features are occasionally implemented, but there is a need for a concerted development that only a more centralized view on repositories may allow us to seize. Furthermore, institutions should clearly state how much they consider the issue of pooling up publication data by redirecting all information gathering needs towards what is actually available there.

5. Dealing with research data and primary sources

In the continuity of the infrastructural work carried out for publication, the academic organisations are more and more turned into considering the importance of dealing with research data as part of their scientific information management duties. Indeed, as exemplified by the report issued to the European commission in October 2010 (Wood 2010), the management of data has become a key issue for the definition of new research funding programs and in particular for the establishment of digital research infrastructures. One of the key issues related to scientific data infrastructures is that of trust, in the sense that researchers may deposit and re-use scientific data with full confidence that these will be properly curated and preserved in the long run.

5.1. Characterising research data

The notion of research data is by far more fluid and heterogeneous than that of publication and shall deserve a specific attention. Contrary to publications, research data cannot be apprehended as a set of clearly identified objects. Data occur at various stage of the research process and may usually be seen at various levels of granularity, from elementary samples to large collections (or corpora). Quality is also an issue where part of the research activity is indeed dedicated to the selection and enrichment of primary data acquired from some equipment, or, as is the case in the humanities (Romary 2011), directly selected from various cultural heritage sources.

Scientific data bear many common features with publications. In essence, they require to be as neatly associated to the originated researcher or institution, as a warrant for the trustfulness of the content, but also in order to allow an adequate citation of the work. More generally, scientific data have to be, even more than publications, associated with precise metadata (in the same way as what we have for publications with bibliographical data). Indeed, whereas one can contemplate the idea of retrieving the source of a paper while looking at it, there is hardly any chance that a collection of numbers may tell anything of its format and origin. The fear with scientific data is that, by not taking the adequate time to document it, researchers may create data cemeteries, which may prevent any further reuse.

From a strategic point of view such issues cannot be dealt with without academic institutions having a real *data curation* strategy, which complements the issues of *data description* outlined above with that of *data selection*. As a matter of fact, it becomes more and more impossible for some scientific fields to keep all the data they produce, since these would go by far beyond the IT capacities as we know them today. To take the most prominent examples, the new Large Hadron Collider at CERN[21] is planned to generate several peta-bytes per year, which in turn forces their technical infrastructure to be organized as a cascading network of computer nodes, where just filtered-out data is percolated down to the individual researchers.

5.2. Pooling data together – the core of the scientific data business

Beyond the local management of research data at the production locus proper, it is important to keep in mind that a major challenge for the further progress of research in many fields is the capacity for research entities to be able to pool together data assets in order not only to avoid duplication of work, but even more to combine complementary evidence to form a wide-coverage informational basis for a given field. This trend has been initiated several decades ago by domains such as astronomy[22], which actually endowed themselves with various means to create networked or even joint databases of observational data (images, frequency range observations), but also of stellar objects and associated bibliographical references. To extend such trend in all fields of science, research communities, but also research organisations and governmental bodies should act in basically three directions:

- At the political level, there should be a strong incentive towards researchers so that they actually make the wide dissemination of data a standard component of their work. Such incentives may take the form of financial support, but also of effective recognition mechanisms for those who take actively part to this endeavour;
- From a legal point of view, communities should identify the least damaging licensing frameworks so that scientific results are of course systematically attributed to their producers, but also no hindrance is made to the combination of information which would have contradictory legal status[23];
- Finally, the compilation of data collections requires a strong technical coherence, to ensure in particular that data assets are made interoperable for

joint queries and combined visualization or processing. This requires that scientific communities be involved in standardisation activities[24] to produce the appropriate requirements and guidelines for the representation of the data relevant for their field of activity.

As can be anticipated, the last point cannot be seen as a short-term endeavour. Each field has to identify its own requirements, link these with the existing international standards and make sure that the corresponding standards will evolve at the same pace as scientific discovery proper. As demonstrated in (Romary 2011) for the general area of the humanities, such standardisation efforts (Gray et al. 2005) relate to the capacity to provide conceptual models for the corresponding data, so that these models are also independent of the contingencies of IT facilities at a given time.

6. Towards a "scholarly workbench"[25] – a vision for the future of scientific information

In the previous sections, we scrutinized the ongoing evolutions of the scientific information processes, and delineated step by step a vision where institutions and scholars would take up the lead in making these changes facilitate the scholarly work. We saw in particular how much the categories of publication and data may deserve in the long run a similar treatment in a context where the separation between the two becomes less and less marked. In this context, we would like to end up our analysis by describing the way we would dream the researcher's workspace of the future, as a target object that we would try to reach when making strategic choices within our organisations. This *virtual research space* (or eSciSpace) would indeed apply for all types of scholarly activities from "hard" physical sciences to the study of primary sources in the humanities.

In the virtual research space scenario, we define a *research asset* as any piece of information that may be manipulated and further made public by a scholar in his working space. These may be pieces of text, data collections of all kinds, annotations to any research asset, or virtual *research folders* grouping together research assets in the course of the scholar's activity.

In his research workspace, the scholar manages all the stages of his scientific information activity, gathering initial evidence by importing data from external repositories, in the form of "publications" or as extractions of existing observations or documents provided by other scholars. Building up from this evidence, he compiles and organises his thoughts in the form of drafts or annotations directly linking to the other documents in his workspace, applying specific software to compute new features from his observations, and organize his thoughts by grouping meaningful sets of information in dedicated research folders.

6.1. Core services of the virtual research space

The virtual workspace outlined above resides on the availability of some basic services, which, not necessarily being specific to the scientific information domain, are mandatory to implement the target scenarios we anticipate. These essential services are the following ones:

- The researcher is automatically identified when connecting to his/her workspace, on the basis of a global authentication process. As a consequence, the various parameters of his/her research environment (affiliations, ongoing funded projects, co-workers) are available, so that the appropriate information can be attached easily to any research asset he/she manipulates;
- His/her workspace allows him/her to create research folders of any kind where he/she can compile various objects (notes, data sets, links to external sources, as well as related persons and events as, e.g., meetings[26]);
- Unique identifiers are provided for all objects in the workspace, which can serve as a stable reference for any further quotation or reuse;
- A precise versioning system is at hand, which allows him to trace the evolution of any research assets and publish any version according to his will.
- Last but not least, the scholar has full liability to select any of his research assets and give access rights to other individuals, groups of individuals or the public at large.

6.2. Community review and certification

In the perspective we want to defend here, the research workspace is also the locus where the validation of the research results takes place, opening up a whole range of new possibilities for a scholar to gain approval or recognition from his peers. Indeed, moving away from the highly focused peer-review process that we currently have for publications, the management of a research workspace combining written documents and data-sets, together with the capacity to provide various levels of visibility to any combination of these objects, allows one to anticipate a much broader community based validation of scholarly content. In the various possibilities we will describe below, the scholar always keeps the initiative to launch a certification process, but the capacity on his virtual workspace allows him actually to receive the level of feedback he actually wants.

The baseline case is not that far from the current peer-review system, as we know it. It is based on the assumption that there exist certification entities associated to a specific topic, and which are articulated around a scientific committee. In this scenario, the scholar gathers research assets in a dedicated research folder and, when he considers the corresponding result to be worth it, he simply makes it accessible to the certification entity, asking for “peer review”.

If the evaluation is positive, the research asset receives a *review certificate*, which can be associated to any data feed related to it. Through this process, and according to the editorial policy – as well as the scientific recognition – of the certification entity, the research asset becomes a "peer-reviewed" publication in the literal meaning of the term.

On the opposite side of the spectrum, the scholar may not want to have an official certification process, but would like to receive a certain level of recognition of his results from a community. To this end, he can publicize a research asset and issue a targeted call for comments, which may range from selected colleagues to a wider scientific community. Such calls would transit through scholarly feeds or social networks to which the researcher subscribed. Feedbacks to the call take the form of *commentaries*, which may be posted in the commentators' research space and linked to the research asset. The collection of commentaries can in turn be used to derive a new version of his work or to assess the recognition of his scientific contribution within the community.

This generalized process of certification incorporates various attempts at exploring new types of peer review in various research communities. Initiatives like open peer review from the EGU (Pöschl 2004), or the attempt at creating data journals[27] for assessing the added value related to the compilation of reference data sets would naturally fit into this picture. Besides, the basically open nature of research workspace would facilitate the compilation of new types of evaluation metrics. For instance, usage statistics, coupled with information concerning the actual scholar, community or institution having used the research asset, would bring usage profiles by far more precise than any kind of citation metrics we have nowadays.

6.3. The library continuum

The problem addressed with this vision of a virtual research workspace is to help the scientist manage the documents he uses and produces in his research, as well as making him feel like he is in charge of his own *virtual research library*, where he can easily retrieve documents, but also disseminate them in isolation or as collections to students, colleagues, or the wide public. Indeed such a vision decentralizes completely the notion of a research library, which would move from a central physical place to a delocalized space in the "cloud" (Aschenbrenner et al. 2009). Still, we may not expect that scholars would take up all the necessary work related to the curation of their research assets, and comprising such a variety of tasks as technical integrity or adequate meta-data description. They will have to be supported, and this is part of the infrastructure by new types of documentary experts (*eLibrarians*, or more commonly digital curators) that will accompany them in their information based research activity.

7. Infrastructures for scientific information

As a word of conclusion, we should situate the conditions under which the evolutions, as well as dreams, outlined in this paper could actually be implemented. Indeed, most of the developments contemplated here, whether related to publication or research data, and comprising what we did not develop here in domains such as bibliometry or digital editions, require long-standing involvements of the corresponding institutions in the domain of scientific information management. Over the years, such an involvement, together with the corresponding investments, has taken the form of real *eInfrastructures* at the service of science. Such infrastructures have shaped differently depending on the country or the academic setting ranging from high-level coordination directorate (at CNRS) or as a continuation of a strong library setting (as at the University of Göttingen, see (Lossau 2004)), it has also often been embodied as autonomous units dedicated to the management of scientific information (as in the Max-Planck Society with the Max Planck Digital Library[28]). All these service or decision units have shown from our own experience that their efficiency and capacity to contribute to the strong dynamics of the field rely on a series of essential factors:

– *Scientific digital libraries,* since they already have a name (Berman et al. 2003), should be close enough to the research communities they have to serve in order to understand both their needs but also their practices in the domain of scientific information; why indeed offer for instance information portals to scholars who would prefer using search engines to find their information sources?
– All the same, they should be constructed around a core of highly competent staff that masters such a wide range of skills as budget negotiation, IT development and knowledge representation. Scholars alone may not always have the adequate sense of the complexity related to certain decisions;
– Even more importantly, the scientific digital library should be closely related to the management of the corresponding academic institution, since from our experience, most strategic decisions in the domain of scientific information directly impact on the research environment of the institution;
– Finally, scientific digital libraries should cover the various fields of scientific information management. If their scope is limited to such a narrow domain as the management of journal subscriptions, they will not be in the position of identifying the adequate tactic moves across the various fields of scientific information.

Considering the current scientific information landscape, it seems that we are not far from having such scientific digital libraries at hand. Various attempts in this direction are taking place and at the same time, we better see the role and evolution of such elementary building block as publication repositories. The work ahead of us is now to bring all this together in a more coherent way.

Notes

1. This paper has strongly benefit from close interactions with Norbert Lossau, director of the State and University Library of Göttingen.
2. McCabe (2004) also demonstrates that mergers among publishers have also been a determining factor.
3. Cf. Brown et al. (2003)
4. The meta-data available from 12 publishers, comprising some of the major companies in the field, have been mapped onto one single standardized structure, in order to be pushed to a group of publication repositories in the context of a large-scale (green) open access experiment.
5. *Text Encoding Initiative*, www.tei-c.org
6. The PEER Depot, implemented in the context of the PEER project is an excellent example of such a clearing house.
7. Depending on the agreement, the coverage can be limited to corresponding authors or (like was the case of the agreement between the Max Planck Society and Springer in 2008) extended to any co-author of the publication.
8. See (Davis 2009) for an overview of the doxa on the subject.
9. http://oa.mpg.de/lang/en-uk/berlin-prozess/berliner-erklarung/
10. usually referred to as institutional repositories (cf. Lynch 2003, and the wider distance analysis in Romary and Armbruster 2010)
11. in the form of deposit mandates (see Sale 2007)
12. see Armbruster and Romary (2010) for such a global comparison.
13. Centre National de la Recherche Scientifique (www.cnrs.fr)
14. *Centre pour la Communication Scientifique Directe* (CCSD) – http://www.ccsd.cnrs.fr/
15. cf. http://www2.cnrs.fr/journal/2546.htm
16. *Institut National de la Santé et de la Recherche Médicale* (www.inserm.fr)
17. http://oa.mpg.de/lang/en-uk/berlin-prozess/berliner-erklarung/
18. See the news (in French) where the French *Agence Nationale de la Recherche* announces that it requires that all publications related to a project it funds should be deposited in HAL: http://www.agence-nationale-recherche.fr/magazine/actualites/detail/?tx_ttnews%5Btt_news%5D=159
19. See for instance Harnad (2011).
20. See in particular the importance of machine learning techniques in this respect (Lopez 2009)
21. http://public.web.cern.ch/public/en/lhc/lhc-en.html
22. See Heck (2003) for an overview, but also Heintz and Jaschek (1982) to get an idea of the visionary mood of the astronomers in the early 80's.
23. When applicable, a basic creative commons CC-BY (http://creativecommons.org/licenses/by/3.0/), where the sole constraint is actually to guaranty the attribution of the work to its source may be considered as a baseline for such licences. Any additional constraints may then become a stumbling stone for the further pooling of information. For a wider discussion on this see Wilbanks and Boyle 2006.
24. See Ochsenbein et al. 2005 for an exemplary community endeavour.
25. This expression appeared in the initial work plan of the eSciDoc project in 2005 (www.escidoc.org/) (Dreyer et al. 2007).

26. Part of the necessary infrastructure would resemble services such as those offered by Calenda (http://calenda.revues.org/) the French diary of events in the humanities (Dacos et al. 2006).
27. In a context where many data journals initiative are taking shape, we can only point to the quite visionary initiative from JISC: *Overlay Journal Infrastructure for Meteorological Sciences* (OJIMS)
28. www.mpdl.mpg.de

References

Aguillo, Isidro F., Jose L. Ortega, Mario Fernandez and Ana M. Utrilla
2010 Indicators for a webometric ranking of open access repositories *Scientometrics* 82:477–486 doi 10.1007/s11192–010–0183-y

Armbruster, Chris and Laurent Romary
2010 Comparing Repository Types – Challenges and barriers for subject-based repositories, research repositories, national repository systems and institutional repositories in serving scholarly communication. *International Journal of Digital Library Systems* 1(4): 61–73. Available at: http://hal.inria.fr/inria-00465197

Aschenbrenner, Andreas, Flavia Donno and S. Senka Drobac
2009 Infrastructure for interactivity – decoupled systems on the loose. *3rd IEEE International Conference on Digital Ecosystems and Technologies. DEST '09*: 643–648

Berman Fran, Geoffrey Fox, Tony Hey and Anne Trefethen
2003 The Data Deluge: An e-Science Perspective In: F. Berman, G. Fox and T. Hey (eds.)*, Grid Computing: Making the Global Infrastructure a Reality* Chichester: John Wiley & Sons Ltd. doi: 10.1002/0470867167.ch36

Björneborn Lennart and Peter Ingwersen
2004 Towards a basic framework for webometrics. *Journal of the American Society for Information Science and Technology* 55(14): 1216–1227.

Brown, Patrick O., Michael B. Heisen and Harold E. Varmus
2003 Why PLoS Became a Publisher. *PLoSBiol*, 1(1):e36. Available at: http://dx.doi.org/10.1371/journal.pbio.0000036.

Davis, Philip M.
2009 How the Media Frames 'Open Access'. *Journal of Electronic Publishing* 12(1).

Dacos Marin, Nancy Murzilli and Inès Secondat de Montesquieu
Développer les usages en ligne: de l'éditeur électronique au lecteur – Le cas de Revues.org *Réseaux Humains / Réseaux Technologiques* 5–6.

Dreyer M., N. Bulatovic, U. Tschida and M. Razum
2007 eSciDoc – a Scholarly Information and Communication Platform for the Max Planck Society *Proceedings of the German e-Science Conference*, Baden-Baden.

Edlin, Aaron S. and Daniel L. Rubinfeld
2004 Exclusion or Efficient Pricing: The "Big Deal" Bundling of Academic Journals. *ABA: Antitrust LJ* 72(1):119–157.

Ginsparg, Paul
1994 First Steps Towards Electronic Research Communication. *Computers in Physics* 8(4): 390–396.
Gray Jim, David T. Liu, Maria A. Nieto-Santisteban, Alexander S. Szalay, Gerd Heber, and David DeWitt
2005 Scientific Data Management in the Coming Decade *Microsoft Research* 34(4): 35–41.
Greenstein Daniel
2000 Digital libraries and their challenges *Library trends* 49(2): 290–303.
Groth Paul, Andrew Gibson and Jan Velterop
2010 The anatomy of a nanopublication *Information Services and Use* 30(1–2):51–56.
Harnad, Stevan
2011 Gold Open Access Publishing Must Not Be Allowed to Retard the Progress of Green Open Access Self-Archiving. *Logos*. (Submitted).
Heck,Andre (ed.)
2003 *Information handling in astronomy: historical vistas, Astrophysics and space science library.* Vol. 285, Berlin/New York: Springer.
Heintz, Wulff and Carlos Jaschek
1982 *Automated data retrieval in astronomy: proceedings of the 64th colloquium of the International Astronomical Union*. Strasbourg France July 7–10 1981, Dordrecht, Boston: Kluwer Academic Publishers.
Ide Nancy and Laurent Romary
2004 International standard for a linguistic annotation framework *Natural Language Engineering* 10: 211–225.
Kling Rob and Geoffrey McKim
1999 Scholarly Communication and the Continuum of Electronic Publishing *Journal of the American Society for Information Science* 50(10): 890–906.
Lagoze, Carl and Herbert van de Sompel
2001 The Open Archives Initiative: Building a low-barrier interoperability framework. *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries*: 54–62.
Lopez Patrice
2009 GROBID: Combining Automatic Bibliographic Data Recognition and Term Extraction for Scholarship Publications *Research and advanced Technology for Digital Libraries* Lecture Notes in Computer Science 5714: 473–474, DOI: 10.1007/978-3-642-04346-8_62
Lossau Norbert
2004 Search Engine Technology and Digital Libraries: Libraries Need to Discover the Academic Internet *D-Lib Magazine* 10(6).
Lynch, Clifford A.
2003 Institutional Repositories: Essential Infrastructure For Scholarship In The Digital Age. *Libraries and the Academy* 3(2): 327–336.
McCabe, Mark J.
2004 Law serials pricing and mergers: A portfolio approach. *Contributions to Economic Analysis & Policy* 3(1).

Ochsenbein François, Pierre Fernique, François Bonnarel, Mark Allen, Thomas Boch and Françoise Genova
2005 Interoperability in Action: the Aladin Experience Interoperability in Action: the Aladin Experience XIV 347:193–197.
Pilat Dirk and Yukiko Fukasaku
2007 OECD Principles and Guidelines for Access to Research Data from Public Funding *Data Science Journal* 6: OD4-OD11
Pöschl, Ulrich
2004 Interactive journal concept for improved scientific publishing and quality assurance. *Learned Publishing* 17:105–113. Available at: www.copernicus.org/EGU/acp/poeschl_learned_publishing_2004.pdf
Romary, Laurent
2011 Stabilizing knowledge through standards – A perspective for the humanities. Karl Grandin (ed.): *Going Digital: Evolutionary and Revolutionary Aspects of Digitization*, Science History Publications. Available at: http://hal.inria.fr/inria-00531019
Romary, Laurent and Chris Armbruster
2010 Beyond institutional repositories. *International Journal of Digital Library Systems* 1 (1): 44–61. Available at: http://hal.inria.fr/hal-00399881
Subramanyam K.
1983 Bibliometric studies of research collaboration: A review *Journal of Information Science* 6(1): 33–38.
Sale, Arthur H. J.
2007 The patchwork mandate. *D-Lib Magazine* 13 (1/2).
Suls, Jerry and René Martin
2009 The Air We Breathe: A Critical Look at Practices and Alternatives in the Peer-Review Process. *Perspectives on Psychological Science* 1:40–50.
Velterop, Jan
2003 Should scholarly societies embrace open access (or is it the kiss of death)? *Learned Publishing* 16(3):167–169.
Wilbanks, John, and James Boyle
2006 *Introduction to Science Commons*. Science Commons. http://sciencecommons.org/wp-content/uploads/ScienceCommons_Concept_Paper.pdf.
Wood, John (ed.)
2010 *Riding the Wave: How Europe can gain from the rising tide of scientific data.* Report of the High-Level Group on Scientific Data, European Commission, October 2010. Available at http://cordis.europa.eu/fp7/ict/e-infrastructure/high-level-group_en.html

12. Verbal communication protocols in safety-critical system operations

Peter Bernard Ladkin

1. Background: Systems, safety, agents, joint behavior

A “system” is best defined as a collection of objects that exhibit joint behavior. An object which exhibits behavior is often known as an agent; so a system is a collection of agents. Engineering systems are designed by humans to fulfil a particular purpose, to achieve a set of goals, thus they may be deemed “teleological”. Besides those other agents in the system with which an agent might interact, system agents interact with – that is, exhibit joint behavior with – objects not in the system. These jointly-interacting but non-system agents constitute the “environment” of the system.

After an initial introduction to the background with definitions of these system types and related concepts, a series of case studies which exemplify crucial aspects of safety-critical system operations in the domains of air and train traffic will be presented, followed by a discussion of formal aspects of safety critical systems. Key distinctions will be made at appropriate places between linguistic and non-linguistic elements in safety-critical system interactions.

1.1. Agents and agent interaction

Consider a desktop or laptop computer. Such a system consists of agents – chips and other electronic components sitting on a motherboard, screen and keyboard, connections to other devices known as “peripherals”. A user can be considered to belong to the environment, through which interaction takes place over very narrow channels: keyboard input on the one hand and screen output on the other. Such interaction is not necessarily best analysed or understood using linguistic constructs. Neither is desktop computer use in its computational function per se either safety-critical or safety-related.

However, there are systems consisting of discrete agents in which joint behavior involves explicit communication, and in which this communication may have implications for safety. Consider a highway system, with a junction uncontrolled by STOP signs or traffic lights. You drive up to the junction and stop; to your right, another vehicle drives up and also stops. The driver gestures with her hand towards you, and moves her hand across the windshield in your direction of travel. You take her to be waving you across, and accelerate. She was, however, attempting to indicate to you the traffic jam on the other side, and that you cannot therefore proceed, expecting you to conclude that she, however, can. She also accelerates; you both collide. The interaction, though performed by means of gesture, admits analysis in linguistic terms because, well, ambiguity is a typical linguistic issue, and that is how it was just explained what the miscommunication consisted in!

A road system is only partially engineered. Some teleological transport systems are more thoroughly engineered, down to the communication amongst agents, for example railways and air traffic. Both involve safety-critical behavior, as does road transport. We shall consider some cases in which technical, partially- or fully-controlled communication amongst system agents and the environment plays a fundamental role critical for safety.

1.2. Safety-critical systems and protocols

What does "safety-critical" mean? The concept "safety" is defined in a number of ways in the engineering literature when talking about the safety of systems. One definition is: freedom from accidents (Leveson 1995), where an accident is defined as unwanted but not necessarily unexpected behavior of some kind. Another definition, in the international standard governing the functional safety of systems involving programmable-electronic components: freedom from unacceptable risk (IEC 61508). Risk is defined here in the usual way as the combination of the probability of harm occurring and the severity of that harm, where harm is considered to be physical injury or damage to the health of people, or damage to property or the environment.

Much behavior which has consequences which may be critical to safety is regulated within a system by human agents – for example, a policeman directing traffic at a busy intersection. Other examples: air traffic control is intended to ensure the physical separation of participating aircraft; train dispatching protocols coordinate the behavior of trains running in opposing directions, or different-speed trains running in the same direction, on single-track lines.

Traditionally, these protocols are purely verbal, but increasing use is made of automated helper systems, such as the ASCII-text CPDLC system for relaying ATC clearances and acknowledgements between computers on board and in air traffic control centers, with initiation and checking performed by humans, or the computer-based protocol-aiding systems on board regional dispatched trains on Austrian railways (Stadlmann and Zwirchmayer 2004). Such protocols are synchronous, in the sense that both initiating agents are interacting in real time, whether the communication is mediated by digital-electronic devices or not.

Not all communications involving human agents take place between two human agents, however. Central monitoring system messages in complex modern aircraft determine the status of various critical systems and relay changes in status to an electronic bulletin board which the pilots read. These communications are different from the verbal-exchange protocols in that one communication channel, from the central monitoring system to the pilot's screen, is designed asynchronously, including its execution. The execution is controlled by synchronous events, but the system adapts according to algorithms which have been designed in the engineering workshop largely before the aircraft was built, and certainly before it flew its present trip.

In the following sections we consider some incidents in which these communications went wrong, and how they may be repaired. Potential or actual solutions to the problems exhibited are often not primarily linguistic but involve technical measures to which linguistic features contribute.

The verbal protocols used in air traffic control are crucial to the safety of air travel, a major mode of transport used by increasingly many. This raises the question why more attention has not been paid by researchers and engineers to this domain. Here are some possible reasons.

First comes a social reason. The skills of safety-critical system analysis and development consist in hazard analysis, safety requirements derivation, risk analysis, risk avoidance and mitigation; understanding and knowing how to use the basic concepts and techniques. To the author's knowledge, there are few linguists, with the possible exception of Cushing (below), who possess these skills to the necessary degree. Similarly, computational-linguistic skills, skills in formal semantics and syntax, and in discourse analysis and the engineering of controlled languages, are generally not to be found amongst safety-critical system engineers.

Second, the verbal protocols in, say, air traffic control work nowadays very well. Major accidents in which air traffic control behavior has played a role most often involve passivity – a controller could possibly have intervened in anomalous aircraft behavior, but somehow did not. Such situations are not primarily linguistic and so are of less interest here. In the first half of 2010, major accidents disproportionately involved a loss of control or of situational awareness, and none involved difficulties with verbal crew-ATC protocols – see for example (Ladkin 2010).

There are two notable commercial air accidents in the past 15 years in which aspects of the verbal ATC-flight crew protocols played a causal role. We analyse the linguistic aspects of both these situations.

The other set of verbal protocols considered are those for train dispatching on non-signalled single-track lines. Such lines are relatively rare in Europe (although not in, say, Australia, where such lines may be hundreds of miles long, and exhibit quite different phenomena), and relatively sparsely used (for otherwise they would be signalled). So the system is not as pervasive as that for air traffic. The protocols are simpler, less involved, and arguably more robust. These characteristics enable technical solutions to reliability and accuracy problems, as will be shown, which solutions could be considered exemplars for how one may go about the task of increasing reliability, accuracy and safety in the air-traffic environment.

2. Case studies and linguistic modelling: Air traffic control

2.1. Case study: The TCAS collision avoidance system in the Überlingen midair collision, 2002

2.1.1. Description of the Überlingen incident

On 1 July, 2002, a Tupolev 154M commercial jet transport aircraft, operated by Bashkirian Airlines (BTC), a Russian airline, was flying westwards at night over Southern Germany towards a destination in Catalunya. A Boeing 757 operated by the cargo airline DHL was flying northbound over Switzerland, at the same Flight Level 360 (representing a nominal altitude of 36,000 feet in a "standard" atmosphere). Both were operating under Instrument Flight Rules (IFR), compulsory at this Flight Level. Skyguide, the Swiss air traffic control organisation, had control of both aircraft in its Zürich center, and accordingly had responsibility for separation of the aircraft.

The controller on duty was operating two positions, some meters apart, because colleagues were on break. He was working primarily with other traffic at one position, and only noted the convergence of the two aircraft close to the limit of the separation he was required to ensure (7 nautical miles lateral and/or 1,000 ft vertical separation). Another air traffic control facility at Karlsruhe had noticed the convergence, but was unable to contact Zürich Center through the dedicated communication channel, which was undergoing maintenance. An automatic "early warning" system installed at Zürich Center was also undergoing maintenance and did not trigger.

The controller issued a verbal avoidance manoeuvre to BTC to descend immediately. However, both aircraft received a Resolution Advisory (RA) from their on-board Airborne Collision Avoidance System (ACAS) devices, both TCAS II Version 7.0 from the company ACSS, virtually simultaneously with this instruction. The avoidance manoeuvres advised by TCAS are strictly vertical: one aircraft is advised to climb, and the other to descend. The rates at which these manoeuvres are to be accomplished are also normed in the TCAS protocol: a smooth 1/4g acceleration to a climb, resp. descent rate of 1,500 feet per minute is to be performed. TCAS advised to BTC an immediate climb, and to DHL an immediate descent.

DHL descended. The BTC commander also instructed his Pilot Flying (PF) to descend. 7 seconds later, the air traffic controller repeated his descend instruction to BTC with an note to "expedite", for traffic which he mistakenly described as at the “two o’clock” relative position. (The direction of flight is “twelve o’clock”; two o’clock is two segments to the right, so two-twelfths of 360°, 60° right of BTC’s direction of flight.) BTC was in fact at “two o’clock” to DHL; DHL was correspondingly at “ten o’clock” to BTC. This was a cognitive slip by the controller. Such slips are not uncommon, and normally inconsequential. In this case, however, it caused the BTC commander to believe he was in a three-aircraft conflict, with DHL, whose lights the crew could see and had identified at their ten o’clock position, and with an unknown aircraft which his TCAS display was not “seeing”, at his two o’clock position (BFU 2004). (Ladkin had speculated that this might have been so already in (Ladkin 2002). The present analysis is based heavily on the final paper (Ladkin 2004).)

7 seconds later, DHL received an “iterated advisory” to “increase descent” (to a normed rate of 2,500 fpm). 9 seconds after that, DHL informed air traffic control that he was in a “TCAS descent”.

Air traffic control procedures are such that the controller is no longer responsible for separating traffic responding to TCAS Resolution Advisories until it is reported to him by the participants that they are “Clear of Conflict”. However, a controller may continue to provide information to participants during the manoeuvres. The air traffic controller conformed with this procedure.

11 seconds after DHL informed the controller of the TCAS descent, the two aircraft collided.

2.1.2. Discussion of the Überlingen incident

TCAS is a system which of which the agents are a visual and aural display, along with “advisories”, issued according to a programmed algorithm. The agents who activate the avoidance manoeuvre are the human pilots of the involved aircraft. The communication between TCAS agent and human is visual, and aural (involving commands expressed as English sentences). What of communication between ATC and TCAS or ATC and aircraft crew? The TCAS “philosophy” is that TCAS only triggers an RA when ATC separation has “failed”; it is an on-demand system implementing a function based on the hypothesis that ATC has failed to fulfil its function. Hence TCAS is considered as a system, by some only involving the two programmed boxes and associated electronics, but this is analytically inadequate because the crucial joint behavior, the avoidance manoeuvre, is initiated and performed by the human crews of the aircraft; the TCAS boxes issue “advisories”. So TCAS boxes plus flight crew of each aircraft belong to the “system”. ATC is not considered to be part of the system: TCAS is predicated on failure of ATC function and one wants to remedy this failure through TCAS, not incorporate the failure into the system itself.

However, this part of the TCAS “philosophy” (i.e. modus operandi as construed by the designers) is also inadequate to the reality, as shown clearly in this accident. The controller issued information from which the BTC crew could infer the presence of a third aircraft with which they also had a conflict. This aircraft was "phantom", the statement containing this information mistaken. But BTC took this information as correct, which is appropriate given the understood roles of the participants, and made a decision based on it.

They were in fact presented with a decision problem: one can only go up or down; one sees an aircraft which TCAS is likely telling one to avoid; there is another aircraft which one does not see (either visually or metaphorically via TCAS display), but which the controller sees, and given the sense of the controller's instruction is likely at the same flight level or above. Two "threatening" aircraft; only two directions (up/down), either choice exacerbates one conflict or the other. BTC choose to descend, and have been widely criticised in the literature for so choosing. Ladkin suggested that decision to have been rational, given the crew's understanding of the state of the airspace and its occupancy, and gave reasons in (Ladkin 2004). This is not the focus of the observations here. Neither is the apparent TCAS algorithm discrepancy which failed to issue a so-called "reverse RA" to DHL when TCAS sensed that separation was not being insured by the first RA, for that is a pure system issue, and not linguistic. (This anomaly has been fixed by Change Proposal 112E in the most recent version of TCAS, Version 7.1.)

The relevant observation here concerns the causal efficacy of mistaken information in a communication that was legitimate according to protocol but involved a not-uncommon cognitive slip. This information gave the BTC crew a mistaken understanding of the state of the airspace around them: they were considering two "threats", not one. In fact, all three participants in this interaction, the two TCAS participants and the Zürich Center controller, had at this point three mutually incompatible understandings of the situation. The situation can be expressed simply as states of parameters:

posn:	vertical and horizontal position; correspondingly *Dposn* for DHL position and *Bposn* for BTC position, with its two components:
hpn:	horizontal position; correspondingly *Dhpn* and *Bhpn*
alt:	altitude (vertical position); correspondingly *Dalt* and *Balt*
RA:	an RA has been issued
sense:	the sense of the RA is known, correspondingly *Dsense* and *Bsense*, which should be opposite
move:	the up/down/level horizontal motion of an aircraft, correspondingly *Dmove*, *Bmove*.
phantm:	there is a (third) aircraft at BTC's two o'clock

The collection of values of these parameters will be called here the *state* of the airspace. To indicate who has what information, and who does not have it, + and – signs are used. The state consists of the values of the parameters thus:

Dposn, Bposn, +RA, Dsense down, Bsense up, Dmove down, Bmove down, --phantom

However, a rational reconstruction of the state of knowledge of the three main actors during the crucial initial few seconds of the encounter yields the following state:

ATC:	*+Dposn, +Bposn, --RA, --Dsense, --Bsense, --Dmove, Bmove level, --phantom*
DHL crew:	*+Dposn, +Bhpn, --Balt, +RA, +Dsense down, +Bsense up (inferred), +Dmove down, --Bmove, --phantom*
DHL TCAS:	*+Dposn, +Bposn, +RA, +Dsense down, +Bsense up (inferred), +Dmove, +Bmove, --phantom*
BTC crew:	*+Dhpn, -Dalt, +Bposn, +RA, +Dsense down (inferred), +Bsense up, --Dmove, +Bmove down, +phantom*
BTC TCAS:	*+Dposn, +Bposn, +RA, +Dsense down (inferred), +Bsense up, +Dmove, +Bmove, --phantom*

A simple observation is that all these five understandings of system state are different. Are they compatible? BTC crew's state is compatible with that of no other participant, because of the value of phantom. DHL crew's understanding is compatible with that of DHL TCAS. ATC's understanding is incompatible with that of no other participant, because of the value of *RA*. DHL TCAS and BTC TCAS are mutually compatible (which we would hope!).

Ladkin called these rational reconstructions of the partial system state, given the knowledge of each agent, the agent's *Rational Cognitive Model* (RCM) of the situation (Ladkin 2009). Note that two of the agents are boxes, DHL TCAS and BTC TCAS.

It is normal that in the design of safety-critical systems states are identified which can be deemed to be *hazardous,* that is, roughly speaking, states in which the chances of an ensuing accident are much increased over those chances in a non-hazardous state (there are in fact various different notions of hazard in various engineering domains, which this loose statement attempts to cover). This state, in which many but not all significant agents have incompatible understandings of the partial state, is obviously hazardous – the aircraft collided seconds later!

Hazard analysis (HazAn) attempts to identify all hazardous situations in system operation, and avoid them or mitigate the possible consequences. HazAn is required by most standards for development of safety-critical systems, as here also. It is clear from the above that hazardous states can issue from communication of mistaken information. The questions which interest communications specialists, and which must be answered, are: Which hazardous states thereby ensue? And what to do about them? Analysis of this example here has indicated that answers may be found using techniques of finite-state combinatorics and analysis, techniques which are well-known to formal linguists.

Ladkin has proposed principles for the design of such systems, which principles can help avoid such hazard states as the above (Ladkin 2009). One such principle is the *Rational Cognitive Model Coherence Criterion*, that all participants' RCMs, their partial understandings of state, are compatible. Another is the *Mutual Cognisance of Relevant Parameters* criterion, that agents who need to know the value of certain parameters should have the true values available. Here, ATC did not know RAs had been issued until well into the interaction; and DHL did not know that BTC was also descending. It is of course a far cry from knowing which principles were violated to being able to devise protocols which adhere to them! For example, one instance of MCRP violation is (we may hope!) solved by CP112E. The other may be solved by downloading RA status, when two are issued, automatically to ATC. However, controllers have legitimate worries about the intended use of such information and their responsibilities upon reception of it, and such worries have not been resolved at time of writing.

2.2. Case study: Controlled collision with terrain on approach into Cali airport, Colombia, 1995

2.2.1. Description of the Cali incident

On 20 December, 1995, an American Airlines Boeing 757 commercial passenger jet aircraft was flying into Cali airport in Colombia at night. Cali airport lies in a long, narrow valley surrounded by high mountains, and the runway lies along this valley, aligned at 10° right of magnetic north, allowing one to land to the south (on the end labelled Runway 19, the 19 standing for 190° magnetic on the compass) or to the north (on RWY 01, at 10° magnetic). To approach, an aircraft has to descend within the valley directly aligned with the runway, towards one end or the other, so at a heading of 10° magnetic or 190° magnetic, depending. The crew were used to overflying the airport, turning back, and landing to the north on RWY 01. However, on a sparse-traffic night they were offered a "direct" "straight-in" approach to RWY 19. They were unfamiliar with that approach, and hurried to find it in their charts (called "approach plates" because each is printed on one standard-format page. They may be found in Appendix C of the accident report in (Aeronautica Civil, Cali 1996).)
There was also some informational confusion. This account follows closely the account in (Gibbon and Ladkin 1996). Here is the pilot-ATC exchange, taken verbatim from the accident report, from the original approach clearance to the end of reported conversation.

> *Approach replied, "Roger 965 is cleared to the VOR DME approach runway one niner, ROZO Number One arrival, report Tulua VOR"*
> *The flightcrew readback was, "Cleared the VOR DME one niner ROZO one arrival, we'll report the VOR, thank you Sir"*
> *Cali approach immediately clarified with, "Report Tulua", and the flightcrew immediately acknowledged, "Report Tulua"*

The flightcrew referred to the cockpit chart package (approach publications) after ATC instructions to "Report Tulua"
Flightcrew discussion took place about the navigational aids to be used in the ROZO 1 Arrival, specifically their position relative to Tulua
About 30 seconds later the flightcrew requested, "Can American Airlines 965 go direct to ROZO and then do the ROZO arrival sir?"
Several radio transmissions then took place: Approach replied, "affirmative direct ROZO one and then runway one niner, the winds calm". The flightcrew replied, "all right, ROZO, the ROZO 1 to 19, thank you, American 965". And the controller stated, "Affirmative, report Tulua and twenty one miles, 5000 feet". The flightcrew acknowledged, "OK report Tulua, twenty one miles at 5000 feet, American 965"

2.2.2. Discussion of the Cali incident

Gibbon and Ladkin remark that AA965's request (after the phrase '*About 30 seconds later'*) demonstrates a linguistic expectation. Going '*direct to ROZO and then do the ROZO arrival'* only makes sense if one expects that navigation fix ROZO is at the beginning of the ROZO One arrival. Which it isn't. A look at the approach plate in Appendix C of the report (*op.cit.*) shows that the ROZO navigational beacon is at the end of the ROZO arrival, not at the beginning. In the US, all such arrivals are named after the start point, not the end point as here.

Gibbon and Ladkin suggest that this reversal of the naming convention common in the US may have induced semantic dissonance in the US crew, which can lead to confusion with respect to temporal as well as spatial location (see, for example, Gibbon 1995). However, they note that the opposite is true with road naming conventions in a variety of cultures, e.g., close to where the author lives in Germany, between-town roads are named with the presumed goal: the very same road is called the *Wertherstrasse* (Werther road) in Bielefeld, and the *Bielefelder Strasse* (Bielefeld road) in Werther. The entry point for the ROZO One arrival is in fact the *Tulua* navigation beacon, some 22 nautical miles before the *Initial Approach Fix* (IAF) for the RWY 19 approach, which IAF is itself 9 nautical miles before the ROZO beacon, which is the *Final Approach Fix* (FAF) for the approach to RWY 19, 2.6 nautical miles before the threshold of RWY 19.

Gibbon and Ladkin note that the ATC response '*affirmative'* to this request is incorrect. AA965 cannot procedurally fly what they requested; therefore the controller's confirmation of that request is wrong. An ATC "*clearance*", which is what the flight crew and the controller are discussing, is both a commitment by ATC to keep the airspace clear for the space and time slots indicated and a mandatory (but revisible) routing instruction. Any clearance must therefore be a possible routing. Which the putative clearance contained in AA965's request is not. (When asked why he replied "*affirmative*", or "yes", when he knew that the correct reply was "*negative*", or "no", the controller cited relative status – he conceived of his to be much lower than that of airline pilots – and a consequent unwillingness to be seen to contradict people of higher status, even when he could see they were mistaken (Aeronautica Civil, Cali 1996)).

Gibbon and Ladkin also note the following words spoken by ATC: ‘*direct ROZO one’* (a complete phrase, as indicated by the following conjunction, ‘*and’*.) and that it is a syntactically incorrect phrase in any grammar of pilot-controller speech. The word ‘*direct’* must be followed by the name of a navigation fix. If “ROZO” is interpreted as the fix, the word ‘*one’* is then superfluous. The entire phrase ‘*ROZO one’* denotes an arrival, and an arrival name cannot correctly follow the word ‘*direct’*. They suggest that the lexical ambiguity between fix (a point) and arrival (a procedure) enabled production of this syntactically incorrect and semantically confused phrase.

The Rational Cognitive Models of the participants may be reconstructed thus, as they evolve through time:

ATC:	*ROZO One arrival* [Tulua fix to ROZO fix]; *then RWY 19 approach* [ROZO fix to RWY 19].
AA965:	Confirm: *ROZO One arrival* [unknown fix to unknown IAF RWY 19]; *then RWY 19 approach* [unknown IAF to RWY 19]
ATC:	*Report overhead Tulua fix* [at start of ROZO One arrival]
AA965:	Confirm: *Report overhead Tulua fix* [?expectation: at end of ROZO One arrival?]
AA965:	Request: *fly direct to ROZO fix and then ROZO One arrival* [ROZO to unknown IAF]
ATC:	Affirm request. {Semantically inconsistent phrase}; *and then ROZO One arrival* [ROZO to RWY 19]
AA965:	Confirm *ROZO fix and then ROZO One arrival* [ROZO to unknown IAF] and then *RWY 19 approach* [unknown IAF to RWY 19]
ATC:	Affirm. *Report overhead Tulua fix.*
AA965:	Confirm. *Report overhead Tulua fix.*

We can represent the data structures here, giving the semantics of the phrases, by pairs: <*x,y*> represents flying from overhead Fix *x* to overhead Fix *y*; and by sequencing: <*x,y*>*;*<*z,w*> represents flying <*x,y*> and then <*z,w*>, and *z* must equal *w* (that is, either identical, or two different designations for the same fix).

We may also compose, namely <*x,y*>*;*<*z,w*> may be represented by <*x,y=z,w*>. Then the syntax and semantics of the interchange proceeds as follows.

Each line has three components, namely agent:syntax:semantics, with the components separated by colons. “Arr” abbreviates “arrival” and “App” “approach”.

Elements of a sequence are denoted by “*.first*”, “*.second*”, etc.

Here, the Initial Approach Fix for the RWY 19 approach is the point at 21.0 nautical miles on the 013° radial of the Cali VOR navigation beacon (identifier CLO) (Aeronautica Civil, Cali 1996). I abbreviate by CLO013/DME21.0:

ATC: ROZO One Arr; RWY 19 App: *<Tulua,IAF=CLO013/DME21.0,FAF=ROZO,RWY 19>*

AA965: ROZO One Arr; RWY 19 App: *<unknown,unknown=IAF,unknown=FAF,RWY 19>*

ATC: Say overhead *Tulua*: say overhead *clearance.first*

AA965: We say overhead *Tulua*: say overhead *clearance.unknown*. Supposition: *clearance.unknown* is *clearance.second* or later

AA965: Request ROZO and then ROZO One Arr: *<ROZO,unknown=IAF,unknown=FAF,RWY 19>*

ATC: Affirm, {inconsistent} and then ROZO One Arr: *inconsistent*

AA965: ROZO and then ROZO One Arr and then RWY 19 App: *<ROZO, unknown=IAF,unknown=FAF,RWY 19>*

ATC: Affirm, Say overhead *Tulua*: say overhead *clearance.first*

AA965: We say overhead *Tulua*: say overhead *clearance.unknown*

It is clear from this representation of the syntax and semantics of these communications that the syntax is ambiguous and the RCMs of the participants are inconsistent, and remain so despite repeated attempts to unify them. This sequence shows a clear violation of RCMC and MCRP.

The following then happened. Because AA965 considered ROZO to be *clearance.first*, they attempted to fly direct. Unbeknownst to them, there was another navigational beacon, with the selfsame abbreviated identifier and the same frequency as ROZO, lying roughly at the nine o'clock position and within reception range. Upon entering the data for ROZO in the flight management computer, the autopilot then began a left turn towards this other beacon, ROMEO, which went unremarked by the crew for some significant portion of a minute. They deviated from the line of the valley, and now had mountains in between them and the airport. However, they continued to descend into the airport. Upon realising the aircraft was flying in the wrong direction, the crew finally initiated flight towards the right fix, but there were now mountains in the way. As the aircraft flew towards a mountain, the ground proximity warning sounded, but the "escape manoeuvre" was not successful and the aircraft hit the mountain.

The above analysis proceeded, as did the analysis of the Überlingen midair collision, by representing the cognitive states of participant agents by simple data structures. Unlike in the Überlingen case, some concerted attempt was made to reconcile the RCMs of the participants through the usual ATC-pilot communication format, but this format was ambiguous (strictly: the same phrases were allowed different semantics). Also, the ATC mistake in the Cali situation was deliberate, engendered by social concerns, rather than being inadvertent as with Überlingen.

An attempt to enforce RCMC and MCRP in this case would involve disambiguating the language. This is formally straightforward: for example, one could require participants to enumerate explicitly the semantics – their understood sequence of fixes – saying "*unknown*" where they do not know, allowing a knowledgeable communication partner to precisify. However, this would involve a major change in ATC-pilot communication. It is not clear whether such a major change would avoid more accidents than it might engender, for every change requires some time for it to mature in use.

We may observe that, once again, simple finite-state methods enable an adequate analysis of the communications and suffice to generate suggestions for avoiding or mitigating hazards.

2.3. The linguistics of ATC-pilot communication: Work of Stephen Cushing and extensions

The linguist Stephen Cushing investigated linguistic phenomena in ATC-pilot communication leading to incidents and accidents on behalf of NASA in 1987–9. His results were presented in (Cushing 1994), predating the accidents above and their analysis.

Cushing compiled a superb collection of examples, largely drawn from NASA's Aviation Safety Reporting System, an anonymised reporting facility guaranteeing immunity for reporters, intended to identify trends as well as individual occurrences or potentially hazardous situations so that these could be pre-empted if necessary by regulators in the hope of reducing accident statistics.

Cushing identified problems of ambiguity (as we considered in the case studies above), homophony (similar-sounding phrases having different meanings, for example similar aircraft call-signs and consequent misunderstandings of a crew that a clearance intended for another aircraft was in fact for them), the influence on meaning of spoken punctuation and of intonation. He identified problems of reference: uncertain reference, unclear addressee and unclear handoff (an air traffic control procedure for assigning an aircraft to a different controller), which overlap with problems with homophony. Then there are problems of inference, including implicit inference, lexical inference, and false assumptions, as we have seen above concerning the Cali accident. Then come problems of repetition, especially involving readbacks, which we have also seen in the Cali case study above. There are problems with numbers and their abbreviations, in particular for radio frequencies, in navigational procedures, for runways, and for altimeter settings (communications of the local barometric pressure). He noted problems with radios, problems of compliance, and general problems of technical communication loss such as lost messages or garbled message contents. Although such problems are formally recognised and adherence to the controlled language of ATC-pilot communication is supposed to be a prophylaxis, Cushing's investigations showed just how big a gap there is between the reality of everyday ATC-pilot communication and the solutions offered by the controlled language promulgated by the aviation regulatory authorities (FAA ATC Phraseology 2010).

One reason for the continuing discrepancies may well be found in work of the organisational sociologist Jens Rasmussen (Rasmussen 1997). He observed that human agents in their everyday work devise ways of getting their job done optimally. That is, optimally for them, and the agent may well have different goals from the designers of the system within which heshe is working. Rasmussen noted in many accidents how these individual optimisations, which can be observed and which he called “migration to the boundary” (MttB), his phrase for coming up against the limits of what one can optimise, conflicted either with overall system safety goals, or with each other to engender hazardous system states and behaviors. Most ATC-pilot communication is routine, and much of it can be expressed by use of much shortened phrases amongst knowledgeable pilots and controllers: elisions and suchlike. And these optimisations, such as elisions, rarely lead to miscommunication, lead even more rarely to hazardous situations, and these themselves lead even more rarely to accidents. So MttB in formalised communication has immediate and pervasive benefits and only occasional, very rare, disadvantages. Which even more occasionally result in accidents, sometimes with heavy loss of life. Which of course is why the formalised communication, not foreshortened by MttB-derived modifications, is promulgated in the first place.

Cushing suggested a “long-term solution” to the disadvantages derived from the pervasive linguistic phenomena which he had enumerated. The solution consisted in a computer-supported “intelligent” voice interface for aviation communications, resting on a formal language. He proposed an architecture for the system, and use of some computer-supported-linguistic tools, such as LEX and BISON, based on a formal grammar of aviation communications. He made extensive use of SW available on Apple Macintosh computers, at that time limiting the applicability of the work since Apple did not have a large market share of personal computers.

Hilbert and Ellermann showed, however, that the grammar proposed by Cushing was formally inadequate for these communications, in that it did not correctly represent them, allowing some phrases that were inappropriate and ruling out other phrases in officially-sanctioned use (Ellermann and Hilbert 2001). They went on to show that accurate pilot-ATC communications could in fact be represented by a formally-simpler grammar in extended Backus-Naur form (EBNF), which makes it suitable for standard computer-linguistic parsing and analysis tools (Ellermann and Hilbert 2002). In this work, Ellermann and Hilbert also implemented a parser for their grammar on a Unix-based computer system. UNIX had a much broader installed base than Apple systems at that time. Nowadays, Apple operating systems are also Unix-based.

Hölz and Hettenhausen designed and partially implemented a limited computer corpus of aviation communications for published accidents (Hölz and Hettenhausen 2001). This task is made particularly difficult in that such communications are not publicly available, in part through agreements with professional-pilot unions for reasons of privacy and employee protection in the workplace. Indeed, such transcripts are not available in most cases even within airlines. They are sometimes made available, say by the US National Transportation Safety Board, in the case of significant accidents, but other investigative authorities, such as the Transportation Safety Board of Canada, redact heavily and summarise, through concerns for the privacy of participants and victims. The chances of building a representative corpus are thereby limited. However, Döring, McGovern and Sanders (Döring, MacGovern and Sanders 2001) showed using the limited Hölz-Hettenhausen corpus that semantic parsing techniques on the corpus enabled unique correlation of spoken phrases in the transcripts with their formal Ellermann-Hilbert equivalents in a high proportion, some 80–90 %, of cases. This work validates Cushing's intuition that computer linguistics can be fruitfully applied to improve aspects of day-to-day controller-pilot communications with respect to the unsafe features he identified in his compilation of examples. A linguistic approach is also taken by Sassen (2005).

The question is, however, still unanswered what sort of computer-supported tools would most benefit these communications from the point of view of enhancing safety, and which such tools stand a reasonable chance of being implemented and successfully introduced into service. Limited versions of such communications are already flying, for example the CPDLC communications, which are digital-computer-mediated communications whereby the sender uses a keyboard, the receiver a screen, with a button for acknowledgement of reception and/or clearance. As of writing, CPDLC has been successfully used for communications on trans-oceanic flights for almost two decades.

One possible future development could see verbal statements of agents parsed by computer, transmitted digitally as text after checking for compliance, consistency, completeness, and common forms of failure, and voice-synthesised at reception. The state of the art is a long way from having such proof-of-concept prototypes, and even further from a practically workable system which would be accepted by the aviation community.

Such systems are feasible. We turn now to a domain in which such tools have been successfully implemented, the linguistically somewhat more simple domain of train dispatching and train announcement procedures.

Train Announcement Procedures (TAP, German: *Zugmeldeverfahren*) are one of the two procedures on German railways for coordinating traffic on single-track lines without signals. The other is train dispatching (*Zugleitbetrieb*, ZLB), for which computer support will be considered, below.

3. Train dispatching and train announcement procedures

3.1. Train control

Train dispatching and TAP are forms of train control used in German railway operations on sparsely-used single-track railway lines without signal support (signals are required by most railway authorities on all tracks which have above a certain frequency of train movements). Trains halt at a stopping point, usually a railway station, and negotiate a verbal protocol with a train controller in order to proceed to the next stopping point. The region of line between two stopping points is called a block, and the protocol must ensure that no two trains may proceed in the same block at the same time under central control. It must be possible, for track maintenance, rescue, and shunting purposes to allow trains to move at line-of-sight braking speeds under certain circumstances, but a train must not be allowed to proceed under normal driving conditions into a block in which some line-of-sight operations are taking place, and line-of-sight operations must not be allowed to proceed in a block in which a train is proceeding under normal driving conditions. Basically, the two allowable states for one block are: one train only or everyone is being slow and careful and can stop before hindrances.

Train dispatching operates with one train controller, who controls the movements of all trains on a specific section of track (a block). A controller using dispatching communicates with train drivers using dedicated radio.

TAP is a protocol in which two train controllers, usually the station masters at the stations at each end of a block, coordinate the dispatch of trains into the block between themselves first, and then communicate permissions with the train drivers. The station masters communicate with each other using dedicated telephone lines, and with the train drivers using dedicated radio.

The verbal protocols for train dispatching and TAP are strict, use defined phraseology, and are part of German rail law.

The protocols do most often support the operational safety requirements, but mistakes are sometimes made.

3.2. Case study: the Warngau single-track train collision, 1975

On Sunday 8th June 1975 18:32h, two passenger trains crashed into each other near Warngau, Germany. The passenger trains both entered the single-track line between Warngau and Schaftlach after receiving permission to continue. One train was lifted off the track and fell to the side. The accident killed 44 people, including the two engine drivers, and injured 122. Damage amounted to about 4 million Deutsche Mark (about € 2 million). Verbal execution of protocols played a significant role.

The single-line track between Warngau and Schaftlach is 4.8 km long. It is part of the line between München (Munich) and Lenggries, in Bavaria. Automatic Train Protection, automatic braking on passing a signal at danger, is installed and functions through inductive-electromagnetic trackside and on-board equipment (Indusi). Indusi was required for the line because it was designed for trains travelling faster than 100 kph. However, traffic density on the line was sufficiently low that compartmentalisation into blocks was not required between stations. Thus the track between Warngau and Schaftlach constituted one block (German *Blockabschnitt*) to which access was controlled by station masters/controllers (*Fahrdienstleiter*) in Warngau and in Schaftlach using TAP. The block contained one level-crossing (grade crossing), protected by warning signals.

Under TAP, trains must be *offered (OFF)*, *accepted (ACC)*, *checked out (PERM)* and *reported back (ACK)* between station masters. These procedures are as follows. Suppose a train at holding point (here, station) A wishes to progress to holding point/station B.

OFF: if he does not regard the block as occupied, station master A tells station master B that he has a train ready to enter a specific block.
ACC: station master B confirms that the train can enter (if the block is not already occupied), and holds the block as occupied until the train arrives at the B side of the block, at his station, and leaves. This arrival event will be assured visually and verbally by an ACK, below.
PERM: station master A notifies the train driver that the train may enter the block and notes the train's time of departure.
ACK: station inspector B notifies station inspector A that the train has cleared the block.

The OFF-ACK pair guards against opposite-direction trains colliding head-on, for station master B will know if he has already sent a train into the block, and station master A will have marked the block as *occupied*. The OFF-ACK pair guards against same-direction trains colliding head-on-tail. I forego a formal analysis of the TAP protocol here; we will see such an analysis below for train dispatching.

On the day of the accident train 3591 was waiting at Warngau, ready for departure. Train 3594 S (scheduled Sundays only) was waiting at Schaftlach, ready for departure. Both station masters offered their trains (OFF), and both interpreted the communication for their respective train as if they had received ACC. Both cleared their trains to depart.

The apparent misunderstanding was determined to be a result of not using the fixed phrases. It was reported that the station masters conversed using a Bavarian dialect and "cut corners" in the execution of the protocol. If so, this would then be an example of MttB. Unfortunately, there is no transcript of the communications publicly available.

According to the written schedule, train 3591 had to wait for two trains coming from Schaftlach: train 3592 and train 3594 S. The station master at Warngau had to handle three trains in a short time (about 9 minutes), and as well as sell tickets and answer passengers' questions. He was the only person in the Warngau station.

Other causal factors than miscommunication played a role. For example, a feature of train timetabling for opposite-direction trains, "air intersections", *Luftkreuzungen*, theoretical positions within a block at which trains following a timetable would cross were the block to consist of two one-way tracks, was used in the pictorial timetable for these stations for these trains. In fact, the EBO regulations explicitly disallowed *Luftkreuzung* in timetable construction. It was nevertheless common practice at this time to use them, to allow station masters some scheduling flexibility in case trains were not running on time.

Also, between Warngau and Schaftlach, there were two signals guarding a level crossing (grade crossing) for trains. The signal indicator lights start blinking when a train's first axle passes a contact point on the track. If a train from the opposite direction passes its respective contact point, then the control light does not blink. In this case the train driver must come to a full stop before the crossing. Train 3591 passed the contact point first; train 3594 S passed its respective contact point after 3591 passed his, but the driver of 3594 S did not react to the lack of blinking signal, which should have indicated to him that he should commence an emergency stop. The trains were in a blind curving section of track, so visual contact was not possible until late in the convergence. Either visual contact or braking of train 3394 S would have reduced the severity of the collision. This is not the only case in German railway history in which safety information is conveyed by a lack of action, here the lack of operation of the signal light, before shortly later an accident ensued. The Brühl derailment accident (Brinkmann and Lemke 2003) is another example. Our concern here is not, however, with the analysis of visual communications.

3.3. Warngau and Cali: Lessons from common features of communication

The Warngau collision and the Cali accident have some communication phenomena in common. Non-standard phraseology was used, in this case local dialect and ambiguous phrasing – MttB at work – and the result was an inconsistent understanding of overall system state by the participants: RCM and MCRP were again violated, obviously.

This raises the question, raised above for aviation communications, whether computer support can help secure such operations. In this case, there are working computer-supportive systems for the alternative to TAP, train dispatching. Train dispatching without local signal support is supported on some Austrian railway lines run by the company Stern & Hafferl, such as the local railway in the region around Linz, using a system developed at the University of Applied Sciences at Wels (Fachhochschule Wels) by a team led by Burkhard Stadlmann (Stadlmann and Zwirchmayer 2004). Stadlmann's work represents the state of the practice in computer-supported train dispatching, but it holds minimal interest from the linguistic point of view, since the system is primarily informational to the agent.

4. Formal grammars for protocol implementation

4.1. Computer support with a proven-correct protocol implementation

If that is state of the practice, then what about state of the art? Sieker has considered the Zugleitbetrieb protocol defined by German railway law for non-federal railways. There are many such lines run by private railway companies in Germany, but only a few, mostly in Sachsen around Dresden, are non-signalled. Most lines, including all those around Bielefeld where we work, are supported by signalling systems to some degree.

In his PhD thesis, in German (Sieker 2010), Sieker started from a formal expression in logic of the above basic requirement of Zugleitbetrieb, the one-block one-train criterion, with the distinction between central control and line-of-sight operations. He proceeded to derive the protocol along the lines indicated in the law (FV-NE – *Fahrdienstvorschriften für die Nichtbundeseigenen Eisenbahnen* – Train Control Regulations for Non-State Railway Operations of the Federal Republic of Germany), by using a technique of computer-system design elaboration known as formal refinement. Sieker derived a series of ever-more-elaborate state machines to implement the verbal protocol formally using computational techniques, and each state machine was formally proven using formal logic to mimic the more abstract state machines higher in the development hierarchy. The penultimate step consisted of factoring a state machine representing the global system state into different state machines for each agent, in such a manner that it could be proven that the interactions of the agent state-machines formed exactly the global state machine from which they were derived.

The final step was implemented by Phil Thornley of SparkSure, who implemented the agent state machines in the annotated programming language SPARK (Barnes 2003) and proved formally, using the required annotations supplementing the executable code, that the executable computer programs implemented exactly and precisely the agent state machines derived by Sieker.

The result is a fully formally analysed SW system, which runs a verbal protocol precisely, and which has been formally proved to satisfy the basic safety requirement of Zugleitbetrieb. That means that running the SW is guaranteed to execute the protocol logically correctly (Sieker discovered during the course of the work that the protocol did not define actions in all reachable states. The required actions were mostly obvious, and were added. The protocol which Sieker's development runs is strictly speaking a formal completion of the legal protocol).

Of course, there can still arise hardware problems with the machine, and it may be that the compiler which translates the SPARK source code into machine language for the machine is inexact, but these are problems which are common to all computer support of any task, are not specific in any way to mimicking verbal protocols, and have been addressed by the computer industry. Indeed the issue of unreliable compilation has been at the center of work by the developers of SPARK, the company Altran-Praxis, for two decades.

4.2. Sieker's multi-level formalisation

The details of Sieker's work here generally follow an account in (Stuphorn, Sieker and Ladkin 2009). This account focused on the safety properties of the development – for example, the novel approach to hazard analysis, derivation of a demonstrably-complete set of safety requirements, and aspects of the development that would and would not generalise to other situations – which are of less interest to us here.

The highest-level language, called Level 0, required to express the basic safety requirement is simple, indeed astonishingly simple. This simplicity allowed a manual selection of safety requirements, which is complete in the sense that one can easily show that they cannot be logically strengthened.

The next language level, Level 1 and further language levels were defined through the usual type of refinement process familiar to computer scientists, in which the extensions of the language were carefully controlled in small steps. The entire functional operation of the system at each level was able to be expressed in terms of a global finite-state machine, with which state machines were formally proved through simple, short propositional-logic arguments to refine each other, sometimes through addition of extra requirements, which then were added to the list of safety requirements.

The final refinement step involved transforming the global state machine into a set of agent state-machines, one representing a driver and one a train controller, which communicate by means of message-passing. The agent state-machines are expressed in a structure called a Message Flow Graph (MFG), for which a formal semantics has already been defined (Ladkin and Leue 1995). By this means the MFG could be formally proved to implement the global state machine from which it was refined.

The MFG agents were then implemented as SPARK procedure skeletons with the appropriate annotations by Phil Thornley of SparkSure, and the annotation proved to implement the MFG.

The entire development ensured complete traceability between very-high-level safety requirements and SPARK source code. Suppose such a system were to be implemented as either an automated dispatching system, with computers replacing the human agents, or, more practically, as a support system which checks that the required steps have been performed by the human agents. Then the risk of using the system resides entirely in the hardware and communications systems used, as well as in the compiler used to compile the source code, and in human factors such as whether the system is used as intended. Any risk inherent in the logic of the program design itself has been eliminated, and shown to be eliminated by the logical proofs. The risk of this computer-based system has thereby been reduced to that of other, generic risks, which data from other, unrelated computer projects may be used to assess.

The architectural scheme of the development is an example of a method called Ontological Hazard Analysis, proposed by myself and developed inter alia by Sieker and Sanders (Stuphorn, Sieker and Ladkin 2009, Ladkin, Sanders and Sieker 2012). The HazAn aspects of the development interest us marginally here. The structure of the development is given in Figure 1.

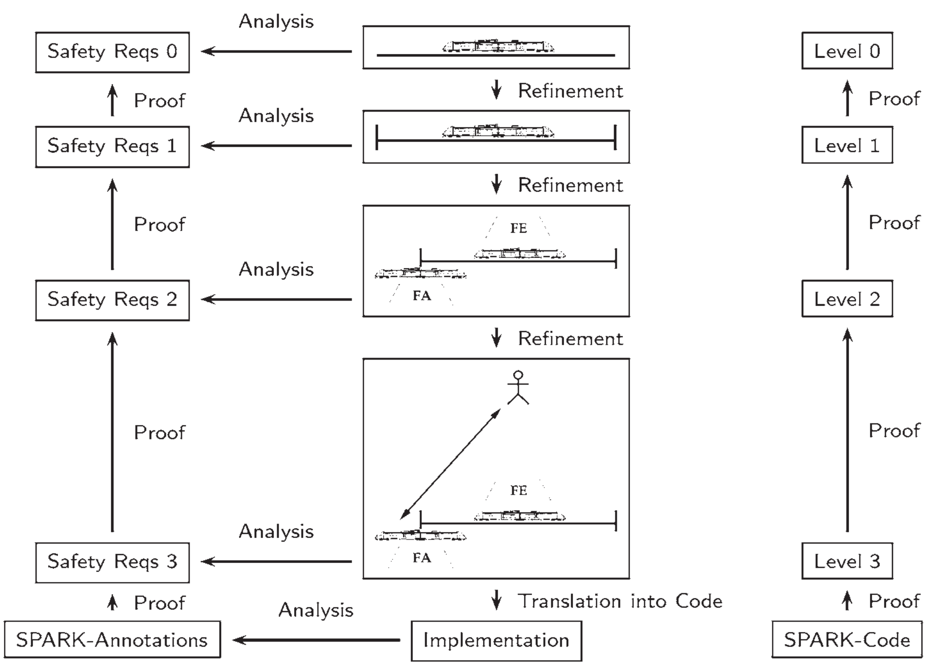


Figure 1. Structure of the OHA.

4.2.1. Level 0

The goal of the highest specification level, Level 0 (Figure 2), is to provide a description language adequate for expressing the intuitive safety requirement of train dispatching, the one-block one-train requirement with exceptions, that is so simple that (a) we can define safety axioms to which all applications experts can assent, and (b) at the same time ascertain that these axioms are both *correct* and *complete* relative to the expressions of the language.

Figure 2. Schematic Representation of Level 0.

The Level 0 language is a logical language, which includes sorts of objects, properties of objects in those sorts, and relations between objects; a basic predicate-logical language with no complex features. Sorts and relations are shown in Table 1 and Table 2.

Table 1. Level 0 Sorts.

Sort	**Description**
Vehicle	Any train or other vehicle operating on tracks
Block	A section of a track inside or outside a station

Table 2. Level 0 Relations.

Relation	**Description**
inA(F,S)	Train F is in Block S
ZV(F,S)	ZV(F,S) Train F may occupy Block S under central responsibility (normal scheduled operation)
LV(FS)	ZV(F,S) Train F may occupy Block S under local responsibility (special case)

Determining Safety Axioms. Using elementary propositional logic as well some semantic domain knowledge, Sieker is able to determine that there turn out to be only 6 safety postulates on Level 0 from consideration of a couple of dozen non-equivalent statements from a total of 256 statements before semantic reduction. We use the following shorthand notation for a train F1 and one block S: V(F1,S) = LV1, ZV(F1,S) = LZ1, inA(F1,S) = in1; similarly for train F2. The

Safety Postulates at Level 0 are enumerated in Table 3. It is not shown here how the safety postulates were derived, for this is not our main concern.

Table 3. Safety Postulates at Level 0.

Safety Postulate	**Description**
ZV1 ⇒¬LV1	If a train is in a block under central responsibility it cannot be there under local responsibility
¬LV1 ∧in1 ⇒ZV1	If a train is in a block and is not there under local responsibility then it is under central responsibility
in1 ∧ZV1 ⇒¬LV1	If a train is in a block under central responsibility it cannot be in that block under local responsibility
(F1≠F2) ⇒(LV1 ⇒¬ZV2)	If a train is in a block under local responsibility another train under central responsibility cannot be in that block
(F1≠F2) ⇒(in1 ⇒¬ZV2)	If a train is in a block another train under central responsibility cannot be in that block
(F1≠F2) ⇒(ZV1⇒¬ZV2)	If a train under central responsibility is in a block, another train under central responsibility cannot be in that block.

4.2.2. Level 1: the first refinement level

The generic block of Level 0 is refined at Level 1 (Figure 3) by introducing the new sorts Track and Station. No other modifications are undertaken, in accordance with the principle of keeping refinement steps simple. This results in the sorts in Table 4.

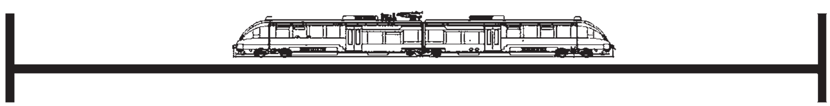


Figure 3. Schematic Representation of Level 1.

Table 4. Level 1 Sorts.

Sort	**Description**
Vehicle	Train or other track vehicle
Block	A track section
Track	A piece of track in the station
Station	A station where messages are exchanged

On Level 1, there are then 10 relations, distributing the sorts over the relations. Meaning Postulates define what each Level 0 sort and Level 0 relation means in terms of the Level 1 language. Using these Meaning Postulates, we arrive at 12 Safety Postulates for Level 1.

4.2.3. Level 2: the second refinement level

At Level 2 (Figure 4), no new sorts are added, but additional relations concerning 'clearances' are added, as shown in Table 5.

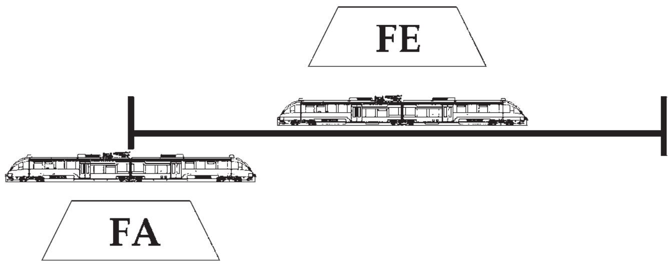


Figure 4. Schematic Representation of Level 2.

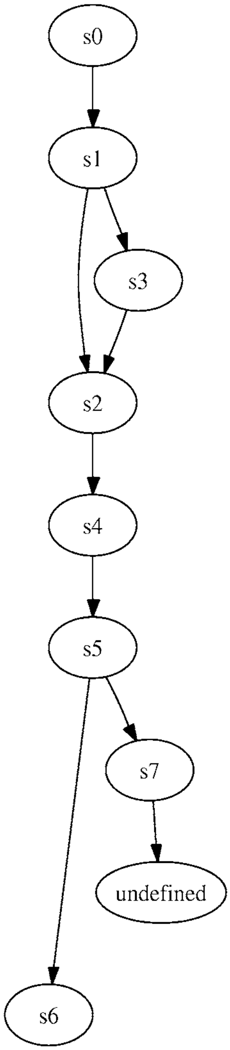
Table 5. Level 2 Relations.

Relation	**Description**
FA(F,A,B)	Train F, in station A, has asked for clearance to go to station B
FE(F,A,B)	Train F, in station A, has received clearance to go to station B
AFE(F,A,B)	Train F, in station A, has been denied clearance to go to station B
KH(F,A,B)	No obstructions are known for train F to go from station A to station B

At this point, Sieker is now able to build a state-machine representing the global states of clearances which represents a train journey. The state-machine is shown in Figure 5, which is presented as a Predicate-Action-Diagram (Lamport 1995).

Three simple Meaning Postulates and elementary logic lead to only two new Safety Postulates, which can be expressed informally as:

1. *If no obstructions are known and clearance has been given, the block can be occupied under central responsibility*
2. *Clearance for a block cannot be given for a second train, if clearance has already been given for a train for the same block in either direction.*



State	Description
s0	inZ(T , A)
s1	∧inZ(T , A) ∧FA(T , A, Next(T , A)) ∧¬ FE(T , A, Next(T , A))
s2	∧inZ(T , A) ∧FA(T , A, Next(T , A)) ∧KH(T , A, Next(T , A))
s3	∧inZ(T , A) ∧FA(T , A, Next(T , A)) ∧¬ FE(T , A, Next(T , A)) ∧¬ KH(T , A, Next(T , A)) ∧AFE(T , A, Next(T , A))
s4	∧inZ(T , A) ∧FA(T , A, Next(T , A)) ∧FE(T , A, Next(T , A)) ∧KH(T , A, Next(T , A)
s5	∧zw(T , A, Next(T , A)) ∧FE(T , A, Next(T , A)) ∧KH(T , A, Next(T , A)) ∧¬ LV(T , S)
s6	inZ(T , A) = s0
s7	∧btw(T , A, Next(T , A)) ∧FE(T , A, Next(T , A)) ∧¬ KH(T , A, Next(T , A)) ∧¬ LV(T , S)

Figure 5. Level 2 State-Machine (a Predicate-Action Diagram).

New hazards identified at this level are simply the negations of the newly-identified Safety Postulates:

1. *Clearance has been given, and no obstruction is known, but the conditions for occupying the block under central responsibility have not been met.*
2. *Clearance has been given for two trains for the same block at the same time.*

4.2.4. Level 3: the third refinement level

Level 3 includes the specific defined communications between train drivers and a train controller (Figure 6).

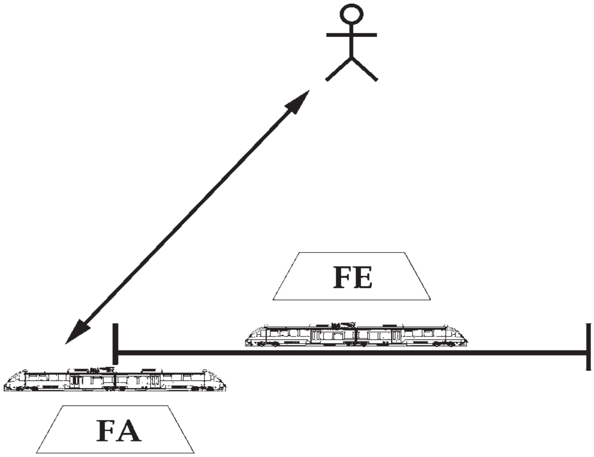


Figure 6. Schematic representation of Level 3.

Message types (Table 6) correspond to the states in which the trains can be, and are designed according to the message types prescribed in the regulations for German non-state-owned railways (VDV 2004).

Table 6. Message Types at Level 3.

Message Type	Description
FA	Request for Clearance (Fahranfrage)
FE	Clearance (Fahrerlaubnis)
AFE	Denial of Clearance (Ablehnung der Fahrerlaubnis)
AM	Notification of Arrival (Ankunftmeldung)

In addition, we define relations to describe sending and receiving of messages, as shown in Table 7.

Table 7. Relations at Level 3.

Relation	Description
Sent(MT,T,A)	Message of type MT, concerning train T and station A has been sent.
Recd(MT,T,A)	Message of type MT, concerning train T and station A has been received.

Note that the sender and receiver of the message are implicit. Messages of type FA and AM are always sent by the specific train driver to the train controller, messages of type FE and AFE are always sent by the train controller.

Through appropriate Meaning Postulates, the state machine of Level 2 can be augmented to include communications. This now more complex state machine can be transformed into a Message Flow Graph (MFG), to make the communications visually clear. The MFG represents the individual agents and their changing states as vertical lines, message passing between agents as angled lines. The MFG can be formally shown to define the same global state machine as the Predicate-Action-Diagram for this level. The MFG is used as the starting point to define the SPARK implementation and the SPARK verification conditions are determined by hand to define the MFG of Figure 7.

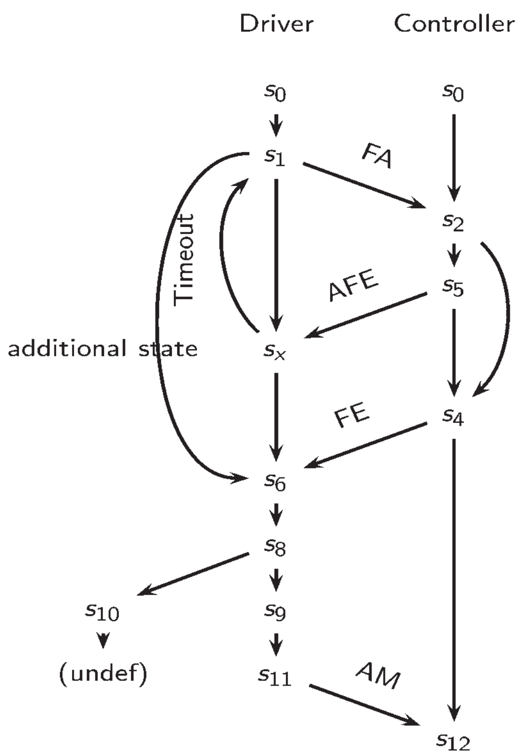


Figure 7. Schematic representation of Level 3.

Table 8. States corresponding to the Message Flow Graph.

MFG-Trans.	Driver-State	Controller State	Global State
s0	inZ(T, A)	–	inZ(T, A)
s0 → s1	∧inZ(T, A) ∧Sent⟨FA, T, Next(T, A)⟩	–	∧inZ(T, A) ∧Sent⟨FA, T, Next(T, A)⟩
s1 → s2	–	Recd⟨FA, T, Next(T, A)⟩	∧inZ(T, A) ∧Sent⟨FA, T, Next(T, A)⟩ ∧Recd⟨FA, T, Next(T, A)⟩

4.3. The step to code: Implementation in SPARK

SPARK is based on a subset of the Ada language. It uses annotations to denote data and information flow and to specify pre- and post-conditions for functions and procedures. The SPARK toolset includes a static code analyser which uses the annotations to prove the absence of run-time errors, such as division by zero, buffer overflows and other bounds violations, before the code is actually compiled. Typical Example of SPARK annotations corresponding to the MFG:

```
procedure Send_FA (DS: in out Driver_State);
--# global out Messages.Out_Queues;
--# derives Messages.Out_Queues from
--# DS
--# & DS from
--# *;
--# pre D_State(DS) = D_S0;
--# post To_S1(DS~, DS);
```

4.4. Conclusions of the FV-NE project

This is another case study in which techniques associated with finite-state systems are adequate for analysing a verbal safety-critical communications protocol. Through the thorough use of formal-refinement techniques, it was possible to prove formally in propositional logic that the derived system expressed in executable SPARK source code fulfilled the basic system safety requirement without exception. One could well imagine a helper computer system, installed on trains and at the train controllers' positions, which reiterated the steps of the verbal protocol as executed by the train drivers and controllers. This system is guaranteed to execute the protocol unambiguously and correctly, and therefore could have helped avoid such an outcome as at Warngau. Furthermore, Sieker's work has shown that the verbal protocol is itself logically correct in that it logically fulfils the basic safety requirement.

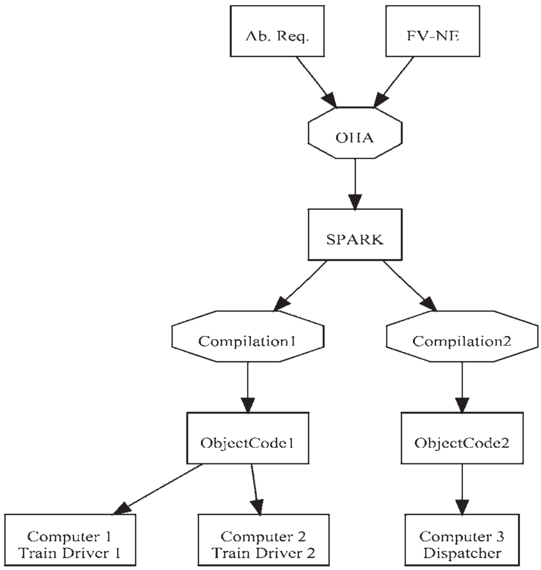


Figure 8. The Development-Flow Graph of the SPARK verbal-protocol simulation.

What is most remarkable about this development in the context of this survey of communication in critical systems is how far the technology has come from the tools generally used by computational linguists.

5. Rigor and accuracy in verbal protocols: General conclusions

We have considered verbal safety-critical communications in transportation systems, namely air and rail travel. This domain forms a small but, it has been argued, clearly delineated subspecialty of safety-critical system science at the borders of general and computational linguistics.

We have seen how techniques from finite-state system science can profitably be used to analyse, as well as in the last instance to help execute, the protocols required for these communications in order to enhance safety.

Two case studies from air travel have been considered in detail, and some safety issues effectively addressed by design principles have been discussed. However, considering work by Cushing as well as extensions by Ellermann, Hilbert and others has shown just how far we have to go in this area to reach a state-of-the-art computer-supported system which would enhance safety in these critical communications by eliminating or mitigating features which have been shown to compromise that safety.

The requirements and protocols of train dispatching are much simpler. In this case, helper systems are already deployed (state of the practice) in Austria, derived from work of Stadlmann et al. But his methods are not based on linguistic analysis, and a state of the art development based on careful linguistic analysis has been exhibited by the research work of Sieker, described broadly above.

All these methods, with the possible exception of those of Stadlman, employ mainly finite-state system techniques.

It has been remarked how far the system developed by Sieker lies from the typical considerations of computational linguistics. One may speculate that, as verbal-protocol systems become more well understood, the phenomena of interest to computer linguists will recede into the background, to be supplanted by standard techniques of reliable- and safe-SW-system development as known to SW safety system engineers. For indeed we should wish to control these phenomena, as Cushing and the case studies have shown. It has been suggested that general techniques are emerging with which we may profitably continue to go about this task: simple abstract data-structures representing state; finite-state-machine techniques; formal refinement; simple formal languages based on the ontology of predicate logic (maybe sorted predicate logic); meaning postulates to relate levels of refinement; and of course formal proof of properties of simulation.

6. References

Aeronautica Civil, Cali
1996 *Aeronautica Civil of the State of Colombia, Aircraft Accident Report, American Airlines Flight 965 near Buga, Colombia, 20 December, 1995.* Available from http://www.rvs.uni-bielefeld.de/publications/compendium/incidents_and_accidents/cali_american_airlines_b757.html

Barnes, John
2003 *High-Integrity Software: The SPARK Approach to Safety and Security.* London: Addison-Wesley.

BFU
2004 *Bundesstelle für Flugunfalluntersuchung / German Federal Bureau of Aircraft Accidents Investigation, Investigation Report AX001-1-2/02 (the midair collision near Überlingen, Germany on 01 July 2002), May 2004*. Available in English and German from http://www.bfu-web.de

Brinkmann, Christian and Oliver Lemke
2003 *Analysis of the Brühl Railway Accident*. Invited talk at the 2nd Bieleschweig Workshop on Systems Engineering, 1–2 July 2003, Braunschweig. Slides available from http://www.tu-braunschweig.de/ifev/veranstaltungen/bieleschweig/bieleschweig2/

Cushing, Steven
1994 *Fatal Words*. Chicago: University of Chicago Press.

Döring, André, Mark McGovern and Jan Sanders
2001 *Computational Analysis of Cockpit-Voice-Recording Transcripts*. RVS Technical Note RVS-Occ-01-07, Faculty of Technology, University of Bielefeld, 11 November 2001. Available at http://www.rvs.uni-bielefeld.de/publications/Papers/doering-01-07.pdf

Ellermann, Martin and Mirco Hilbert
2001 *Review of the Cushing Grammar*. RVS Technical Note RVS-Occ-01-02, Faculty of Technology, University of Bielefeld, 23 July 2001. Available at http://www.rvs.uni-bielefeld.de/publications/Papers/hillermann-critique.pdf

Ellermann, Martin and Mirco Hilbert
2002 *Building a Parser for ATC Language*. RVS Technical Note RVS-Occ-01-05, Faculty of Technology, University of Bielefeld, 18 February 2002. Available at http://www.rvs.uni-bielefeld.de/publications/Papers/hillermann-parser.pdf

FAA ATC Phraseology
2010 *Federal Aviation Administration of the United States of America, Order JO 7110.65T, Air Traffic Control, February 11, 2010*. Available from http://www.faa.gov/documentLibrary/media/Order/ATC.pdf

Gibbon, Dafydd
1995 *The Parable of Next Thursday, 12 November 1995*. Available at http://coral2.spectrum.uni-bielefeld.de:/~gibbon//nextthur.html

Gibbon, Dafydd and Peter Ladkin
1996 *Comments on Confusing Conversation at Cali, 7 February 1996*. Available at http://www.rvs.uni-bielefeld.de/publications/Incidents/DOCS/Research/Rvs/Misc/Additional/Reports/cali-comment.html

Hölz, Oliver and Thomas Hettenhausen
2001 *Building a Corpus for Cockpit-Voice-Recorder Transcripts*. RVS Technical Note RVS-Occ-01-06, Faculty of Technology, University of Bielefeld, 23 October 2001. Available at http://www.rvs.uni-bielefeld.de/publications/Papers/hettenhoelzreport.pdf

IEC 61508
2010 *International Electrotechnical Commission, International Standard 61508, Functional Safety of Electrical/Electronic/Programmable-Electronic Safety-Related Systems*. 2nd Edition. Available at http://www.iec.ch/functional-safety/

Ladkin, Peter Bernard
2002 *ACAS and the South German Midair*. RVS Technical Note RVS-Occ-02-02, Faculty of Technology, University of Bielefeld, August 2002. Available at http://www.rvs.uni-bielefeld.de/publications/Reports/ACAS-Midair-www.html

Ladkin, Peter Bernard
2004 Causal Analysis of the ACAS/TCAS Sociotechnical System. In: *Keynote Paper at the 9th Australian Workshop on Safety-Related Programmable Systems*, volume 47 of Conferences in Research and Practice in Information Technology. Available at http://crpit.com/abstracts/CRPITV47Ladkin.html as well as http://www.rvs.uni-bielefeld.de/publications/Reports/SCSS04.pdf

Ladkin, Peter Bernard
2009 Securing the Interface: Safety-Critical Interaction Between Humans and Robots. In: *Extended Abstract of Keynote Talk, 4th IET System Safety Conference*, 2009.
Ladkin, Peter Bernard
2010 *Fully Automatic Execution of Critical Manoeuvres in Airline Flying*. Available at http://www.abnormaldistribution.org/2010/09/03/fully-automatic-execution-of-critical-manoeuvres-in-airline-flying/
Ladkin, Peter Bernard and Stefan Leue
1995 Interpreting Message Flow Graphs. *Formal Aspects of Computing* 7(5):473–509. Available through http://www.rvs.uni-bielefeld.de/publications/abstracts.html#FACMSC
Ladkin, Peter Bernard, Jan Sanders and Bernd Sieker
2012 *Safety Analysis of Computer-Based Systems*. Berlin: Springer, forthcoming.
Lamport, Leslie
1995 TLA in Pictures. *IEEE Transactions on Software Engineering* 21(9): 768–775. Available through http://research.microsoft.com/en-us/um/people/lamport/pubs/pubs.html
Leveson, Nancy
1995 Safeware. London: Addison-Wesley.
Rasmussen, Jens
1997 Risk Management in a Dynamic Society: A Modelling Problem. *Safety Science* 27(2/3):183–213.
Sassen, Claudia
2005 *Linguistic dimensions of crisis talk: formalising structures in a controlled language*. Amsterdam etc.: Benjamins,
Sieker, Bernd Manfred
2010 *Systemanforderungsanalyse von Bahnbetriebsverfahren mit Hilfe der Ontological-Hazard-Analysis am Beispiel des Zugleitbetriebs nach FV-NE (System Requirements Analysis of Railway Operations with Ontological Hazard Analysis: An Example of Train Dispatching in German Non-State-Railway Law)*. Doctoral Dissertation, Faculty of Technology, University of Bielefeld. Available through http://bieson.ub.uni-bielefeld.de/volltexte/2010/1782/
Stadlmann, Burkhard and Helmut Zwirchmayer
2004 Einfaches Zugleitsystem für Regionalstrecken, *Signal+Draht* 96(6), 6–12.
Stuphorn, Jörn, Bernd Sieker and Peter Bernard Ladkin
2009 Dependable Risk Analysis for Systems with E/E/PE Components: Two Case Studies. In: Chris Dale and Tom Anderson (eds.), *Safety-Critical Systems: Problems, Process and Practice, Proceedings of the Seventeenth Safety-Critical Systems Symposium, Brighton, UK, 3–5 February 2009*. London: Springer. Also available at http://www.rvs.uni-bielefeld.de/publications/Papers/StupSiekLadSSS09.pdf

13. Multimodal Communication

Jens Allwood and Elisabeth Ahlsén

1. Why is the topic of multimodal communication interesting?

Face-to-face communication is the evolutionary primary and most common way of communicating for humans. It is not unlikely that vocal-verbal language with more or less intentional control, but still to some extent systematic, developed from being action- and gesture-based to being vocally based (Arbib 2005). In contrast to the hundreds of thousands of years of natural development of face-to-face communication, communication technologies have been artificially developed in order to supplement, enhance or replace certain types of face-to-face communication and today we have a situation of steadily increasing use of communication technology with possibilities for using multiple modalities. In this chapter, we will give an account of human-to-human multimodal communication, in order to have a background for a discussion of the use of multimodality in new communication technology (see also Lücking and Pfeifer as well as Gibbon both in this volume). We will primarily deal with face-to-face communication, rather than multimedia in relation to written communication. First, we turn to a brief overview of human-human communication in order to see what features can potentially be used in human-computer interaction in general and more specifically in dialog and tutoring systems. Then, we will discuss some problems in making computer based systems multimodal and thereby more human like.

2. Human-human communication

2.1. What is multimodal communication?

The word "modal" has the fairly abstract meaning "pertaining to manner or mode" (e.g., Collins English Dictionary 2009), so it is not surprising that the term "multimodal communication" has been used in many different ways. For example, it has been used for the three phenomena that might more appropriately be called "multi-medial communication", "multi-representational communication" and "multi-mass-medial communication" (cf. Allwood 2008). Here we will take "multimodal communication" to mean communication involving more than two of the sensory modalities (sight, hearing, touch, smell and taste). For practical purposes, we will also, besides "sensory or perception modalities", talk about "production modalities", by which we mean human bodily means that normally produce information for the different sensory modalities, that is, gestures, speech organs etc. Thus, gestures can produce information for the visual modality and the speech organs can produce speech sounds for the auditory modality. By "multimodal information" we mean information for several sensory modalities. Below, we will now consider some of the relevant features of multimodal communication under the four headings production, reception, interaction and context.

2.2. Production

In face-to-face human-human communication, a number of production modalities are used in interaction. From a linguistic-communicative point of view, they can be divided into:

(i) *Communicative body movements* (other than movements of speech organs), for example, gestures, facial expressions, body posture, gaze direction etc.,
(ii) *Aspects of the system of a language, which to some extent are common to speech and writing*, such as *phonemes* (to some extent corresponding to *graphemes*) *morphemes, lexemes (words)* and *phrases*, which make up the store and basis for the structure of meaningful units to be used in the language, *syntax* (word order), *semantics* and *pragmatics* (the meaning and use of language).
(iii) *Aspects of the system of a language*, such as *prosody* (variations in pitch, intensity and duration carrying information concerning word identification, emphatic stress and information structure, attitudes and emotions), *which are usually not directly captured in written language*. This, of course, is not to deny that information structure, attitudes and emotions in writing can be expressed by other means, such as bold face, underscoring, capitals, smileys or by non-prosodic means common to speech and writing, such as word order or rhetorical strategies.

The different communicative expressions that are produced can have different status with regard to what type of semiotic or representational sign relation they (or their various features) are based on (cf. Peirce 1931).

An *index* is a sign based on contiguity (closeness in time and space and by extension causality) to what it refers to. In this sense, clouds are an index of rain and a pointed index finger is an index of what it points to. An *icon* is, in addition to an indexical relation, based on a similarity relation, for example, photos and diagrams are icons, while a *symbol*, in addition to an indexical substratum, has an arbitrary (conventional only) relation to what it refers to. Most of the verbal vocal or written communication we engage in is mainly symbolic. But icons are, for example, used in the picture-like signs found in many interfaces to computers and cell phones. Indexical communication can, for example, be seen in arrows for pointing or so called deictic expressions that depend on their context, like *I, you, he, here, there and now.*

Communicative expressions can also have different status with regard to the degree of intentionality and awareness involved in their production: *indicate*, *display* or *signal* (cf. Allwood 1976). When we *indicate* something, the signs we are producing are not intentional, but are still informative and convey something; blushing, for example, can convey nervousness. When we *display* something, we show it intentionally. A smile, for example, can display friendliness. A pointing gesture is a display of what it points to. Thus, indices (in the Peircean sense) do not need to be indicated, they can also by displayed, that is, a smile can be an indicated unintentional and perhaps unaware expression of friendliness, but it can also be a displayed, more intentional expression of the same emotion. An interesting question for multimodal communication research is to try to find out whether people react differently to indicated and displayed smiles and whether observable, registrable differences exist between them and could be picked up not only by humans but also by automatic sensors.

Most of what has traditionally been studied by linguists has, however, been focused on expressions that are *signaled* (i.e., shown to be shown). Typically this can be achieved by using spoken or written words, phrases and sentences or other types of communication with the help of symbols (i.e., signs with an arbitrary relation to their content – cf. above). The three types of representational relation (index, icon and symbol) and the degrees of intentionality and awareness (indicate, display and signal) provide us with two different ways of distinguishing what has been called "natural meaning" from "non-natural meaning" (cf. Grice 1957).

1. The first way suggests that natural meaning is provided by expressions (signs) that have a motivated non-arbitrary representational relation that is primarily motivated by contiguity or similarity (indices and icons). Non-natural meaning is provided by expressions (signs) that are not primarily motivated by contiguity or similarity, but by arbitrary convention (symbols).
2. The second way instead suggests that natural meaning is provided by expressions that are not exposed intentionally (indicative behavior). Non-natural meaning is manipulable, that is, produced by expressions that are expressed with some intentionality (displayed and signaled behavior). The two ways of drawing the distinction are not equivalent, even if there is an overlap. Indicated indices are natural and signaled symbols are non-natural in both variants of the distinction, while the remaining categories shift their belonging depending on the criteria the distinction is based on.

In face-to-face communication, the three types of representational relation and the three different degrees of intentionality and awareness often occur in a mixed or simultaneous fashion, which makes it possible, at the same time, to communicate factual main messages, attitudes and emotions as well as to manage your own communication and the interaction. In this way, a great deal of the indicated and displayed communication, as well as some of the signaled communication, is mainly conveyed by other modalities than spoken words, phrases and sentences (Allwood 1976, 2008). We can, for example, smile, nod and speak, using both words and prosody at the same time.

The most common combinations of the three types of basic semiotic relations (indexical, iconic and symbolic) with the three degrees of intentionality and awareness (indicate, display and signal) are illustrated in Table 1 below. In principle, all the combinations are possible; the table only shows the most common combinations. Since all types of signs are based on causal relations, they are in this sense indexical.

Some examples of the other combinations are the following. If a smile is an unintentional automatic reaction, it is an indicated index. If it is expressed intentionally to show a type of attitude, it is a displayed index. If it is expressed with the intention that it be recognized as a display, it is a signaled index. This could occur, for example, when a friendly smile is a sign to an accomplice to initiate some action.

A sign is an indicative icon, when a likeness is produced accidentally: for example, I look sad because a grain of sand blew into my eye and made my eyes water. A sign is a displayed icon when the likeness is produced intentionally (e.g., a painted picture). A sign is a signaled icon when the intention to produce a likeness is intended to be recognized. This occurs very often with gestures when they are used to illustrate what is being said.

A sign can be an indicated symbol if it is used unintentionally, for example, spoken words in language X while sleeping might indicate a relation between language X and the sleep talker, for example, that the sleep-talker knows language X. This indicated connection can also be displayed, if I use a language where I have learned an utterance or two by heart, without understanding the meaning, in order to show my connection with the language (see Searle 1970).

Finally, a symbol may be signaled, which is its most common use, when we use words with the intention that it should be recognized that we are communicating something.

Table 1. Basic semiotic relations and degrees of communicative awareness and intentionality.

	Index	Icon	Symbol
Indicate	X		
Display		X	
Signal			X

If we compare *Human-Human Communication* (HHC) with *Human-Machine Communication* (HMC), it could perhaps be claimed that machines do not have awareness and intentions, they can only indicate, and that for this reason only humans exhibit the full range of variation in awareness and intentionality shown in Table 1. Even if this is correct, machines can, however still exhibit variation when it comes to employment of the basic semiotic relations of representation, that is, they can communicate with their users using indices (e.g., activated flashing lights), icons (e.g., pictures) and symbols (e.g., words) or combinations of all three relations.

2.3. Reception

Reception, just like production, includes processes on different levels of control (intentionality) and awareness. One reason for this is that human-human communication probably contains an element of what is often called "mirroring", that is, a kind of automatic reproduction or activation of the communicative actions of the interlocutor, either internally or externally visible (Ahlsén 2008; Le Bel et al. 2009). This means that all the aspects of production described above can be relevant for reception. This, in turn, means that reception can take place on different levels of conscious control or intentionality. What is indicated (and sometimes what is displayed), by body movements and the words and sentences produced, is often received and reacted to more or less *subconsciously.* This includes many aspects of emotions and attitudes as well as of communication management.

More generally, reception includes several degrees of processing from *subconscious reactions* to *perception* characterized by more conscious discrimination and identification, that is, hearing or seeing that certain signs are produced.

The receptive processes can also lead to *understanding* in which incoming information is *interpreted* in relation to the recipient's own stored background, for example, the communicative context and the activity context. One example of how a message is understood, interpreted and reacted to attitudinally is the recipient's *impression of the trustworthiness of* the person producing the message. This impression often has an influence on the effect the understood content has on the recipient (Komiak et al. 2004; Ruttkay and Pelachaud 2004).

In face-to-face communication *words, gestures and prosody* are all interpreted in interaction between two or more people and it is, thus, the multimodal totality of a contribution that is normally received, interpreted and reacted to. Let us therefore briefly consider this interaction.

2.4. Interaction

An important aspect of human-human face-to-face communication (as well as human-computer communication) is how the interaction between communicators is managed. The study of *Interactive Communication Management (ICM)* is the study of different systematic means for this. Often, interactive communication management makes use of different modalities that help to manage the progression of successful interaction.

The *turn management* system helps interlocutors manage the distribution of their contributions to the interaction. When a participant has the turn, he or she has the right to contribute by speech or by some other modality, that is, has "the floor". Participants manage turn distribution, for example, by showing (i.e., indicating, displaying or signaling) when they want to speak, when they accept an invitation to speak, when they want to continue, when they want to stop and often to whom they want to give the next turn (Sacks, Schegloff and Jefferson 1974). Turn distribution can be signaled, for example, in a formal meeting, where a chairman distributes the turns explicitly by means of names or gestures or by questions directed to a specific, named person (*What do you think, Bill*?). Turn management, however can also be less consciously achieved by changes in body movement, gesture, voice quality or facial expression.

The *feedback* giving system (Allwood, Nivre and Ahlsén 1992) helps interlocutors to express ability and willingness with regard to contact, perception, understanding and attitudinal reactions in relation to what is being communicated. This is done continuously, usually mainly by small, unobtrusive contributions, such as head nods, head shakes, smiles, and words like, *yes*, *no* and *mm*. These reactions from the listener guide the speaker concerning whether he or she can go on (continued contact), whether the communication is perceived, understood and concerning the attitudinal and emotional reactions of the interlocutor. When there is a need, feedback can also be elicited by special communicative means, such as use of tag questions or question intonation.

A third mechanism for interactive communication management is *sequencing*. Contributions often occur in fairly set sequences, Such sequences extend from "exchange types", sometimes also known as "adjacency pairs" (e.g., Levinson 1983), such as "question-answer" or "greeting-greeting" and preference organization (e.g., Pomerantz 1984), where a certain type of contribution activates a preference for a particular response among a certain selection of possible responses, for example, after an offer it is often preferred to express gratitude, to "scripts" identifying what is typically said in a specific type of activity, like a restaurant or a travel agency. This structure helps structure how the participants can interact (Shank and Abelson 1977).

2.5. Context

A relevant issue for how and when multimodal communication strategies can be used in human-human as well as in human-computer interaction is the *activity dependency* of the communication. Some aspects of human-human communication are relatively stable across different social activities, whereas other aspects show considerable variation and activity dependence (cf. Allwood 1995, 2000). This activity dependence also applies to multimodality.

In *most ordinary face-to-face communication* (also in distance video communication), multimodality is present in an integrated way. This applies to communication in everyday social activities, work, leisure activities and games as well as to communication through avatars in virtual reality environments etc.

Different communication functions can be more easily handled by one modality rather than by another. Depending on context, gestures, map drawing or diagrams, for example, can be used for giving explanations or instructions concerning how to find the way (e.g. Kopp et al. 2007; Tversky et al. 2008) and spoken words or other sounds can be used for alerting someone quickly.

Different social activities provide different contextual options for multimodal communication and in ordinary human-human communication, the participants more or less automatically adapt to these options. They communicate with actions, objects, tools, gestures, speech etc., according to what is possible and efficient (cf. Allwood 1995, 2000) in a particular activity context.

In communicative contexts where communication to a great extent is handled by *other modalities than spoken or written words*, multimodality is essential, for example, when movements need to be visualized by gesturing.

In a number of contexts, where multimodality is necessary, *flexibility in the choice of modality is needed*. This includes contexts where specific modalities are temporarily or permanently blocked. Examples of this are voice-only communication in phone calls, where the visual modality cannot be used or when sensory modalities are blocked as when one or more participants are blind. In communication with deaf persons, the auditory modality is blocked, so communication needs to be visual (see Kubina and Abramov in this volume). When a person's hands are occupied by other actions, communication needs to employ the auditory channel. In contexts such as these, one modality needs to compensate for loss of communication in another modality. Many of the contexts where communication needs to be multimodal are, thus, *communicatively challenging*, for example, because the participants do not share the same language and/or cultural background or because one of the participants has a communication disorder. Flexibility in choice of modality and the possibility of *redundancy* by using several modalities in interaction is often needed. Also when human-human communication takes place in *complex communication environments*, for example, in technically advanced activities, such as controlling airplanes, nuclear power plants etc., enhanced efficiency makes multimodality involving the *distribution of communication over several modalities* necessary (Goodwin and Goodwin 1997).

3. Which features from multimodal human-human communication (HHC) are relevant for human-machine communication (HMC)?

One answer to this question is that all or most features of multimodal human-human communication are or will over time become increasingly relevant for human-machine communication. However, not all of the multimodal HHC features are relevant for all HMC systems. In the past, the dominating view has been that HMC is so different from HHC that perhaps it works more efficiently if it does not try to model all of the complexity of HHC (cf. Becker 2006). But this view is changing. Even though the complexity of HHC is considerable and not completely understood, described and explained, the rapid technical development enabling more complexity and multimodality, and the demands relating to usability and accessibility for all, together with the increasing dependency on HMC for all kinds of everyday communication activities drive development in the direction of trying to make HMC more HHC-like in general and in this, to make use of the available possibilities for multimodal communication strategies.

A more cautious answer than the one just given is that, although all or most features of HHC are relevant, some of them are probably more important than others. Above all, speech is better understood than other modes of production (cf. Partan and Marler 2005) and might for this reason come into more general use.

A further aspect is that the quality of the features of multimodal communication implemented in HMC systems can vary considerably and that there is often a threshold effect, so that a certain quality has to be fairly fully achieved in order to make a particular communication strategy more useful than disturbing. For some features, however, rough estimations can be sufficient for some communicative purposes, whereas other features need to be fine tuned in order to be useful. A friendly smile can work for many purposes, while a rough estimation of synchronization of lip movements and speech sounds can be more disturbing than helpful.

A complication is that in order to achieve a reasonable degree of "naturalness" in multimodality, the coordination and integration of different features has to be good and this is still not easy to achieve. (cf., however, Oberzaucher et al. 2008; Boukricha et al. 2009). The goals of a "natural" implementation are, for example, to make a virtual *Embodied Communicative Agent* (ECA; see also Kopp and Wachsmuth as well as Martin and Schultz both in this volume) have movements that are smooth and coordinated and do not violate any of the constraints on movement in humans, so that, for example, lip movements and speech sounds should be perfectly synchronized. Facial expressions should also be fine tuned and have many variations, which then must be generated in coordination with the spoken message, since small deviations can be very distracting for the user. Some of the more abstract goals that have been expressed for the design of ECAs are to achieve “addressability”, trust, “personality”, “believability”, “naturalness” and “flow of communication”. All of these goals have to be seen in a long term perspective since, in many cases, it is not clear how they are achieved in human-human, face-to-face communication and even less how they can be achieved in human-machine communication.

Let us now consider some of the features of multimodal Human-Human Communication that are relevant for Human-Machine Communication and therefore are important to consider in further research and development.

3.1. Activity dependency

Since human-human communication shows rapid adaptation to differences between social activities, this ability is also in a long-term perspective important to achieve in human-computer communication. If the ECA is supposed to be involved in more than one activity type, it should be able to flexibly adapt to a new activity type, including the multimodal strategies that would be expected and beneficent in HHC. If the ECA is only intended for one activity type, it could, of course be designed or adapted for only this activity. Achieving this kind of flexibility in HMC involves analyzing the relevant influencing activity factors (e.g., purpose, roles, instruments and environment – cf. Allwood 2000, 2007) and implementing the relevant type of multimodal behavior that they are connected with.

3.2. Cultural variation

Besides differences between activities, differences between cultures are often also relevant. Most societies of today are multicultural and agents on the Internet get involved in communication with humans who have widely different cultural backgrounds. So, like with activities, if the ECA is supposed to be involved in more than one culture, it should be able to flexibly adapt its communication to the culture at hand, including the relevant multimodal strategies. For example, Jan and Traum (2007) and Allwood and Ahlsén (2009) present models with parameters that can be varied for different cultures for types of conversational behavior, such as proxemics, gaze and overlap in turn taking. These parameters can be set for a specific culture. If only one culture is involved, the ECA should be adapted to the more specific and local purposes that are relevant in this culture.

3.3. Flexibility and adaptability in choice of modality

Accessibility for all or most people to the services provided through an ECA also involves the need to adapt the system to the needs of users who can perhaps not make use of all modalities (e.g., blind users, deaf users). It is, thus, necessary to have the possibility to present information in a way that compensates for this. Like with the adaptation to activity and culture discussed above, flexibility and the possibility to choose and set different parameters for multimodal strategies are required. Although strict "translation" between different sensory modalities is not often possible, the possibility to communicate with a system via speech, writing, or sign language means a substantial increase in accessibility. Attempts at rendering the content of pictures in web pages via text, which is presented as speech, are today obligatory in many contexts. When an ECA, for example, is pointing to something on the screen, like a piece of merchandise or a map of a shop, this information has to be accessible also in an auditory (text to speech) version. Another challenging task is to use alternative modalities more creatively. For example, (non-speech) sounds of different intensity, duration and "corresponding" variations in tactile interfaces can be used for presentation of otherwise visual information for persons with visual impairment and an increase in picture and auditory support for text presentation can be used for persons with reading difficulties. The use of systems with multiple combinations of modalities for presenting different types of information to user groups with and without specific disabilities is an interesting area for research in cognitive science, communication and information technology (cf., e.g., Caporusso, Mkrtchyan and Badia 2010).

3.4. Body communication on different levels of awareness and intentionality (gestures, body posture, eye gaze, facial expression etc.)

Besides vocal-verbal, written information and pictures, communicative body movements are the most important type of multimodal communicative behavior in both HHC and HMC. Communicative body movements are often coordinated with prosody, so special attention is needed with regard to prosody in speech. Other challenges are posed by the need for integration of modalities, and for being able to handle this integration, both in the recognition and understanding of multimodal information (sometimes known as "fusion") and in the production and distribution of multimodal information (sometimes known as "fission"), in this way, making multimodal input and output available to HMC systems.

Some of the most important communicative body movements and other means of expression that need to be modeled both for production and reception of communicative behavior include the following (see also Allwood 2002):

– *head movements* (forward movement, backward movement; both nods and jerks, shakes; both left turn and right turn, forward movement, backward movement, tilts)
– *facial gestures* (smiles, frowns, wrinkle, mouth movements other than speech)
– *direction of eye gaze and mutual gaze*, eye movements
– *eye brow movements*
– *pupil size*
– *lip movements*
– *posture and posture shifts*
– *arm and hand movements*
– *shoulder movements*
– *movements of legs and feet*
– *spatial orientation movements*
– *clothes and adornments*

For speech, intonation or prosody (measured and observed through variations in intensity, pitch and duration) are very important. For example, prosody often provides information like which part of the utterance is in focus and/or provides new information. Compare, for example, *It is an old WOMAN* (as opposed to *man*) and *It is an OLD woman* (as opposed to *young*). Prosody also often identifies the type of communicative act, for example, if the utterance is a statement or a question. Compare the utterance *It is raining*, pronounced with a statement or question intonation. Nonverbal sounds also have certain communicative functions. Examples of nonverbal sounds with communicative functions are laughter or sighs and smacking sounds, which can both, for example, indicate or display a reaction of boredom or dissatisfaction in relation to another person's utterance. All of these also provide interesting challenges for automatic recognition in automatic systems.

3.5. Robustness

Multimodality can contribute to making interactive systems increasingly more able to handle *Own Communication Management* (OCM) (cf. Allwood 2007), that is, communication processes that enable choice of what is to be communicated and change of what has been expressed (sometimes also called repair, editing or correction), and *Interactive Communication Management* (ICM), that is, communication processes that enable turn-taking, feedback and sequencing. The system also needs to handle difficulties in text understanding, difficulties in speech recognition and difficulties in picture/gesture recognition in a sensible way. One of the hopes in constructing multimodal systems is that robustness could be increased and that some of these difficulties could then be handled by making use of the assumed redundancy of multimodal systems, that is, since similar and closely related information is expressed in several modalities, failure of recognition in one modality could be compensated by recognition in another modality and the two data streams can then be integrated into a more holistic multimodally based recognition.

3.6. Interaction and interactive features

In order to handle interactive communication in a robust manner, the system has to be able to flexibly handle both own communication management and interactive communication management and these functions need to be potentially adaptable to different activities and contexts. The system, thus, needs to handle:

– *Turn management*

How do we indicate, display or signal that speaker change is about to occur? Is it OK to interrupt other speakers? If so, when should interruptions occur? How long should the transition time be from one speaker to the next speaker? Is it OK to do nothing or be silent for a while? What can the speaker or the system do to keep a turn? How can they signal that they don't want the turn, but rather want the other party to continue? (Sacks, Schegloff and Jefferson 1974; Allwood 1999; De Ruiter, Mitterer and Enfield 2006; Magyari and De Ruiter 2008).

– *Feedback*

How do speakers indicate, display and signal to each other that they can/cannot perceive, understand or accept what their interlocutor is communicating (cf. Allwood 2002). Is this done primarily by auditory means (small words like *mhm, m, yeah* and *no*) or by visual means (head nods, head shakes, posture shifts etc.)? What emotions and attitudes do primarily occur in giving and eliciting feedback? Is very positive feedback preferred or is there a preference for more cautious feedback? (See Kopp et al. 2008.)

– *Sequencing*

What opening, continuing and closing communication sequences are preferred in a particular activity or culture, for example: What is the preferred way of starting an interaction in different activities (opening sequence)? What is the preferred way of closing (closing sequence)? When and how are greetings used? (See also Allwood et al. 2006.)

In studying naturalistic interaction, the role of "mirroring", imitation, automatic alignment, priming, contagion etc., that is, relatively automatic coordination of behavior between interactive individuals has received considerable attention in recent years (cf. Wachsmuth, Knoblich and Lenzen 2008). Making ECAs that can mirror and align requires more research on human behavior as well as design and programming of the type of perception and expressive behavior that is required for relatively automatic coordination. Work on designing such systems has, for example, been done by Grammer and Wachsmuth (cf. Oberzaucher et al. 2008).

3.7. Types of content

Different types of content require different combinations of modalities to be expressed. An emotional content in human face-to-face interaction is easier to perceive, and usually more trustworthy, if words, prosody and facial gestures reinforce each other. Information about train departure times, on the other hand, is less dependent on reinforcement by prosody and facial gestures. The situation might here be different in an interactive multimodal system, where other kinds of multimodal information could be available and information about train departure times would best be given by a visual presentation of a train table, accompanied by highlighted parts being read aloud. More generally, interactive multimodal systems, for example, ECAs, are employed for various functions and, depending on the function, to a varying degree need specific combinations of modalities. However, the quest for naturalness in interaction continuously makes the issue of how different types of content are expressed more complex. In human face-to-face interaction, all modalities are available and the possibility to oscillate between different topics and even different activities and subactivities is often utilized (Allwood 2000, 2001, 2007). So, in the process of making an ECA appear as a reliable, natural communication partner, it is important to notice that most institutional human-human interactions involve confidence promoting and social alignment strategies, such as jokes and references to personal experiences, remarks about the weather, news etc. and that multimodality plays an important role in this process (e.g., smiles, body posture, prosody, facial expressions etc., in addition to fluent speech). If we want ECAs to be as natural as possible, these elements should be included. This means that also content usually expressed by indexical and iconic communication is important and that the ECA has to be able to deal with a number of topics, other than the topic that is the focus of the actual task being pursued. It therefore has to be able to change between different domains in a flexible way. As we have seen, this, besides the often task-specific factual main message content, also includes variation concerning culture and activity. This variation can then, for example, concern the content and functional areas of:

– *Identity*. How should the body and body movements indicate, display or signal who the agent is?
– *Physiological states*. What "physiological state" should be indicated, displayed or signaled by the agent and how?

- *Emotions*. What emotions are acceptable and appropriate in different activities?
- *Attitudes.* What attitudes, for example, regarding epistemic stance, politeness or respect, are appropriate?
- *Factual information.* What information is communicated? To some extent, this question is related to what is often called "information structure". What information is explicitly verbally in focus? What information is backgrounded and presupposed (perhaps multimodally available) in a particular situation? I can point to a car and say *new brake system* or perhaps say *it has a new brake system* or perhaps without nodding or pointing *that car has a new brake system.* In another situation, I might be answering a question, *Which cars have new brake systems?* and answer *that* or *that car*, perhaps accompanied by nodding or pointing. The factual information in all these examples is on some level the same, but the way it is focused and presented differs.
- *Everyday topics.* Included in factual information, we can ask what topics are regarded as neutral and possible to address even for strangers (e.g., politics, the weather, job, income etc.)?
- *Common speech acts*. Finally, we can consider what types of speech act are the most appropriate to convey the above types of information: What types of speech acts are commonly used in different activities, e.g. greetings, farewells and other typical exchange types?
- *Communication management*. How should the different types of communication management be accomplished? (Cf. Section 3.5 and 3.6 above).

4. Types of human-machine interaction

4.1. Non-digital machines

Multimodal communication has probably always been characteristic of the human species. In paleolithic times, stones were, for example, decorated by carvings, made by cut stones and painted with colors. Images of animals, persons, objects, often guiding rites and other activities, were created. Later on, tools for imprints in clay and pens, brushes, printing, typewriters etc. were developed, as writing gradually became an important way of communicating over distances in time and space. In the last 150 years or so, communication technology for images, photography, film and video as well as telegraphy, telephony and different types of audio recording has made multimedia communication, for example in newspapers and films, possible. Finally, in the last 20–30 years, with the advent of digital computer based technology, different media channels have been combined in many different ways (see Waltinger and Breuing in this volume).

4.2. Computers

4.2.1. Text only interfaces

At first, communication with and via computers occurred using text only. This is still the case for some applications and can pose problems, like when students are not able to represent a figure in taking notes at a lecture or, perhaps more seriously, when applications are not accessible to persons with low literacy skills.

4.2.2. Interface with pointer (or touch screen)

At a more recent stage, user friendliness was increased by introducing the desktop interface, together with menus and icons as well as interfaces using pointers, for example, mouse, pen or screen pointing. This added iconic and indexical dimensions to human-computer interfaces.

4.2.3. Voice communication

Voice communication with computers, for a long time has posed and still poses a real challenge to developers and users and is, for this reason, a very critical issue in trying to achieve naturalistic HCI. Speech synthesis has progressed from monotonous, metallic sounding robot speech with poor assimilation of adjacent sounds to more natural sounding concatenated speech, based on recordings of humans. But still many problems remain in generating naturalistic speech, not least in the areas of prosody for expressing affect and information structuring. Speech recognition remains to a great extent a so far poorly solved but very challenging problem. While systems for very limited domains, such as booking trains or flights or enquiring about telephone numbers, are in use and work sufficiently well, although not perfectly, voice dictation systems still need adaptation to individual speakers and tend to produce many errors. So, better recognition of naturalistic speech by computers is an important problem, which remains a challenge in the quest for on-line speech communication between humans and computers (Edlund et al. 2008).

4.2.4. Communication with a multimodal agent

While voice communication would make human-computer interaction similar to telephone conversation, and this would be adequate for many purposes, it is still incomplete, especially concerning speech recognition, as we have mentioned above. The goal of more naturalistic HCI is to enable humans as well as computers to produce and recognize multimodality in similar ways to what humans do. For this, simulations and representations of functions of the human body are needed. The design of embodied communicative agents (ECAs) is therefore an important area for multimodal communication technology dealing with "face-to-face-like" spoken and more or less multimodal communication. It is, in fact, even more challenging than voice communication, since the task of recognizing and producing naturalistic body communication is even less developed than the production and recognition of speech.

Simple "embodied" communicative agents used as interfaces to databases, today occur fairly frequently. Most typically, such systems have a human-looking face and upper body connected to a written dialog system for enquiries about services, merchandise, prices etc. The face of the "embodied" communicative agent can often add a few facial expressions linked to key words in the written input and output. The multimodality of this type of communication is, therefore, still extremely limited. One example of this type of ECA is IKEA's interface "Anna", which comes in a few slightly varying versions, but is essentially the same in most countries, showing three possible facial expressions accompanying text answers to queries about buying furniture (cf. Allwood and Ahlsén 2009).

Examples of AI-based, more advanced screen based embodied Communicative Agents, which can handle many HHC-like aspects are REA (Cassell et al. 1999) and GRETA (De Rosis et al. 1999), MAX (Wachsmuth 2005) and GANDALF (Thórisson 1997) (cf. also André et al. 1999; see also Kopp and Wachsmuth in this volume). In such ECAs, features like emotions (based on a PDA, i.e., pleasure, dominance, arousal, model), interaction management, and gestures have been modeled.

There are a number of application areas for ECAs and different forms of multimodal communication strategies, for example, in tutorial systems for education and learning, in systems of interaction for children with autism spectrum disorders or other learning difficulties, in systems for giving various types of advice, in social care-giving services and in entertainment. Another interesting area of HMC, where human-human multimodal communication strategies are relevant and could serve the purposes of providing more communicatively efficient, naturalistic interfaces is the area of mobile dialog systems.

4.2.5. Mobile dialog systems for ECAs and robots

Mobile dialog systems mostly rely on text or menu based communication, especially on the human input side, since, as we have mentioned above, speech recognition is hard to achieve, except for in very limited domains. In recent years more and more research has therefore concerned speech interfaces, with the purpose of making the interaction with different systems more naturalistic.

The introduction and increasing use of embodied communicative agents (and robots) in dialog systems has, as we have seen, led to greater interest in phenomena related to face-to-face spoken interaction, such as interaction management and communication of emotions and attitudes. One of the consequences of this is that designers of dialog systems now strive to make systems, which for a long time have been *asymmetric with respect to initiative* (in order to facilitate the processing tasks of the system), more symmetric in this respect, so that they give more equal opportunities for initiatives to user and system (e.g., Johnston et al 2001; see below). Another consequence is that making a dialog system multimodal means that the system, like in human-human communication, should be able to make use of different modalities *in coordination* (cf. Wahlster 2003). This in turn means that in order to achieve human-human like properties, dialog systems should possess *symmetric multimodality*, so that the different modalities can be used both for input and output. An example of a multimodally symmetric system is SMARTKOM, which uses an embodied communicative agent that handles a number of spoken interaction phenomena, for example, emotional prosody, gestures and backchannel feedback. Another example is the dialog system of the mobile robot BIRON, which is based mainly on the speech modality but can augment semantic representations by hypotheses based on other modalities, for example, gestures (cf. Toptsis et al. 2004). Recent systems are moving from the "interface metaphor" to the "human metaphor" in exploiting more characteristics of human-human communication. This provides advantages concerning what the system can achieve, for example, increased naturalness and socially oriented communication, as well as challenges concerning what to achieve, for example, on-line prosodic analysis, communication management and context sensitivity etc. (Edlund et al. 2008).

Mobile dialog systems for robots, for example, the WITAS dialog system for multimodal communication with a robot helicopter, involve processing highly dynamic environments. Multimodal telecommunication systems (Lemon et al. 2003), like the MOBILTEL (Cismár et al. 2009), involve the use of different input mechanisms in a handheld device with a speech and graphical interface that includes an integrated VoIP (Voice over IP; see also Waltinger and Breuing in this volume) client as well as a pen, touch-screen, keyboard input and display including icons, emoticons, hyperlinks and scrolling menus, but without the usual HHC face-to-face features, such as gestures. Similar systems, like MATCH (*Multimodal Access to City Help*) (Johnston et al. 2001), provide a mobile speech-pen interface where the user can choose the modalities of communication, which are then integrated by the system in a speech-act based multimodal dialog manager which is symmetric with respect to initiative, that is, allows mixed-initiative dialog.

4.2.6. Tutoring systems

Another challenging and interesting area is the area of *Intelligent Tutoring Systems* (ITS) (Self 1998) or *Intelligent Learning Environments* (ILE) (Fernandez-Manjon et al. 1998). Such systems are often based on BDI (*Beliefs, Desires and Intentions*) cognitive models that traditionally have as their basic components (i) *domain knowledge*, (ii) a *user* (student) *model* and (iii) *pedagogic strategies*. Sometimes the systems have an "agent paradigm" from cognitive AI (Rao and Georgeff 1991). They can include *Multi-Agent Systems* (MAS) involving more than one user. Advanced systems for teaching in dynamic multi-agent virtual worlds also exist (e.g., Marsella and Johnson 1997). A student model can be "affective" as well as "cognitive" in a more narrow sense. Pedagogical "agents" are either modeled as cooperative agents who work in the background as part of the educational system or as personal, animated agents for human-computer interaction using voice and gesture and showing emotional attitudes (e.g., *Vincent* (Paiva and Machado 1998), *Steve* (Rickel and Johnson 1999, 2000), and *Cosmos* (Lester et al. 1997, 1999; cf. also Person et al. 2001; Vicari et al. 2009). An "agent paradigm" can be used for exploration of interaction and dynamic changes in the environment, for teaching and learning and there is an ambition to make the software more flexible in relation to the user's needs and preferences. Some features that are often needed in tutoring systems are strategies for human-computer interaction and for handling of multimedia information. To achieve a pedagogical human-computer interaction is extremely important for the teaching and learning process. Traditionally the main pedagogical functions of explanation, education and diagnosis have been implemented as a one-way mechanism, that is, the system is in control. So in tutoring systems, the ambition is to replace an asymmetric communication mode with respect to initiative with a symmetric communication mode between human and computer, including conflict solving by real negotiation, when this is needed. The BDI models, mentioned above, are today being extended with more social aspects, where, for example, expectations, confidence, planning and emotion are also modeled and interaction modes like negotiation, competition and cooperation are focused on more. This calls for more social behavior, which in turn entails the use of multimodality in communication.

We can see that dialog systems, including those used in tutoring systems, are increasingly being directed toward resembling human-human communication by including more human-like features.

5. Potential problems with human-like naturalistic ECAs for some applications

We have seen above, that there is considerable complexity and difficulty in designing naturalistic multimodal ECAs and that many tasks lie ahead of us in trying to achieve this. All the same, the research is being pursued and striving for increasing naturalness is a research agenda that is both important and fairly generally adopted. In the meantime, more limited multimodal strategies are being implemented in simplified and not always so natural artificial agents in different interfaces and applications, for example, in tutoring systems.

However, in relation to the development of "social agents", that is, agents that often have as a main purpose to provide, teach or train social interaction, it is important to consider the ethical ramifications of making ECAs optimally naturalistic. Research on social agents is described, for example, by Beazeal (2002) and Louwerse et al. (2008). According to Becker (2006), there are clear and insurmountable problems related to the goal of naturalness. In some contexts where ECAs are (successfully) used, for example, for communication with children with autism spectrum disorders (Dautenhahn 2007; Dautenhahn and Werrym 2004) or for communication with elderly persons, there are problems connected with simulating natural interaction, since it is not the same thing as providing real natural interaction. The agent might simulate emotions by mirroring etc, but actually does not have them, and this is very different from a human conversation partner. Becker especially points to the importance of eyes and voice, where there is no true multimodal symmetry between the ECA and the user. Becker poses the question of what a person is learning with respect to human-human interaction, by interaction with an ECA. The possibility of mixups of a real person, that is, skyping on the computer screen, with a very naturalistic ECA, by elderly persons suffering from dementia (which is one of the target groups for the design of supportive ECAs) can potentially lead to problems, involving over-confidence in the abilities of the ECA, which could be risky in some situations. Ethical concerns have to be addressed in using naturalistic ECAs, in some applications. As a result, Becker advocates the alternative of keeping to more limited tasks and domains where not all (or as many as possible) of the features of a human communicator are implemented. The potential ethical problems involved in using naturalistic ECAs are a controversial area where more research is needed.

6. Conclusion and outlook

The requirements on multimodal communication systems, if they are to be optimally naturalistic with respect to human-human communication, are considerable, as we have seen above in the description of features of human-human communication.

There are huge challenges for the enterprise of providing naturalistic communication. One of them is to provide automatic speech recognition which is not just limited to a restricted domain and which includes some processing of expressions for emotions and attitudes (see also Martin and Schultz in this volume). Another is the achievement of naturalistic gesturing, facial expressions and body posture as well as the recognition of these features.

Two lines of research and development are already obvious and are likely to persist in future work:

(i) the pursuit of greater understanding of natural features of human-human communication in different modalities, that is, basic research on what can then be potentially available to be modeled;
(ii) the development and implementation of application specific multimodal ECAs with limited repertoires of features that are judged to be efficient for specific purposes.

Both lines of research and development will benefit from an exchange of ideas, methods and findings. In both types of research, ethical considerations have to be taken into account. In the development of applications, research on how actual and potential users in fact respond and react attitudinally and emotionally is important for the success of the application and can probably also lead to a more realistic appreciation of the possible ethical problems by providing more specified insights and guidelines.

There is no doubt that multimodal communication technology will be increasingly used in our everyday social life, both in professional tasks and leisure activities. Especially in computer gaming, multi-modal interfaces are being developed rapidly (cf., for example, Sargin et al. 2008; Liu and Kavakli 2010) and also address the challenge of special needs (e.g., Caporusso, Mkrtchyan and Badia 2010). Finally, it is likely that in this area based on features of human-human communication, in turn, enhanced by the developments of computer- and internet-based functions, will bring about new and exciting ways of communication.

References

Ahlsén, Elisabeth
2008 Embodied communication – aphasia, apraxia and the possible role of mirroring and imitation. *Clinical Linguistics and Phonetics* 22(4–5): 1–5.

Allwood, Jens
1976 *Communication as Action and Cooperation. Gothenburg Monographs in Linguistics* 2. Gothenburg: University of Gothenburg, Department of Linguistics.

Allwood, Jens
1995 *An activity based approach to pragmatics. Gothenburg Papers in Theoretical Linguistics* 76. Gothenburg: University of Gothenburg, Department of Linguistics.

Allwood, Jens
1999 Are There Swedish Patterns of Communication? In: Hiroshi Tamura (ed.), *Cultural Acceptance of CSCW in Japan and Nordic Countries*, 90–120. Kyoto: Kyoto Institute of Technology.

Allwood, Jens
2000 An Activity Based Approach to Pragmatics. In: Harry Bunt and Bill Black (eds.), *Abduction, Belief and Context in Dialogue: Studies in Computational Pragmatics*, 47–80. Amsterdam: John Benjamins.

Allwood, Jens
2001 Capturing Differences between Social Activities in Spoken Language. In: Istvan Kenesei and Robert M. Harnish (eds.), *Perspectives in Semantics, Pragmatics and Discourse*, 301–319. Amsterdam: John Benjamins.

Allwood, Jens
2002 Bodily Communication – Dimensions of Expression and Content. In: Björn Granström, David House and Inger Karlsson (eds.), *Multimodality in Language and Speech Systems*, 7–26. Dordrecht: Kluwer Academic Publishers.

Allwood, Jens
2007 Activity Based Studies of Linguistic Interaction. *Gothenburg Papers in Theoretical Linguistics* 93. Gothenburg: University of Gothenburg, Department of Linguistics.

Allwood, Jens
2008 Dimensions of Embodied Communication – towards a typology of embodied communication. In: Ipke Wachsmuth, Manuela Lenzen and Günter Knoblich (eds.), *Embodied Communication in Humans and Machines,* 257–284. Oxford: Oxford University Press.

Allwood, Jens, Elisabeth Ahlsén., Johan Lund and Johanna Sundqvist
2006 Multimodality in Own Communication Management. In: *Proceedings from the Second Nordic Conference on Multimodal Communication. Gothenburg Papers in Theoretical Linguistics* 92(2): 10–19. Gothenburg: University of Gothenburg, Department of Linguistics.

Allwood, Jens and Elisabeth Ahlsén
2009 Multimodal Intercultural Interaction and Communication Technology – A conceptual framework for designing and evaluating Multimodal Intercultural Communicators. In: Michel Kipp, Jean-Claude Martin, Patrizia Pag-

gio and Dirk Heylen (eds.), *Multimodal Corpora*, 160–175. LNCS 5509. Berlin: Springer.

Allwood, Jens, Joakim Nivre and Elisabeth Ahlsén
1992 On the Semantics and Pragmatics of Linguistic Feedback. *Journal of Semantics* 9: 1–26.

André, Elisabeth, Thomas Rist and Jochen Müller
1999 Employing AI methods to control the behavior of animated interface agents. *Applied Artificial Intelligence* 13: 415–448.

Arbib, Michael
2005 From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. Behavioral and Brain Sciences 28(2): 105–124.

Beazeal, Cynthia L.
2002 *Designing Sociable Robots*. Cambridge, Massachusetts: The MIT Press.

Becker, Barbara
2006 Social robots – emotional agents: Some remarks on naturalizing man-machine interaction. *International Review of Information Ethics* 6(12): 37–45.

Boukricha, Hana, Ipke Wachsmuth, Andrea Hofstätter and Karl Grammer
2009 Pleasure-Arousal-Dominance driven facial expression simulation. In: *3rd International Conference on Affective Computing and Intelligent Interaction, ACII 2009*, 119–125. Amsterdam: IEEE Press.

Caporusso, Nicholas, Lusine Mkrtchyan and Leonardo Badia
2010 A multimodal interface device for online board games designed for sight-impaired peope. *IEEE Transactions Information Technology Biomedicine* 14(2): 248–254.

Cassell, Justine, Joseph Sullivan, Scott Prevost, Scott and Elizabeth Churchill (eds.)
2000 *Embodied Conversational Agents.* Cambridge, Massachusetts: The MIT Press.

Cismár, Anton, Matús Pleva, Ján Papaj, Lubomir Dobos and Jozef Júhar
2009 MOBILTEL Mobile multimodal telecommunications dialogue system based on VoIP telephony. *Journal of Electrical and Electronics Engineering* 2: 134–137. *Collins English Dictionary – Complete and Unabridged 10th Edition*
2009 Fulham Palace: Harper Collins Publishers.

Dautenhahn, Kerstin
2007 Socially intelligent robots: dimensions of human – robot interaction. *Philosophical Transactions of the Royal Society B: Biological Sciences* 362(1480): 679–704.

Dautenhahn, Kerstin and Iain Werrym
2004 Towards Interactive Robots in Autism Therapy: Background, Motivation and Challenges. *Pragmatics and Cognition* 12(1): 1–35.

DeRosis, Fiorella, Catherine Pelachaud, Isabella Poggi, Valeria Carofiglio and Berardina DeCarolis
2003 From Greta's mind to her face; modeling the dynamics of affective states in a conversational embodied agent. *International Journal of Human-Computer Studies* 58(1–2): 81–118.

De Ruiter, Jan Peter, Holger Mitterer and N. J. Enfield
2006 Projecting the end of a speaker's turn: A cognitive cornerstone of conversation. *Language* 82(3): 515–535.
Edlund, Jens, Joakim Gustafson, Mattias Heldner and Anna Hjalmarsson
2008 Towards human-like spoken dialogue systems. *Speech Communication* 50(8–9): 630–645.
Fernandez-Manjon, Baltasar, Juan Cigarran, Antonio Navarro and Alfredo Fernandez-Valmayor
1998 Using automatic methods for structuring conceptual knowledge in intelligent learning environments. In: *Proceedings of Intelligent Tutoring Systems, 4th International Conference, ITS' 98 San Antonio, Texas, USA, August 16–19*, 264–273. LNCS 1452. Berlin: Springer.
Gibbon, Dafydd
2012 Resources for technical communication systems. In: Alexander Mehler, Laurent Romary and Dafydd Gibbon (eds.), *Handbook of Technical Communication*. Berlin/Boston: De Gruyter.
Goodwin, Charles and Marjorie Harness Goodwin
1997 La Coopération au travail dans un aéroport [Cooperation at the workplace in an airport]. *Réseaux* 85, 129–62. (Special issue "La Coopération dans les Situations de Travail" edited by Dominique Cardon).
Grice, H. Paul,
1953 Meaning. *The Philosophical Review* 66: 377–388.
Jan, Dusan and David Traum,
2007 Dynamic Movement and Positioning of Embodied Agents in Multiparty Conversations. In: *Proceedings of ACL 2007 Workshop on Embodied Language Processing*, 59–66.
Johnston, Michael, Srinivas Bangalore and Gunarajnan Vasireddy
2001 Match: Multimodal Access to City Help. In: *Proceedings of the Automatic Speech Recognition and Understanding Workshop, Madonna di Campliglio*, 2001.
Komiak, Sherrie Y. X., Weiquan Wang and Izak Benbasat
2004 Trust building in virtual salespersons versus in human salespersons: similarities and differences. *e-Service Journal* 3(3): 49–64.
Kopp, Stefan, Paul A. Tepper, Kim Ferriman, Kristina Striegnitz and Justine Cassell
2007 Trading spaces: How humans and humanoids use speech an gesture to give direction. In: Toyoaki Nishida (ed.), *Conversational Informatics: An Engineering Approach*. Chichester: John Wiley & Sons.
Kopp, Stefan, Jens Allwood, Karl Grammer, Elisabeth Ahlsén, and Thorsten Stocksmeier
2008 Modeling Embodied Feedback with Virtual Humans. In: Ipke Wachsmuth and Günther Knoblich (eds.), *Modeling Communication with Robots and Virtual Humans*, 18–37. LNAI 4930. Berlin: Springer.
Kopp, Stefan, Kirsten Bergmann and Ipke Wachsmuth
2008 Multimodal communication from multimodal thinking – Towards an integrated model of speech and gesture production. *Semantic Computing* 2(1): 115–136.

Kopp, Stefan and Ipke Wachsmuth
2012 Artificial interactivity. In: Alexander Mehler, Laurent Romary and Dafydd Gibbon (eds.), *Handbook of Technical Communication*. Berlin/Boston: De Gruyter.
Kubina, Petra and Andy Lücking
2012 Barrier-free communication. In: Alexander Mehler, Laurent Romary and Dafydd Gibbon (eds.), *Handbook of Technical Communication*. Berlin/Boston: De Gruyter.
Le Bel, Ronald M., Jaime A. Pineda and Anu Sharma
2009 Motor-auditory-visual integration: The role of the human mirror neuron system in communication and communication disorders. *Journal of Communication Disorders* 42(4): 299–304.
Lemon, Oliver, Anne Bracy, Alexander Gruenstein and Stanley Peters
2003 An information state approach in a multi-modal dialogue system for human-robot conversation. In: Hannes Rieser, Peter Kühnlein and Henk Zeevat (eds.), *Perspectives on Dialogue in the new Millennium*, *Pragmatics and Beyond*, 229–242. Amsterdam: John Benjamins.
Lester, James C., Sharolyn A. Converse, Susan E. Kahler, S. Todd Barlow, Brian A. Stone and Ravinder S. Bhogal
1997 The persona effect: Affective impact of animated pedagogical agents. In: *Proceedings of CHI '97*, 359–366.
Lester, James C., Jennifer E. Voerman, Stuart G. Towns and Charles, B. Callaway
1999 Deictic believability: Coordinating gesture, locomotion, and speech in like-like pedagogical agents. *Applied Artificial Intelligence* 13: 383–414.
Levinson, Stephen, C.
1983 *Pragmatics.* Cambridge: Cambridge University Press.
Liu, Jing and Manolia Kavakli
2010 A survey of speech-hand gesture recognition for the development of multimodal interfaces in computer games. In: *Proceedings of the 2010 IEEE International Conference on Multimedia and Expo, ICME 2010, 19–23 July, Singapore*, 1564–1569.
Louwerse, Maxim M., Arthur, C. Graesser, Danielle McNamara and Shulan Lu
2008 Embodied conversational agents as conversational partners. *Applied Cognitive Psychology* 23(9): 1244–1255.
Lücking, Andy and Thies Pfeiffer
2012 Framing multi-modal technical communication. In: Alexander Mehler, Laurent Romary and Dafydd Gibbon (eds.), *Handbook of Technical Communication*. Berlin/Boston: De Gruyter.
Magyari, Lilla and Jan Peter De Ruiter
2008 Timing in conversation: The anticipation of turn endings. In: Jonathan Ginzburg, Pat Healey, and Yo Sato (eds.), *Proceedings of the 12th Workshop on the Semantics and Pragmatics Dialogue*, 139–146. London: King's college.
Marsella, Stacey C. and W. Lewis Johnson
1997 An instructor's assistant for team-teaching in dynamic multi-agent virtual worlds. In: *Proceedings of the Fourth International Conference on Intelligent Tutoring Systems*, 464–473. Berlin: Springer.

Martin, Jean-Claude and Tanja Schultz
2012 Multimodal and speech technology. In: Alexander Mehler, Laurent Romary and Dafydd Gibbon (eds.), *Handbook of Technical Communication*. Berlin/Boston: De Gruyter.
Oberzaucher, Elizabeth and Karl Grammer
2008 Everything is movement: on the nature of embodied communication. In: Ipke Wachsmuth, Manuela Lenzen and Günter Knoblich (eds.), *Embodied communication*, 151–177. Oxford: Oxford University Press.
Paiva, Ana and Isabel Machado
1998 Vincent, an autonomous pedagogical agent for on-the-job training. In: Valerie Shute (ed.), *Intelligent Tutoring Systems*, 584–593. Berlin: Springer.
Partan, Sarah R. and Peter Marler
2005 Issues in the classification of multimodal communication signals. *The American Naturalist* 166(2): 231–245.
Peirce, Charles Sanders
1931 Principles of Philosophy. Vol. 1, Book 2. § 369, p. 195. In: *Collected Papers of Charles Sanders Peirce, 1931–1958*, 8 vols. Edited by Charles Hartshorne, Paul Weiss and Arthur Burks. Cambridge, MA: Harvard University Press.
Person, Natalie K., Arthur C. Graesser, Roger J. Kreuz and Victoria Pomeroy
2001 Simulating human tutor dialog moves in AutoTutor. *International Journal of Artificial Intelligence in Education* 12: 23–29.
Pomerantz, Anita M.
1984 Agreeing and disagreeing with assessment: Some features of preferred/dispreferred turn shapes. In: J. Max Atkinson and John Heritage (eds.), *Structure of Social Action: Studies in Conversation Analysis*, 57–101. Cambridge: Cambridge University Press.
Rao, Anand and Michael Georgeff
1991 Modeling Rational Agents within a BDI-Architecture. In: *Proceedings of Knowledge Representation and Reasoning (KR&R 1991)*, 473–484.
Rickel, Jeff., and W. Lewis Johnson
1999 Animated agents for procedural training in virtual reality: Perception, cognition, and motor control. *Applied Artificial Intelligence* 13: 343–382.
Rickel, Jeff and W. Lewis Johnson
2000 Task-oriented collaboration with embodied agents in virtual worlds. In: Justine Cassell, Joseph Sullivan, Scott Prevost and Elizabeth Churchill (eds.), *Embodied Conversational Agents.* Cambridge, Massachusetts: The MIT Press.
Ruttkay, Zsófia and Catherine Pelachaud (eds.)
2004 *From Brows till Trust: Evaluating Embodied Conversational Agents*. Dordrecht: Kluwer.
Sacks, Harvey, Emanuel A. Schegloff and Gail Jefferson
1974 A simplest systematics for the organization of turn-taking for conversation. *Language* 50: 696–735.
Sargin, Mehmet E., Yucel Yemez, Engin Erzin and Ahmet M.Tekalp
2008 Analysis of head gesture and prosody patterns for prosody-driven head-gesture animation. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30(8): 1330–1345.

Shank, Roger and Robert Abelson
1977 *Scripts, Plans, Goals and Understanding*. Hillsdale, HJ: Erlbaum.
Searle, John R.
1970 *Speech Acts: An Essay in the Philosophy of Language*. Cambridge, Massachusetts: Cambridge University Press.
Self, John
1999 The defining characteristics of intelligent tutoring systems research: ITS care, precisely. *International Journal of Artificial Intelligence in Education* 10: 350–364.
Thórisson, Kris R.
1997 Gandalf: An Embodied Humanoid Capable of Real-Time. Multimodal Dialogue with People. In: *Proceedings of the first ACM International Conference on Autonomous Agents, Marina del Rey, California, February 5–8*, 536–537.
Toptsis, Ioannis, Shuyin Li, Britta Wrede and Gernot A. Fink
2004 A Multi-modal Dialogue System for a Mobile Robot. In: *Proceedings of the International Conference on Spoken Language Processing*, 273–276.
Vicari, Rosa Maria, Patricia Augustin Jaques and Regina Verdin (eds.)
2009 *Agent-Based Tutoring systems by Cognitive and Affective Modeling*. Hershey, PA, USA: IGI Global.
Wachsmuth, Ipke
2005 "Ich, Max" – Kommunikation mit künstlicher Intelligenz. In: Ch. S. Herrmann, M. Pauen, J. W. Rieger and S. Schicktanz (eds.), *Bewusstsein: Philosophie, Neurowissenschaften, Ethik,* 329–354. München: Wilhelm Fink Verlag.
Wahlster, Wolfgang
2003 SmartKom: Symmetric Multimodality in an Adaptive and Reusable Dialogue Shell. In: *Proceedings of the Human Computer Interaction Status Conference 2003*, 47–62. Berlin: DLR.
Waltinger, Ulli and Alexa Breuing
2012 Internet-based communication. In: Alexander Mehler, Laurent Romary and Dafydd Gibbon (eds.), *Handbook of Technical Communication*. Berlin/Boston: De Gruyter.

14. Decentralized Online Social Networks

Nils Diewald

1. Introduction

Social Network Sites (SNSs) are becoming an increasingly important communication platform on the World Wide Web. According to Nielsen Media, in June 2010 Americans spent nearly a quarter of their online time on SNSs, which therefore supersede email as the most important communication medium on the Internet (Martin 2010). As of September 2011 *Facebook*[1], at this time the biggest SNS worldwide with more than 750 million users (Facebook 2011), was the second most visited site on the web (Alexa 2011).

In this chapter we will discuss the benefits of SNSs as well as potential problems for their users, especially associated with the centralized nature of current SNSs.

SNSs allow for the establishment of direct relations to other members of the service, and hence creating an online social network (OSN)[2]. These networks have known benefits regarding efficiency in the dissemination of information, often associated with the so-called "small world" phenomenon proposed by Milgram (1967): Travers and Milgram (1969) conducted a study, asking randomly selected individuals to send letters to unknown target persons, mediated only by chains of persons they personally know. They found these chains to be unexpectedly short on average (a little greater than five; Travers and Milgram 1969, pp. 432, 437), in case they were completed[3]. The assumption was that there is only a small distance between any two individuals in a social network in terms of direct acquaintanceship, which leads to potentially good target-oriented dissemination of information (Dodds, Muhamad, and Watts 2003). A popular hypothesis in social science claims an approximated average value of "six degrees of separation" between any two people on earth[4]. And recent analyses of global technical communication networks – like Email (Dodds, Muhamad, and Watts 2003) or contact lists in Instant Messaging systems (Leskovec and Horvitz 2008) – surprisingly supported this number roughly.

The "six degrees" also gave one of the first Social Network Sites on the World Wide Web its name: *SixDegrees.com* was launched in 1997. It allowed its users to build personal profiles, to send direct messages to other users, to create a list of contacts and – starting in 1998 – to surf the so established OSN (boyd and Ellison 2007). After SixDegrees.com was shut down in 2001, services like *Friendster*[5] (2002), *MySpace*[6] (2003), *Orkut*[7], Facebook (both 2004), *Twitter*[8]

(2006), or *Google+*[9] (2011) adapted its basic principles and introduced new features and forms of relationships and communication.

Boyd and Ellison (2007) define SNSs as

> … web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system. (boyd and Ellison 2007, p. 211)

While this definition may be arguable (Beer 2008), it can serve as a blueprint for the architecture of prototypical SNSs as examined in this chapter: A SNS provides tools for the construction of an OSN, with users as nodes and social relationships as edges, the presentation of the user by means of profile pages, and the ability to explore the network along these links. However, this definition may lack a significant aspect of SNSs, which makes them important communication platforms and should be introduced as a fourth item: The ability to interact with each other within the social network[10]. Popular channels for interaction in SNSs are, among others, direct and instant messaging, micro-blogging, sharing of multimedia content, and profile pages. Some of these channels are interwoven in the user profiles, and all are embedded in the context of the egocentric OSN. In addition to this modification, the limitation of "web-based services" seems to be outdated as non-web applications become more and more popular in recent years to participate in OSNs, especially on mobile devices (Ziv and Mulloth 2006; Lugano 2007).

In the first part of this chapter, we will illustrate the characteristics of SNSs, OSNs, and social networks in general from a sociological point of view and discuss their importance as growing communication platforms (Section 2).

Despite their short history, SNSs have become the most popular and dominant communication tool of many users online. But unlike preceding technical communication systems, for example email or telephone, their infrastructure is closed and centralized. Individuals who do not participate are not able to communicate with members. And if they participate, they are not able to communicate with members of other sites. Furthermore, they have not much control on their personal information and how it is disseminated. All information presented and exchanged is under the control of the website's service provider, which may result in serious privacy concerns. This turns these sites into "information silos" (Au Yeung et al. 2009) and their underlying OSNs into "walled gardens" (Fitzpatrick and Recordon 2007).

In the second part of this chapter, we will examine these problems regarding the closed and centralized architecture of current SNSs in both technical and privacy related terms, and will discuss the different network topologies (Section 3).

Tim Berners-Lee, one of the inventors of the World Wide Web, made use of the 20th anniversary of the web in 2010 to warn about this, in his eyes, threat to the open web – and talked about “continued grassroots innovations” to anticipate this development, mentioning projects like *GNU social*[11], *StatusNet*[12] and *Diaspora*[13] (Berners-Lee 2010). The latter was announced only a couple of months before, asking for funding to create a decentralized SNS on the open crowdfunding platform *Kickstarter* (Kickstarter 2010). While it was not the first project of its kind, it raised a lot of media attention (including an article in “The New York Times”) and at the end of its fundraising round it had accumulated twenty times its goal of $10,000 dollars capital, proclaiming the funders’ need for the decentralization of OSN based communication.

In the third part of this chapter, we will discuss these ongoing developments. Open and royalty-free formats and protocols, that are necessary for interoperable and vendor independent systems, are presented (Section 4) as well as projects focusing on decentralizing OSNs (Section 5).

We will conclude with a summary on the topic (Section 6).

2. Characteristics and sociometrics of Social Network Sites

“The web is more a social creation than a technical one.”
Berners-Lee (2000, p. 123)

The socialization of the web has many forms: In addition to social networking there are applications of social sharing, social news, social bookmarking, social gaming, or social commerce (Schneider 2008). These applications of “social software” have one special thing in common: they support group interactions (cf. Allen 2004). Users of social software interact with each other to the benefit of the service. Take, for example, the collaborations and discussions on the website *Wikipedia* and other wikis (Mehler and Sutter 2008). SNSs, as a subcategory of social software, focus on these interactions (rather than on the object the interactions are about like in, e.g., online bulletin boards; James, Wotring, and Forrest 1995) and allow for an explicit construction of the social relations.

The OSN, based on these relations, can be seen as part of the topmost conceptual layer of a technical communication network (see Figure 1)[14]. The basic layer, as described by Berners-Lee (2007), is the Internet, which he calls the “International Information Infrastructure” (III) (see Figure 1, bottom). Starting in the 1980s[15], it changed the perspective on digital information exchange, realizing “It isn’t the cables, it is the computers which are interesting” (Berners-Lee 2007). The underlying communication network was formed by computers as nodes and cables as the edges. The “World Wide Web” (WWW), as the second technical layer (see Figure 1, center), then sharpens the view to “It isn’t the computers, but the documents which are interesting” (Berners-Lee 2007). In the

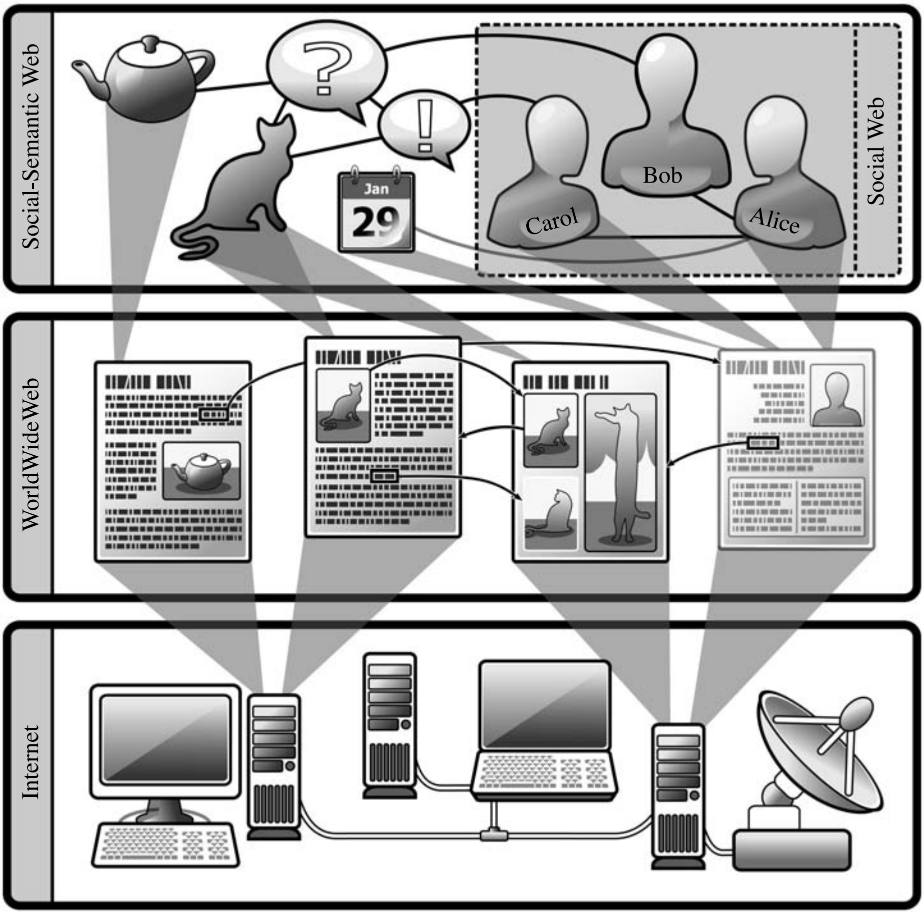


Figure 1. The historically evolved three layers of linked data: The Internet (*III*), representing a network of computers, the web (*WWW*), representing hyper-linked documents, and the social-semantic web (*GGG*), representing interlinked data of objects. The social web is an application of the semantic web.

WWW, documents form the nodes of the network with hyperlinks as the edges. Now, the semantic web (see Figure 1, top) is meant to model the realization "It's not the documents, it is the things they are about which are important" (Berners-Lee 2007), for which Berners-Lee coins the term *Giant Global Graph* (GGG). In this technical and conceptual layer, all "things" form nodes in the graph and are related to each other: Texts and photographs as well as humans, products, dates, and thoughts.

There are two perspectives on what the nodes and edges of the social graph as embedded in the GGG can be: A narrow and explicit as well as a wide and im-

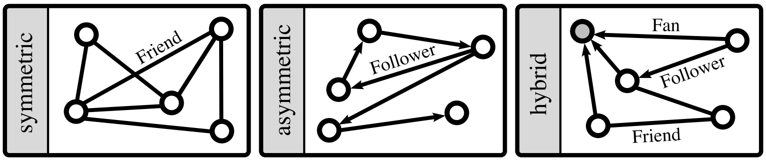


Figure 2. Symmetric and asymmetric relations in SNSs.

plicit one. From the explicit perspective, the social graph is the subgraph of the GGG where all nodes are humans or institutions and their relations are manually established by the members of the social network (e.g., by adding a "friend" to the list of friends on a SNS). In Figure 1 (top), Alice and Bob are connected in this explicit manner, while Carol and Bob are not.

From the implicit perspective, two individuals can also be mediated by different objects in the GGG, forming an "object-centered social network" (Engeström 2005; Breslin and Decker 2007; Breslin, Passant, and Decker 2009). That means, although Bob and Carol are not explicitly connected, they have a mediated relation by taking part in the same online discussion (see Figure 1, top).

When talking about OSNs in the domain of SNSs in this chapter, we will refer to the narrow, explicit perspective, where all relations are direct and manually established by the individuals in the network.

2.1. Profiles, relations, and interaction tools

The information provided by SNSs consist of three basic components: *profiles*, *relations*, and *interactions*. Users of SNSs provide a huge amount of such information on themselves, their social network and their activities[16].

Profile pages represent meta information on the nodes of the social graph, connected via relations that form the paths on which interactions like direct messaging can happen. These relations are manually created by the members of the SNS. Their relation type can form different shapes of networks (see Figure 2): In most cases they are symmetric, as for example in Facebook's relationship model (see Figure 2, left). If the user Alice adds the user Bob to her list of "friends", Bob is asked for acceptance of this friendship invitation. In the case of rejection, there is no relationship between both of them established. In case of acceptance, the contacts are displayed in each others' "friends" lists. Twitter, a microblogging[17] based SNS, on the other hand, has an asymmetric relationship model (see Figure 2, center). Whenever the user Alice wants to read status updates (or "tweets") of the user Bob, she simply "follows" him – his short notes then will be part of Alice's news stream, unless Bob has protected his status updates from being displayed to the public. In that case, Bob has to accept or reject the request of Alice for "following". After that, Alice has Bob in her "following" list, while Bob has Alice as a new member in his "follower" list. Some SNSs allow for both relationship models: *ResearchGate*[18], for example, a SNS for academics, allows users to symmetrically relate with collaborators as well as to follow scientists whose work the user is interested in (see Figure 2, right). Facebook allows partial asymmetric relations with the ability to be a "fan" of special nodes in the social network, such as celebrities, shows, or other objects or events.

As for the wide diversity of SNSs, in the following sections we will focus on symmetric OSNs and refer to Facebook as being a prototypical SNS of this type, due to its concept of reflecting real life social networks: "A social graph is a model for Facebook, we're not trying to make new connections, but mirror the real world", says Mark Zuckerberg, founder and CEO of Facebook (Riley 2007). This is a new model of contemporary SNSs (Lampe, Ellison, and Steinfield 2006). Early services like Friendster had their origin in online dating platforms and thus were aiming at meeting new acquaintances, relying on the idea that an existing social network based on real life acquaintanceship is a good basis for making new connections (boyd 2004).[19] This leads boyd and Ellison (2007) to the preferred term of "Social Network Site" instead of the commonly used "Networking Site", which in their minds emphasizes on the creation of new connections in the social graph.

While meeting new friends still can be part of a user's activity in a SNS, the actual variety of activities on these platforms is huge. As discussing this variety in detail is beyond the scope of this chapter (refer to Thimm 2008; and Waltinger and Breuing 2012, in this issue, for different forms of online communication), we will focus on the impact of the embedding social layer on the different forms of interaction in SNSs.

In Facebook the tools for interaction can be subdivided into three groups: *Communication tools* (including "Wall" postings, status updates, chat, direct messages, comments, and events), *presentation tools* (including profile data, groups, and photo albums), and *entertainment tools* (including games, tests, pokes, and gifts) (Kneidinger 2010).

The *communication tools* vary in the size of the addressed audience as well as regarding their synchronism, from inter-personal (one-to-one, e.g., direct messages) to public communication (one-to-many, e.g., "wall" postings with no specific addressee), and from nearly synchronous (e.g., chats) to asynchronous communication (e.g., status updates). Most of the tools, mainly because of the extent of short messages, bare specific communication characteristics, such as informal use of language, non-orthographical use of punctuation and capitalization, and wide use of abbreviations and emoticons, as known in online communication (Werry 1996).

The *presentation tools* form the basic elements of the profile pages (see Figure 3). Most noticeable elements include the profile data, usually enriched by a

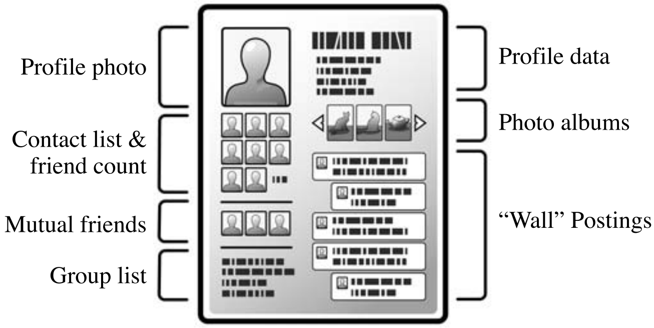


Figure 3. Prototypical elements of SNS profile pages.

photograph, contact and personal information like age, birthday, location, interests, and a short "about me" prose text. Additionally, photo albums and groups – if existing – are part of the presentation of a user's online identity.

The *entertainment tools* make use of the different communication channels and data provided by the presentation tools of the user and other members of the OSN.

In more and more SNSs the profile sphere aggregates further elements, as part of the communication and entertainment tools, as well as information about the user's position in the social network. The so-called "activity stream" merges various status updates of a user, including short messages addressed to friends, messages from friends to the user's "wall", notifications on activities of the user (e.g., changes of profile data), or information from external services (e.g., games and polls the user participates in).

Profile information is public by a certain degree: The position of the user in her egocentric OSN can provide access authorization to this data. In Facebook, the user is able to adjust the privacy level, deciding that only direct friends can retrieve certain information. An initial function of Friendster was the restriction to be disabled to view profiles which are more than three mediating friends away from the observing user (boyd and Ellison 2007). This authorizing principle also applies to the various communication tools on these platforms, as a user is only able to write direct messages or comment on status updates she is allowed to view.

The connections themselves are part of the profile, too, and, as already noted, their visibility and traversability are crucial aspects of SNSs. In Facebook, a random part of the user's friends is displayed by thumbnail sized images of their photographs, associated with a number indicating the amount of friends the user has (see Figure 3). In addition to that, some SNSs visualize the path between the observer and the owner of a profile by means of mediated connections, or a list of mutual friends.

2.2. Sociometric popularity

A main difference between SNSs and other forms of social software is the embedding of the user's profile in social and communicational contexts. The visibility of the egocentric network is supposed to have an important impact on the user's online reputation (Donath and boyd 2004): While the owner of a profile is in control of her profile data, like name, age, photograph, or her membership in groups, her influence regarding the surrounding social environment and the public communication (e.g., "wall" postings of friends) is rather weak. These contributions from persons other than the profile owner distinguish SNS profiles from other applications of Internet communication, like web pages, emails, or chats, where the owner is in full control of her creation (Walther et al. 2008).

To separate aspects of this visibility in SNSs, we will refer to different perspectives a user can have on displayed elements: An *owner-centric view*, applied when watching the user's own profile, and an *observer-centric view*, applied when watching the profile of another user. The display of connections on the profile pages has different benefits depending on these perspectives. The display of one's own contact list can help to stay in touch with friends and acquaintances and the display of a foreign person's contact list gains her credibility (Donath and boyd 2004).

Based on the aforementioned assumption that most connections reflect relationships in the real world, facts in a user's profile data are validated by the companionship of her friends, and are barely to manipulate in this way. In consequence, this makes SNSs easy for the exploration of a person's real life sociometric data, leading to the assumption *"I am whom and how many I'm connected to"*. If a user is married, for example, it is nearly impossible to pretend being single. The same is true for gender, age and other claims, often doubtful in online communication (cf., e.g., Stoll 1995, pp. 56–59). Hence, friendship links can serve as "identity markers" (Donath and boyd 2004). Due to this, profile pages in SNSs like Facebook and the German *StudiVZ*[20] have shown to be close depictions of the persons in real life (Back et al. 2010), rather than presentations of idealized views on their selves (as suggested by, e.g., Manago et al. (2008) in the context of MySpace).

Thus, the online behaviour of the user is connected to her real life social reputation: By embedding herself in real life context, the user signals the "willingness to risk one's reputation" (Donath and boyd 2004), online as well as offline. In anonymous contexts, like in online guestbooks, the user does not risk any reputation by her behaviour. This can be an important advantage when expressing opinions or asking awkward questions (McKenna and Bargh 2000). In pseudonymous contexts, for example in online bulletin boards, a user is expressing the will to only risk her online reputation that is connected with the corresponding pseudonym. The negative aspect of anonymous or pseudonymous contexts is that bad behaviour has only small negative effects on the user, because it does not effect real life social reputation (see Levmore and Nussbaum 2011, for a collection of articles on this topic). That's why communication in SNSs is expected to raise more polite behaviour online (cf. Schonfeld 2011).

Beside reliability, contact lists also affect the perception of a user's social and physical attractiveness. The more attractive the friends of a user are, the better is her perception (Walther et al. 2008). This is also true for all influences on the user's profile that are not in her control, for example, "wall" postings or other forms of external content, and the number of friends (see Section 2.4).

2.3. Quality of social relations

The relational term "friend", as used by Facebook, indicates a strong and somehow intimate relationship. But more often, a lot of these relationships, especially when dealing with implausible high numbers, are rather weak (boyd 2004). Beside the aforementioned reputational aspects of contact lists, political and technical reasons may play a role for gathering more and more contacts: Politically, as the rejection of a "friend's" request may seem impolite or may result in a social disprofit, and technically, because the "friend" relationship widens the authority of a user, for example, it allows her to browse more protected areas in the profiles of other members (see Section 2.1) or to advance her level in a social network game (Wohn et al. 2011). Additionally, users rarely remove established "friends" from their contact lists, although their real life relationship may have passed.[21]

That means that although it is rather likely that the contacts of a person in modern SNSs like Facebook are based on real life relationships, the strength of these relationships is arguable. Thus, when talking about the impact of social networks on technical communication in SNSs, the characteristics of the interpersonal relations may be of high importance.

Granovetter (1973) argues that the strength of a social connection is a

> … (probably linear) combination of the amount of time, the emotional intensity, the intimacy (mutual confiding), and the reciprocal services which characterize the tie. (Granovetter 1973, p. 1361)

A strong tie can exist between close friends or family members, with lots of time spent on the maintenance of the relationship, while weak ties exist between persons who are related in a specific context, for example co-workers or persons who share a common interest and spend time together only occasionally.

In the model of Granovetter (1973), strong ties between an individual and two others indicate a high probability of at least a weak tie between the other two. In Figure 4, Alice has a strong tie with Bob as well as with Carol. Now it may be the case that Bob and Carol are related by a strong tie, too, but at least there is a high probability that they know each other and have a weak tie. In a so-

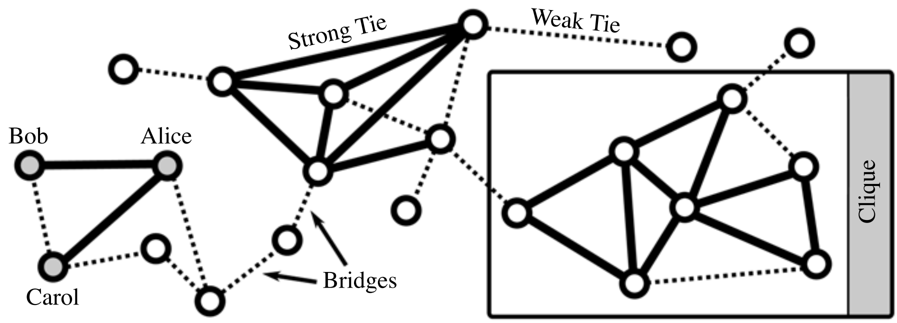


Figure 4. Illustration of a small social network with three cliques connected via bridges. There are strong ties between the individuals Alice and Bob, and Alice and Carol. Based on the definition by Granovetter (1973), there is at least a weak tie between Bob and Carol.

cial network based on this assumption, strong ties form dense, fully connected subgraphs within the social network, so-called "cliques"[22].

Weak ties, on the other hand, are able to connect these clusters. Ties in a network that provide the only path between two individuals by a certain degree (so-called "Local bridges"; by definition always weak), are significant for social networks, as they connect cliques and lead to more and shorter paths in the network (see Figure 4). That means that the better a network is connected by "bridges", the more effective it is in terms of communication as they provide an important impact to the small world phenomenon. That makes weak ties the more critical connections in social networks: "the removal of the average weak tie would do more 'damage' to transmission probabilities than would that of the average strong one" (Granovetter 1973, p. 1366). The weakness of the majority of ties in the egocentric OSN of SNSs can in this way be profitable in terms of dissemination of information. The reason is that cliques gather people that are structurally more equivalent (Burt 1992, pp. 18–19), that is, they have lots of connections in common. In consequence, information will tend to circulate inside the clique. Individuals, that connect otherwise separated cliques and thus fill "structural holes" (Burt 1992) in the network, have a significant control on the flow of information.

For example, when a rumor is told exclusively to close friends and these friends will also retell the rumor exclusively via strong ties, the circulation in the cliques will lead to repetitive transmission of the same information addressed to the same person. By using weak ties, a larger number of people can be reached and a longer path in the social network can be traversed (Granovetter 1973, p. 1366). This is important, when actively intending a widespread dissemination of information as well as for getting novel information (see Section 2.5). Granovetter (1995) showed, for example, that it is a lot more likely to get a new job through weak ties, by means of co-workers or less known acquaintances, than by friends or family members.

Weak ties, as they are able to bridge between cliques, are important for informational spread with high social distance (i.e., the path length in social networks), reaching more individuals. On the other hand, strong ties "have greater motivation to be of assistance" (Granovetter 1983, p. 209) and are more likely to be trusted.

Before the advent of SNSs, online communities were often formed by a common interest (James, Wotring, and Forrest 1995), and most members of an online community, as in bulletin boards, newsgroups or chat rooms, did not know each other in person (Wellmann and Gulia 1999, p. 335). These communities were based most completely on weak ties. "Friends" on SNSs nowadays, as they are based on real life relationships, are both: weak and strong. That makes them more effective on the dissemination of information.[23]

2.4. Quantity of social relations

Additionally to the quality of social relations, the quantity has also to be taken into account. Especially as weak ties, by means of local bridges, are important contributors to information flow in social networks by virtue of their numbers – not based on their individual efficiency (Friedkin 1982).

The average member of Facebook has 130 "friends" (Facebook 2011). While this may sound a lot, it is far below the average of 500 to 2.500 acquaintances proposed in Milgram's seminal paper on the small world phenomenon (where acquaintanceship was defined as "known on a first-name basis", Milgram 1967, p. 64).

Anthropologist Robin Dunbar hypothesized that there is a biological limit of social contacts that can be reasonably handled by humans, predicted by around 150 persons (Dunbar 1993, p. 682). The number, Dunbar says, is limited due to the size of the human brain's neocortex, the part of the brain in mammals which deals with complex and logical thoughts. This is meant to be roughly the number of contacts a human "can keep track of within its social group" (Dunbar 1995, p. 287) and is able to have a "genuinely social relationship" (Dunbar 1996, p. 77) with by knowing the other person and knowing how they relate to each other.[24] Dunbar (1993) showed that 150 is a functional size of working communities throughout history. Therefore he claims having more than 200 friends in a SNS can be seen as implausible regarding the term "friend" (Dunbar 2010, p. 22). As the number of friends a user has is displayed in most SNSs, Tong et al. (2008) studied the influence of this sociometric value regarding the influence on her perception. The interesting result was, that the number is not proportional to the user's social attractiveness. Although too few friends indicate less attractiveness (closely related to physical attractiveness), an implausible amount leads to a loss of social attractiveness. The research measured a decrease of social attractiveness by around 300 “friends” (Tong et al. 2008).

Taking the emotional dimension of strong ties into account (following the definition by Granovetter), psychologists argue that there is also a limit on emotional capacity to humans, limiting the number of persons we are strongly emotionally related to. Therefore a person’s “sympathy group” (e.g., a group of “persons whose death would cause you anguish” as defined in a study by Buys and Larson 1979) is rather small, consisting of typically less than 10 to 20 people (Parks 2007; Wellmann and Potter 1999). A similar small subgroup, as a number of active ties with regular interactions and thus greater invested time for maintenance, can be found in Facebook as well (Marlow 2009).

Thus, while there seem to be natural limits of ties a human can have, this does not limit the size of the egocentric social network of a person, as it only holds true for a given point in time (boyd 2005)[25]. Donath and boyd hypothesize, that, although the number of strong ties will not increase by technology provided by SNSs, the number of weak ties that can be maintained “may be able to increase substantially” (Donath and boyd 2004, p. 80). Dunbar concurred with this hypothesis, saying that while biological limits will not be crossed, the communication in SNSs may make the maintenance of relationships more time efficient and less dependent on geographical distance (Dunbar 2010; Krotoski 2010).

Therefore, it is no surprise that keeping in touch with acquaintances and geographically distant friends is the most important benefit of SNSs for the majority of users (Joinson 2008; Kneidinger 2010).

2.5. Commercial use of Social Network Sites

Albeit this user-oriented benefit of SNSs, the most important feature of social networks in general regarding communication is the effective dissemination of information, that is, the acting of a social network as a communication network. It is important in terms of “learning” from trusted friends and acquaintances with implications on various fields, including how individuals “find employment, but also about what movie they see, which products they purchase, which technologies they adopt, whether they participate in government programs, whether they protest, and so forth” (Jackson 2008, pp. 71–72). But – to no surprise – most recent research on diffusion of information in social networks focus on marketing.

Dissemination of information in SNSs happens in a number of ways. Atomic forms include, for example, the “like” button in Facebook, that, by expressing a positive rating on a status message of a friend, a website, a photo or something similar, allows for the quick diffusion of this information with all direct friends in the OSN. The “liked” item then shows up in all friends’ news streams. In Twitter users can “retweet” status messages to make them visible to their followers[26].

In times of online retailers, traditional mechanisms of economics change: Due to the enormous size of potential customers, infinite shelf space and simplified logistics, online retailers can offer a wider variety of products than brick-and-mortar stores. While traditional retailers are most profitable by offering "best-sellers" only[27] online retailers can benefit from selling niche products, making the "long tail" of their product-lineup profitable (Anderson 2006)[28]. To advert niche products effectively, "using traditional advertising approaches is impractical" (Leskovec, Adamic, and Huberman 2006).

Because of its user-centric, information rich approach and the underlying network structure, SNSs are popular for targeted advertising. Like Google's *AdSense*[29], these sites can provide context sensitive advertisements by analyzing content on profiles, messages, groups etc. So, when reading a message in a group with a computer-related topic, the user may see computer-related ads. But SNSs can additionally rely on *who the consumer is*: The advertisements can be filtered regarding profile information like age, sex, location, and interests. Thus, if the system has further information on the location of the user and her employment status, it can show an advert for a computer-related job when reading in the computer-related group. This user-centric approach to online advertisements can be compared to automated recommendations as on *Amazon*.

By additionally applying social network analysis (SNA), SNSs do not have to rely on the explicit information in the user's profile – it can also rely on the information from the user's contacts. Because users in cliques do not only tend to be structurally equivalent (see Section 2.3), but also tend to be similar regarding several properties like age, education, religion, ethical values, behaviour patterns and so on, the system can guess, who the user is and what she is interested in. This phenomenon is called "homophily" (Lazarsfeld and Merton 1982) and describes "the principle that a contact between similar people occurs at a higher rate than among dissimilar people" (McPherson, Smith-Lovin, and Cook 2001, p. 416). In that way, homophily can be helpful for successfully guessing the background of an individual. On the other hand it can act as a "barrier to diffusion" (Rogers 2003, p. 306), as a high degree of homophily in a subnet of a social network will tend to form circles in information propagation. Bridges in social networks are thus more of a "heterophilous" nature, that means, relations of less similar individuals.

But SNA is not limited to target marketing based on the *intrinsic value* of the user: it also can take her network value into account (Domingos and Richardson 2001). Further filters can be set to target structurally important nodes in the network (see Section 2.3) or individuals with high reputation (see Section 2.2) to identify opinion leaders within the network. Taking in advantage the effective dissemination of information in social networks, these individuals are preferentially triggered to recommend products to their acquaintances, hopefully starting a cascade of positive "word-of-mouth" propagation (Kempe, Kleinberg, and

Tardos 2003). As recommendations by friends and acquaintances are still the most trusted ones (The Nielsen Company 2009), these viral marketing strategies are popular among SNSs[30].

Although these principles are best studied in the field of marketing, the effectiveness of dissemination of information is – as Jackson (2008) noticed – also significant in political discourse. In early 2011, for example, protesters in several countries of Northern Africa were using SNSs to organize their actions. It helped them to quickly diffuse information concerning governmental attacks on demonstrators in form of *Youtube*[31] videos or blog posts.[32]

3. Centralization versus decentralization

> "The trick here, though, is to make sure
> that each limited mechanical part of the
> Web, each application, is within itself
> composed of simple parts that will
> never get too powerful."
> Berners-Lee (2000, p. 183)

The benefits of SNSs by means of effectivity of information diffusion and maintenance of social relations however may have some drawbacks.

Concerning the maintenance of weak ties, the benefit depends on the participation of all individuals in the same SNS. That is because for now, each of these SNSs are relying on their own separated OSNs. Speaking in terms of social network theory, these sites form cliques with no local bridges – they are "independent, isolated and incompatible" (Mitchell-Wong et al. 2007). Members of a SNS can interact with each other, but not with members of other SNSs. The centralized scenario is often parallelized to the Internet's past in the "walled gardens" of *AOL* and *CompuServe* (Recordon 2007; Li 2008). In these early years of the private usage of the Internet, some companies provided email services and discussion forums for their customers, without interoperability with the systems of others – until the emergence of the open, decentralized World Wide Web.

This architecture of the web guarantees the interoperability between different systems by means of open standards (see Section 4) and thus the independence for a customer to choose her provider.

Regarding SNSs, users are free to choose their provider as in the early days of the Internet. But furthermore, they can participate in more than one SNS, to be able to connect to as many friends as possible. That means, their identities in the web's "social ecosystem" (Mitchell-Wong et al. 2007) are not unique.

Figure 5 juxtaposes this centralized scenario for SNSs and a decentralized form. In the centralized scenario (Figure 5, top), the user Alice participates in

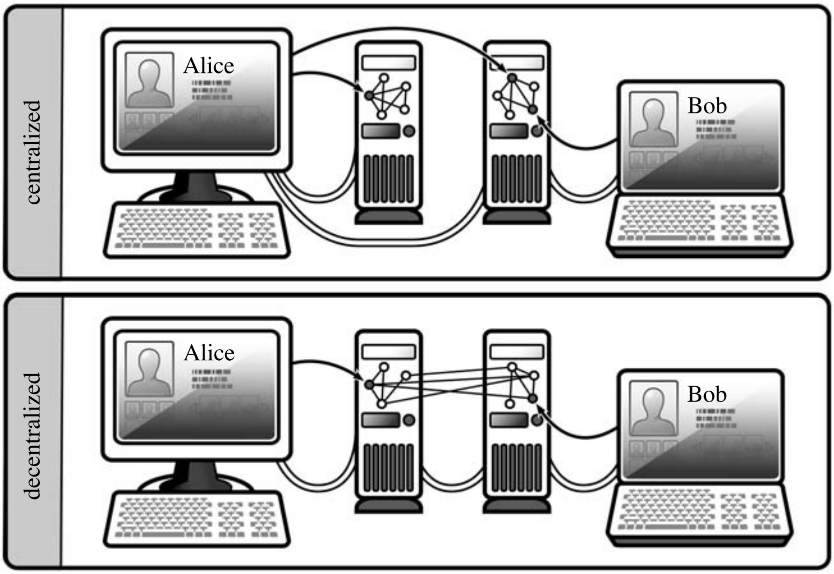


Figure 5. The architecture of centralized and decentralized OSNs. While in both scenarios two SNSs exist, the underlying OSNs in the centralized scenario are separated, while there is an interconnected OSN in the decentralized scenario.

two separated SNSs, having two separated identities with separated profiles and separated contact lists in separated OSNs. Bob participates in one of these SNSs and in this site's OSN he is a friend of Alice – while he is not a friend of Alice in the other one. That means, the OSNs of both SNSs are separated.

In the decentralized scenario (Figure 5, bottom), Alice and Bob participate in one and the same OSN, although they are members of separated SNSs. Their identities in the web's social ecosystem are unique. They each have one profile and one contact list in one common OSN. That means, their SNSs are no "walled gardens" anymore.

As SNSs become more and more important as communication platforms, vulnerabilities regarding *technical* and *privacy* issues gain significance. This Section will discuss these issues in the context of centralized SNSs and will juxtapose centralized and decentralized scenarios, as they have different properties in these aspects.

3.1. Technical issues

The technical architecture of communication networks is always a critical issue. The main objective is to avoid scenarios where failure of small parts of the architecture have a huge impact on the stability of the whole network. When

Paul Baran outlined possible topologies of computer networks in 1964, he stated that a "centralized network is obviously vulnerable as destruction of a single central node destroys communication" (Baran 1964, p. 1). Although Baran was talking primarily about enemy attacks on the communication infrastructure of a country, the nature of single points of failure (SPOFs) in network systems is risky in many ways.

In 2009, during the Iranian election, supporters of the political opposition communicated heavily via SNSs like Twitter or Facebook to inform about protests and organize their campaign. The communication broke down as the services were blocked by the Iranian regime. Because of their centralized nature, there was only the need to block access to two sites to disrupt the availability of the necessary information and the whole communication structure. Shortly before this blocking action, due to its responsibility regarding the protesters in Iran, Twitter refused to take an operation break for maintenance, that would otherwise have interrupted the communication worldwide (Stone and Cohen 2009).

In September 2010, Facebook went offline for two and a half hours because of a software bug (Johnson 2010), and in the hours right after the death of musician Michael Jackson in June 2009, Twitter almost collapsed[33] because of the high frequency of new tweets on this topic (Bates 2009).

But also enemy attacks can still influence the availability of services in a communication network: On August the 6th, 2009, both Facebook and Twitter were facing "Denial of Service" attacks, forcing Twitter to go even offline (Van Buskirk 2009).

Enemy attacks as well as software bugs, performance problems, maintenance downtime, and governmental blocks make SPOFs critical in terms of technical communication networks. Sometimes services simply go out of business, as the example of the previously mentioned SixDegrees.com has shown, or the defunction of Twitter competitor *Pownce*[34] in 2008. As SNSs become more and more popular, people change their communication behaviour and rely on these centralized communication platforms.

Figure 6 visualizes centralized, decentralized, and distributed scenarios of OSNs following the subdivision of communication networks by Baran (1964).

In the centralized scenario (Figure 6, top-left), Alice participates in the SNSs 1 and 2, while Bob only participates in 2. That means, Bob is not able to connect with users who only participate in SNS 1. In the decentralized and distributed scenarios on the other hand (Figure 6, top-right and bottom-left), every user can connect to every other user, regardless which SNS they participate. The decentralized scenario has Bob participating in SNS 1 and Alice participating in SNS 3. Because these SNSs are technically connected, Alice and Bob can be socially connected in the same OSN, although they participate in different SNSs.

While centralized scenarios are highly vulnerable to SPOFs (think of SNSs 1 and 2 as being Facebook and Twitter), a decentralized scenario is less vulner-

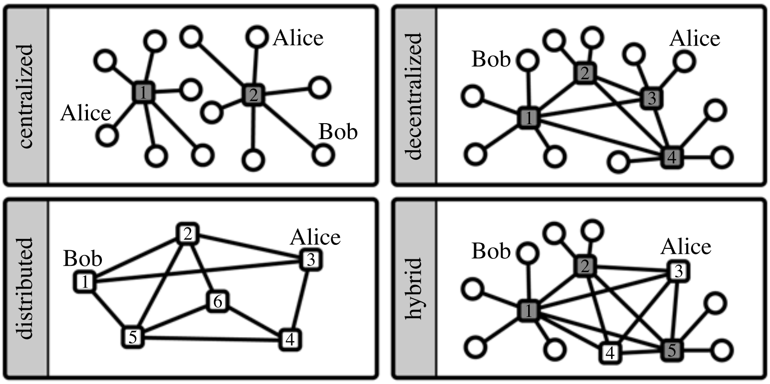


Figure 6. The topologies of centralized, decentralized, distributed, and decentralized-distributed-hybrid SNSs (cf. Baran 1964, p. 2). Circles are representations of participants in the social network, boxes are SNS providers. In the distributed and hybrid scenarios, white boxes indicate no distinction between users and providers.

able, as the defunction of one SNS only affects its users without affecting the communication network as a whole. The distributed scenario is the least vulnerable.

If SNS 2 in the centralized scenario (Figure 6, top-left) goes offline due to technical problems, Bob, Alice, and all other users on this SNS are no more able to communicate with each other. If SNS 2 in the distributed scenario (Figure 6, bottom-left) goes offline, this would only have an impact on one participant in the network, while all other participants would still be able to communicate.

The distributed scenario does not follow a *client-server* like model with a distinction of user and provider. Instead, it follows the principles of a *peer-to-peer* model (P2P; see Heyer, Holz, and Teresniak 2012, in this issue), in that all instances are equal in terms of technical connectedness with links based on social relations. Beside these strict architectures, hybrid scenarios of decentralized and distributed SNSs with a common OSN are possible, where, for example, Bob participates as a client of the decentralized SNS 1, while Alice uses the dedicated SNS 3 in a distributed way (see Figure 6, bottom-right).

By following this taxonomy, contemporary SNSs like Facebook can be seen as centralized and separated from other networks. Email services are decentralized, as servers can communicate with each other, allowing for provider independent communication of their clients. *Skype*[35] (in 2011 the most successful Voice-over-IP service on the Internet) or Instant Messaging systems like *ICQ*[36] are distributed, using P2P technology[37]. Weblogs can be seen as hybrid, as there are dedicated blogs as well as popular services like *WordPress.com*[38]. That means in consequence, for example, although when access to the site WordPress.com is limited in China, dedicated instances of WordPress (the open source blogging software running WordPress.com) can still be available (Newey 2009).

Admittedly, as most decentralized architectures allow for dedicated instances as well as servers for multiple clients (e. g., every user is allowed to set up her own email server), decentralization can be seen as a range: From a few interoperable instances, serving as providers for lots of users, to distributed systems with dedicated instances for each user. We will use the term “decentralized” in this chapter in contrast to “centralized” and will refer to “distributed” as being fully decentralized. In recent literature, the term “federated” was used to describe the same concept of an interoperable, decentralized social web (Prodromou 2010).

3.2. Privacy and legal issues

In centralized scenarios, the user is dependent on the separated OSN of the SNS she participates in, as well as on the provider of the service. These providers are in full control of all information on the users, her relations and her interactions. Regarding the diffusion of information under commercial aspects (see Section 2.5) and against the background that SNSs in most cases enable third-party applications use on this information, this can have a significant impact on the members’ privacy.

But a lot of the benefits from SNSs depend on the privacy the user can expect. When Facebook launched in 2004, the service was restricted to students having a harvard.edu email address. While this fact limited the value of the network[39], it “contributed to user’s perceptions of the site as an intimate, private community” (boyd and Ellison 2007). This privacy, however, is doubtful.

As the amount of data a user provides to the public, her friends, and her SNS provider is immense (see Section 2.1), the protection of this information is a significant challenge. This challenge is not in the user’s hands alone but also in the hands of the provider, who (in most cases) has a commercial interest in making use of the data.

The question, to which degree the service provider has the right to make use of the data, is mainly open and can be paraphrased to the question, if the data provided in centralized SNSs is of public or private nature. Due to the recency of SNSs, legal practice in judging the privacy of this new technique is relatively vague. In case of the major SNSs, the protection of privacy and data has to follow the jurisdiction of the USA.[40] Hodge (2006) discusses court decisions regarding (mostly technically) mediated communication based on the Fourth Amendment to the United States Constitution and their impact on privacy issues on Facebook and MySpace[41]. He analyzes that, if a communication act is private or public depends on a couple of factors: The mentioned “expected privacy” of a user is important, but this expectation has to be “one that society is prepared to recognize as reasonable” in order to be protected (Hodge 2006, p. 113). In fact, the Supreme Court "consistently has held that a person has no legitimate expectation of privacy in information turned over to a third person" (Hodge 2006, p. 114), as is the case with centralized SNSs.

Another aspect concerning the privacy of the data is the intended recipient of the communication. While phone numbers are intended to be of value for the phone company to connect the user with the intended conversational partner, the content of the call is meant to be addressed only to the partner. That means, the phone company is the intended recipient of the phone number, while the addressee is the intended recipient of the content. The same is true for the address on a mail envelope in opposite to the letter: The postal service has a legitimate business purpose to use the address on the envelope, as it is necessary for providing the desired service. The user of a mail service thus has no expected privacy regarding the address on the letter's envelope (Hodge 2006).

While a mailman has no business interest in opening and reading a letter, Facebook, Google and the like, on the other hand, have a business purpose to analyze the content of profile information or emails, or apply SNA on the user's contact list to advance their services, including their advertising model. The same aspect is important for storing this information: While an Internet provider has no business purpose to store information of their customers aside their needs for billing, Facebook advances its advertising with this knowledge (Hodge 2006). Because of this legitimate business purpose it is arguable if a user has a reasonable expected privacy in these services.

Noticeably, most SNSs allow for changing the degree of publicity of their profiles and contact lists in regards to other users, limiting it to be viewable, for example, only to direct friends (see Section 2.1). By explicitly limiting a profile, an expected privacy may be seen as being expressed. In this case a limited profile "would be more like a phone call or a sealed letter in this aspect" (Hodge 2006, p. 118), and by this matter being protected by the Fourth Amendment.

But, in the end, information on SNSs "can only be known to be private if the information they contain is not at any point given to an intermediate service provider" (Lucas 2007), as restrictions on data access is not effective to the SNS provider. That means, privacy can be best protected in distributed scenarios.

In addition to the use of personal data for commercial interests, SNSs also have control over the provided information. The provider is allowed to ban users if their profiles do not fit its interests. Take, for example, the rebellion of users against Friendster regarding its "Fakester Genocide" politics (boyd 2004), when entertaining fake profiles were rigorously deleted, or the defunction of user accounts with pseudonyms in Google+ (Pfanner 2011). The same can be true for content: The algorithm Facebook uses to rank messages in the "top news" stream (the so-called "EdgeRank" algorithm) is, for example, in full control by the provider to personalize the experience of the SNS[42], as is Facebook's filter mechanism to prevent the transmission of unwanted messages (Singel 2009).

The provider can control the diffusion of information within the centralized OSN of its SNS and has the right to change its general terms and conditions about all data whenever needed to optimize the service. In a distributed OSN, on the other hand, the user has potentially full control on her data and the degree of personalization of her SNS.

Beside already mentioned technical charges like "Denial of service" attacks, censoring, and governmental blocking of SNSs to limit the accessibility of data provided by SNSs (see Section 3.1), there also exist attacks to incriminate the privacy of SNSs' users. Various of these attacks are not restricted to centralized scenarios, as they are based on social engineering mechanisms (see Remark 23.1 in Cutillo, Manulis, and Strufe 2010, p. 510) – however, in distributed SNSs it is up to the user to protect her data and her privacy. And in a decentralized scenario the user can at least choose the provider she trusts the most[43].

4. Open protocols and formats

> "This is not a story of startups and entrepreneurs.
> This is not a story about who will become the next
> ~363kg gorilla.
> This is not a tale of who will next be crowned king.
> This is a story about … Wait for it …
> Server-side software implementations and open,
> documented protocols."
> Chisari (2011)[44]

The idea of decentralized OSNs is based on the principles of the decentralized conceptual layers of the technical communication networks (remember Figure 1). The Internet as the first conceptual layer is a decentralized computer network. It relies on open standards for information transmission by means of the Internet Protocol Suite. The World Wide Web as the second conceptual layer is a decentralized document network. It relies on open standards for document linkage by means of HTTP, URIs, and HTML. The third conceptual layer is meant to be a decentralized object network. Like both of the underlying layers, it relies on open, royalty-free standards.

As seen in Section 2.1, a SNS consists of three building blocks: *profiles*, *relations*, and *interaction tools*. To make these components vendor independent interoperable in a decentralized or distributed way, they need to be specified by means of open protocols and formats. In addition to that, secure mechanisms for the authentication of a user and the authorization for the access of resources are necessary.

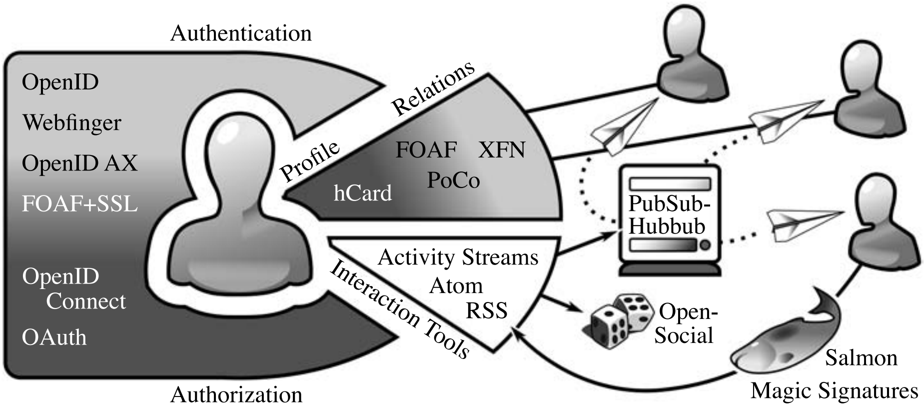


Figure 7. A user in the social web needs mechanisms for authentication and authorization (Section 4.1), tools for presenting profile information and social relations (Section 4.2), and protocols for real-time publishing and interactions (Section 4.3).

Figure 7 outlines the architecture presented in this Section. The following discussion on these topics can only show an excerpt of the wide variety of open protocols and formats in the realm of SNSs. For lots of the discussed use cases there are alternatives – gaining small differences up to completely different approaches.

4.1. Identification

Internet Protocol addresses are necessary for the identification of computers in the Internet, and URIs (Uniform Resource Identifier) are necessary for the identification of resources in the WWW. In the same manner identifiers for individuals are crucial for the social web[45]. Nowadays, as most of the social activities on the web happen on SNSs, users have unintentionally multiple identities on the social web: They have accounts on Facebook, Twitter, Google and so on. Beside the aforementioned technical and privacy related issues, this can lead to practical problems. When participating in several SNSs, the user initially has to create an account for each one of these, which means to type in the same personal information over and over again. Additionally, the user has to maintain her multiple accounts occasionally when information changes, like location or relationship status. Status messages (e.g., *Tweets*) have to be send multiple times to multiple SNSs to reach all possible contacts, which can, in consequence, lead to repetitive transmission of information to the same person in different SNSs. Another drawback is to keep track of replies on all these platforms. A lot of users therefore use aggregation services to bundle streams from various

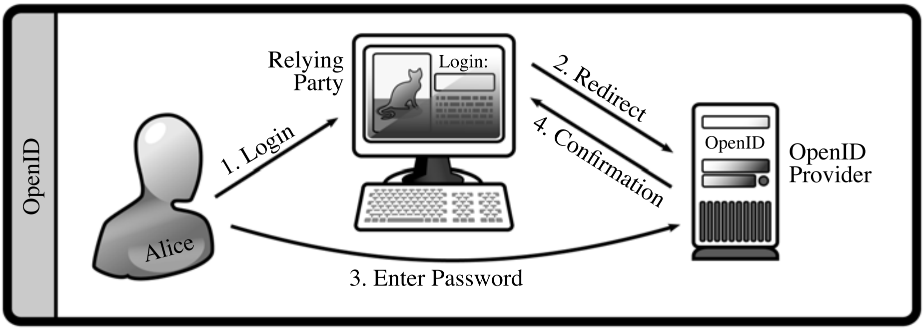


Figure 8. Four basic steps in a (simplified) OpenID authentification: 1) the user types her OpenID into the login form of a relying party site, 2) the relying party site redirects the user to the associated OpenID provider, 3) the user authentifies herself to the provider by inserting, for example, a password, then 4) after authentifying successfully, the OpenID provider tells the relying party that the user is rightfully associated to the given OpenID.

SNSs, showing last tweets from Twitter, recent videos from their YouTube channel, last uploaded pictures on *Flickr*[46] and recent bookmarks on *Delicious*[47].

This practical challenge often also leads to security problems, as users tend to use the same password on several sites (Riley 2006). Security vulnerabilities on one of these sites then can make all other sites vulnerable as well – and identity theft an easy task.

A solution to this problem would be a unique identity in the OSN, that can be used across all sites. To provide such a solution, two aspects have to be taken into account: First, a way for the user to authenticate herself to the service, and second, to grant the service the allowance, for example, to retrieve information about the user, that is, a way for authorization.

OpenID (The OpenID Foundation 2007) is an open standard for decentralized authentication. It is meant to be a single-sign-on solution, that is, the user can apply her OpenID to sign-on for several services instead of creating new identities for each one. The OpenID therefore is an URI. With this mechanism, a user needs only one URI and one password throughout the whole social web, having one single unique identity.

In practice, when using the OpenID to log in to a service (the so-called "relying party"), the user enters her URI first. Alice, for example, has the OpenID https://openid-server.example.net/~alice. The relying party then redirects Alice to her OpenID provider where she inserts her password for authentification. The OpenID provider afterwards redirects back to the relying party and notifies it that the given OpenID is correctly associated with Alice (see Figure 8).

Because it is rather uncommon for users to accept URIs like https://openid-server.example.net/~alice as their unique online identities, the second step in Figure 8 (the redirection to the login form of the OpenID provider) is rarely applied directly. OpenID allows for indirect identifiers, forcing the relying party to discover the OpenID provider of the user based on the OpenID that does not necessarily have to lead to the provider. Assume Alice has a blog at https://example.org/~alice, she can use this URI as her OpenID. In the HTML document of her blog she has to give meta information on where her OpenID provider can be found (by using a <link />-tag in the <head /> section of the HTML document), the relying party then discovers this information and redirects to her provider[48]. In this way, Alice's unique identity is directly related to information about her that can be found on her blog.

Although, URIs are neutral regarding the resource they locate to, people usually don't think of themselves as being represented as URIs. This led to other forms of identifiers in the discovery process of OpenID, with using email-like addresses as being the preferred variant.

The *Webfinger* protocol (Fitzpatrick et al. 2010; Jones, Salgueiro, and Smarr 2011) provides email like addresses as identifiers and allows for the discovery of *OpenID*. It was designed based on the *Finger* protocol (Harrenstien 1977), that was used to provide information on user accounts on different computers in a network. Webfinger introduces a new scheme prefix "acct:" (to differ from the email scheme prefix "mailto:") and allows for the discovery of meta information on an account name like acct:alice@example.org by requesting a document from the given domain called "host-meta" (Hammer-Lahav and Cook 2011) that can be found in a known location on the server (the "/.well-known" location; Nottingham and Hammer-Lahav 2010). The returned document is an *Extensible Resource Descriptor* file (XRD; Hammer-Lahav and Norris 2010), that contains meta information on the requested domain[49].

A possible link of type *lrdd* (*Link-based Resource Descriptor*) in this document can contain a template for the account URI to be requested, for example http://example.org/webfinger?q={uri}. By retrieving http://example.org/webfinger?q=acct:alice@example.org then, a new XRD file is returned, containing account specific information like the user's OpenID endpoint or her profile information.

While OpenID is the dominant decentralized single-sign-on system for authentication on the web, it has many drawbacks, mainly due to the diversity of specifications for implementers and the minor awareness and acceptance by users. Proprietary centralized approaches like *Facebook Connect* or *Sign in with Twitter* are currently more successful and widely adopted. In addition to OpenID, they do not limit their functionality to authentication but also allow for authorization in the same step, using an open protocol called *OAuth* (Hammer-Lahav, Recordon, and Hardt 2011). Facebook and Twitter are using this protocol also for their application programming interfaces (APIs). OAuth enables a user to authorize a third party with limited access to her data without sharing password information.

OpenID Attribute Exchange (OpenID AX; Hardt, Bufu, and Hoyt 2007) is an extension to OpenID to combine authentification and data exchange in a similar way to, for example, Facebook Connect. A more recent approach is *OpenID Connect*[50].

The World Wide Web Consortium summarizes solutions to authentication under the term *WebID*[51]. A current approach in this workspace is *FOAF+SSL* (Story et al. 2009), that uses SSL (*Secure Sockets Layer*) and *FOAF* (*Friend of a Friend*; see next Section) for the establishing of a *web of trust* based authentication infrastructure.

4.2. Profiles and relations

Information on both the profiles and the relations has to be accessible in a human as well as in a machine readable format. The World Wide Web introduced HTML as a standard for hypertext documents for this purpose. In the semantic web, now, there are several different formats to define different resources (Waltinger and Breuing 2012, in this issue). In this Section we will give a brief overview on formats relating to profile and relationship data in the social web (cf. Mika 2007). These formats differ in their expressiveness, their objective, and their embeddedness in the WWW.

Early electronic standards for personal data sets were defined independently from the web. *vCard* (Internet Mail Consortium 1996), for example, defined an electronic business card format. Listing 1 shows an example vCard of the user Alice.

Listing 1. vCard of Alice Example.

```
BEGIN:VCARD
VERSION:4.0
FN:Alice Example
ORG:Example Inc.
URL:https://example.org/~alice
EMAIL;PREF=1:alice@example.org
END:VCARD
```

To place this electronic business card information on the WWW, *hCard* (Çelik and Suda 2005) was developed as a 1:1 representation of vCard by means of microformats[52]. Microformats are semantic annotations in HTML, using established HTML attributes like "class" or "rel" to formulate relations between elements and contents on a webpage, or between the document and linked resources. Listing 2 shows the vCard of Alice as an HTML-embedded hCard.

Listing 2. hCard of Alice Example.

```
<div class="vcard">
<a class="url fn" href="https://example.org/~alice">
AliceExample</a>
<a class="email" href=
"mailto:alice@example.org">alice@example.org</a>
<div class="org">Example Inc.</div>
</div>
```

Microformats provide vocabularies for a wide range of domains. Regarding relations in OSNs they define the *XHTML Friends Network* vocabulary (XFN; Global Multimedia Protocols Group 2003). The declared target of XFN is the annotation of links, especially as part of “blogrolls” (see Lenhart 2005, pp. 122–133), to form social networks in the so-called “blogosphere” with elaborated relationships.

XFN allows for fine grained annotations in respect of the type of the tie between two individuals, regarding the strength (defined values are “Friend”, “Acquaintance”, and “Contact”), the domain of relationship (e.g., “Professional”, “Romantic”, or “Family”) and even types of relationships within the domain (e.g., “Sibling” in the “Family” domain). In addition to the basic directed and undirected graph models of SNSs, this introduces asymmetric relations, as a “Child” relationship is inverse to a “Parent” relationship.

Microformats were invented to enhance HTML with semantic information instead of rebuilding the web by means of new semantic technology, primarily the *Resource Description Framework* (RDF; Brickley and Guha 2004). RDF is one of the fundamental components of the semantic web. It is a formal language to model decentralized knowledge by means of <subject, predicate, object> triples, like “Alice knows Bob”. With this method, RDF allows for defining directed, labeled graphs with multiple edges. The distributional character is introduced by allowing URIs for the graph nodes and namespaces for different predicate vocabularies.

The *Friend of a Friend* (FOAF; Brickley and Miller 2010) namespace defines an RDF vocabulary to provide knowledge about an individual. On the one hand, this is profile-like information, as the name, title, gender, homepage, or the workplace. On the other hand, FOAF enables to formulate relationships with other individuals.

In a decentralized SNS the user is in control of her FOAF information. With the “knows” predicate she is only able to define an asymmetric relation: If “Alice knows Bob” there is no conclusion that “Bob knows Alice”. This is similar to the relationship model in Twitter. To reason a symmetric or bidirectional relation between Alice and Bob (like Facebook’s “Friendship”), a system has to follow the URI of the object, analyze the FOAF information of Bob and has to check for the triple “Bob knows Alice” [53].

Additional namespaces for personal information that can be embedded in RDF are biographical (Davis and Galbraith 2002), calendaric (Connolly and Miller 2005), or geographical information (Brickley 2003).

With RDFa (Adida et al. 2008), an extension to the syntax of XHTML[54], a specification was suggested that makes it possible to enrich current webpages with semantic information using these RDF triples likewise to microformats.

A similar approach to the same data but with a different focus is the *Portable Contacts* specification (PoCo; Smarr 2008). While it deals with profile data and contact data in a comparable fashion like FOAF (with using a vocabulary widely borrowed by vCard for profile data and XFN for contact relation types), the main objective is to provide a standard mechanism for the secure access and exchange of this information between different services, for example, to allow a SNS to access the address book of a user's email provider to search for known contacts in the OSN or for the export of a user's digital address book.

4.3. Interaction

Most interactions in SNSs are time-aligned, that means, the information is shared and aggregated in chronological order of publishing, for example, status updates or "wall" postings. A common way to distribute time-aligned information on the semantic web are so-called feeds, for example, in RSS or Atom formats (Winer 2009b; Gregorio and de hOra 2007). These formats are based on XML and serialize entries of small information. A specification to describe entries typical in the context of SNSs is *Activity Streams* (Atkins et al. 2011). It is an Atom or JSON (Crockford 2006) based format, where each entry holds information on the activity's author, the kind of activity (e.g., "post"), the object of activity (e.g., a "note"), and possibly the target (e.g., a blog).

For SNSs, real-time notifications of new activity events like these are important. When, for example, the user Bob is a friend of the user Alice and writes a new status message on his activity stream, Alice is immediately notified by her news stream. In centralized scenarios, all information is stored at one point, so the stream of information Alice gets is directly served by her SNS provider. In a decentralized scenario, the activity streams of Alice's friends can be stored completely separated, which makes synchronization of all feeds more complicated. The SNS needs to fetch all her friends' activity streams from the web to aggregate it in her news stream. When combining this with real-time information, the server has to fetch all streams regularly, which results in bad performance if Alice has a lot of friends (see Figure 9, top). In favour of regular requests to Bob's server ("pull"), the system can alternatively send notifications to all subscribers ("push"), which are, in a SNSs context, Bob's "friends" or "followers". These server-to-server notifications in favour of regular pulling have lately been named *WebHooks*[55]. *PubSubHubbub* (or *PuSH*) (Fitzpatrick, Slatkin, and At-

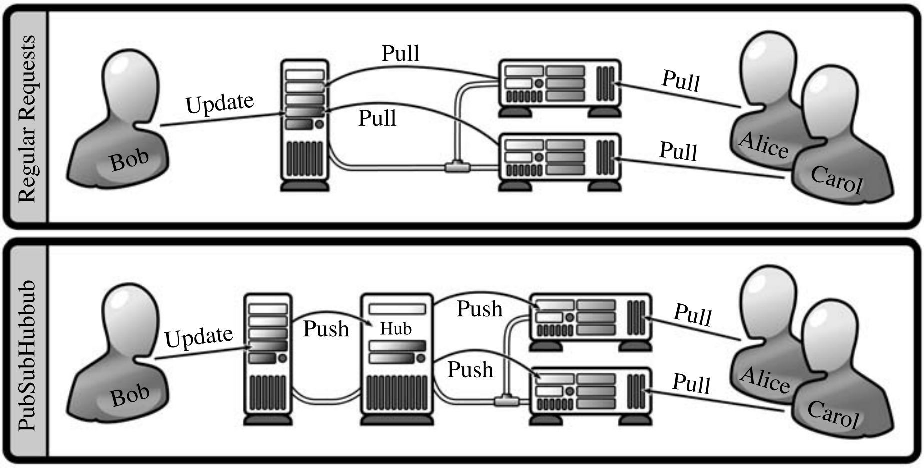


Figure 9. Regular request mechanism and PubSubHubbub.

kins 2010) is a push service using webhooks with a mediating hub server[56]. Instead of fetching all streams of Alice's friends regularly, Bob's server, after he published a new status message, sends a notification to a hub, where all his friends subscribed to his activity stream (see Figure 9, bottom). The hub then fetches Bob's new status message and forwards it to all subscribers. Afterwards the subscribers are able to update their news streams in real-time. In this way, PubSubHubbub delegates time-consuming distribution mechanisms to a dedicated system. A similar approach to real-time feed publishing with mediating hubs is *rssCloud* (Winer 2009a).

For user-to-user interactions (e.g., sending responses to status updates or to "like" a posting) in a decentralized way an additional mechanism is needed, as a system has to know, where and how to send a response. The *Salmon* protocol (Panzer 2010) defines a specification for such a mechanism. It allows for sending Atom entries to specified endpoints as, for example, replies to feed entries by another user. If the user Bob wants to respond to an Activity Streams entry by Alice, and this entry serves a Salmon endpoint in form of an URI, he can post the entry directly to Alice's feed[57].

For some events, like the notification a user was tagged in a photo on an otherwise non-related resource, the user can be notified without responding to a resource by using a Salmon endpoint directly associated to the user's account, discovered using Webfinger. If the user Bob, for example, wants to mention Alice in a posting, the Salmon generator discovers her Salmon endpoint by applying Webfinger and posts the message to Alice.

To authenticate Bob as the author of the message, Salmon allows for signing the message with *Magic Envelopes* (Panzer, Laurie, and Balfanz 2011). Magic Envelopes provide a Public Key Infrastructure (PKI) with a private key for signing messages and public keys for author verification. By applying Webfinger to discover the public key of Bob and verifying the signature of the Magic Envelope of the message, Alice can ensure that Bob was the original author of the message. This helps to establish a reliable association between an authenticatable author (see Section 4.1) and her publication.

The Diaspora project as well as *Friendika*[58] provide additional envelope format specifications for the Salmon Magic Envelopes to fully encrypt the messages, based on PKI as well.

All these protocols and formats are able to provide the functionality of both communication as well as presentation tools for decentralized SNSs (see Section 2.1).

Regarding decentralized entertainment tools, the *OpenSocial* (OpenSocial and Gadgets Specification Group 2010) framework provides a bundle of standardized APIs to create social applications, games and widgets (e.g., gifts) with access to profile information, contacts and several other features of SNSs (Le Blanc 2011).

For a more comprehensive overview regarding the technical architecture of decentralized SNSs, including various use cases and further information on aspects like privacy standards, refer to the publications of the W3C Social Web Incubator Group (2010).

5. Projects

Although necessary building blocks for the infrastructure of a decentralized OSN by means of open protocols and formats already exist, the social web is still dominated by centralized SNSs. In this Section, we want to give an overview on several projects (currently) working on decentralizing SNSs. This list is not meant to be comprehensive, nor does it focus on approaches to decentralized SNSs only, but it shall reflect different flavors and directions of ongoing work in this field with a special emphasize on the aforementioned protocols and formats.

In 2010, *StatusNet Inc* – the company behind the microblogging SNS *Identi.ca*[59] – proposed an open standard for distributed status updates as a suite of open protocols called *OStatus* (Prodromou et al. 2010)[60]. It included the previously introduced specifications of the Salmon protocol, Activity Streams, Webfinger, a subset of Portable Contacts, and PubSubHubbub for publishing of and subscribing to status updates in real-time.

Nowadays OStatus is implemented in several software products[61], considerably in the StatusNet open source software, that runs several instances on the web – with Identi.ca being the most popular among the public ones. While it was originally designed for the decentralization of SNSs based on the asymmetric model, projects like GNU social extend its functionality, aiming for decentralized, symmetric, Facebook-like SNSs. The already mentioned Diaspora project was launched in 2010 with the identical goal. While initially starting with specialized protocols and formats, in recent development stages Diaspora adapted the specifications of OStatus as well, allowing to interact with users of StatusNet or Friendika, which is another OStatus compliant SNS.

The *DiSo*[62] project does not aim on creating a SNS, but provides plug-ins for blogging software and content management systems like WordPress, *Movable Type*[63], or *Drupal*[64] to support features for the social web in terms of open protocols and formats. This is closely related to numerous efforts to provide integration of (decentralized) social networking functions into existing blog and content management software, for example *socialriver*[65], a WordPress package based on OStatus.

These approaches are embedded in the basic infrastructure of the World Wide Web, using URIs, HTML, and HTTP. *OneSocialWeb*[66], a decentralized SNS effort by *Vodafone*, supports a lot of the aforementioned open formats and protocols (including OStatus), but has a different approach regarding server-to-server communication: The communication does not rely on HTTP but on *XMPP* (Jabber Software Foundation 2004), an open protocol for decentralized instant messaging formerly known as "Jabber". As the XMPP infrastructure is build for decentralized real-time communication, this setup gains more and more popularity. Another project using XMPP for SNSs is *buddycloud*[67].

Although most of the current projects working on decentralized SNSs aim to support the OStatus specification, competing mature protocols exist as well. For example the *Appleseed*[68] project uses its own specification (called *QuickSocial*) as does *NoseRub*[69].

Most projects in the field currently focus on decentralization in favor of distribution. That is, hubs in forms of SNSs provide an entrance to the social network for multiple users, while not every user is in need for a dedicated SNS portal (see Section 3). This is along the lines with email, as not every user has to set up her own email server. In some ways, this can be risky as, for example, the provider still has control of the user's data. As seen in Section 3, full distribution gives a better protection regarding privacy concerns as well as regarding technical issues. Projects trying to establish real distributed SNSs focus on private servers for every individual (like the *FreedomBox*[70] project) or use applications on mobile devices (Tramp et al. 2011), that even work without a connection to the Internet.

Some of these systems even emphasize full distribution of the OSN over embedding the social network decentralized in the World Wide Web[71]. Systems of this focus make use of P2P technologies (see Heyer, Holz, and Teresniak 2012, in this issue) and mostly emphasize privacy issues with the use of encryption. Projects include, for example, *Safebook*[72] (Cutillo, Molva, and Strufe 2009) and *PeerSoN*[73] (Buchegger et al. 2009).

Beside these projects, major companies now implement several of the illustrated specifications in their services and in some cases even initiated the development. For example, accounts on Google.com or Yahoo.com can be used as OpenIDs and are discoverable via Webfinger; *Google Buzz*[74] had a similar architecture to OStatus; Facebook's stream API provides the Activity Streams standard; Orkut uses XFN for some annotated relations in a user's profile; and, together with MySpace and several others, it relies on OpenSocial for its applications.

6. Conclusion

In recent years, the social web is gaining momentum on the Internet. Especially Social Network Sites (SNSs) like Facebook made online social interactions immensely popular to a wide range of online users.

This chapter outlined the characteristics of SNSs from both a sociological as well as a technical point of view (see Section 2). SNSs were introduced as services that allow for interaction within an Online Social Network (OSN). The success of these platforms was shown to be reasonable due to social benefits like keeping in touch with friends or acquaintances, and being able to monitor their activities without much effort.

Currently, most of the social activities on the World Wide Web happen on centralized SNSs. Critical aspects regarding this architecture were presented, especially the implications regarding technical vulnerabilities and privacy (see Section 3). SNSs based on decentralized OSNs were introduced as a less vulnerable and privacy-aware alternative approach. Open protocols and formats were illustrated as necessary building blocks for vendor independent interoperable SNSs in a decentralized scenario (see Section 4) and a brief overview on currently developed projects in this field was given (see Section 5). These projects try to offer the benefits of SNSs without having most of their drawbacks.

For the moment, future development in this field is not foreseeable. Mostly all specifications of protocols and formats shown in this chapter are still in development or even in draft status. Projects are in early stages to implement common specifications for interoperability in the social web, and although major companies support these developments, the future of a decentralized OSN is open.

But the attention in the Diaspora project has shown that online users are interested in alternatives to centralized SNSs. And the movement within the pool of projects to come up with an interoperable standard, like the mentioned OStatus, is promising. The diversity of all these projects has the potential to result in a wide variety of interoperable clients of one OSN – instead of one platform – that could be able to allow each user to individually choose her wanted level of privacy and features of her SNS. This development may lead to new forms of communication in OSNs other than described in this chapter, with new possibilities for the diffusion of information, the maintenance of social ties, and the forms of collaboration.

The SNSs described in this chapter can be seen as only one specific application of this social web. New social software may be developed based on a decentralized OSN that is less vulnerable regarding technical and privacy related issues, and could be more useful for target groups, for example, with a business or dating focus, for different ages, or with a focus on barrier-free access (see Kubina and Lücking 2012, in this issue) without the need for separated OSNs.

This scenario – while at this moment highly speculative – would possibly replace the "walled gardens" of the current social web and lead to a decentralized and open "new social web" (Salzberg et al. 2011).

Notes

1. https://www.facebook.com/
2. In this chapter we will distinguish between the terms "Social Network Site" (SNS), as introduced by boyd and Ellison (2007), and "Online Social Network" (OSN), as the more recent term for – in most publications – the same subject (see, e.g., Datta et al. 2010). We use the term SNS to refer to the service portals (e.g., Facebook or Twitter) that allow for participation in a digital social network of individuals and social relations, which we will refer to as an OSN.
3. Similar studies confirmed these findings (see, e.g., Dodds, Muhamad, and Watts 2003, for an email-based approach). However, there is some critique (e.g., Kleinfeld 2002) regarding the conclusion of a *"small world"* based on the results of the experiments by Travers and Milgram (1969) and others, as the communication paths were broken in the majority of experiment runs, so only successful acquaintance chains were taken into account.
4. The term "Six Degrees of Separation" was especially popularized as the title of a play written by John Guare (1990).
5. http://www.friendster.com/
6. http://www.myspace.com/
7. http://www.orkut.com/
8. https://twitter.com/
9. https://plus.google.com/
10. This extension is along the lines with Datta et al. (2010).
11. http://foocorp.org/projects/social/
12. http://status.net/
13. http://joindiaspora.com/
14. In *Weaving the Web*, Berners-Lee conceptualized a four layer infrastructure of the

web, consisting of "the transition medium, the computer hardware, the software, and the content" (Berners-Lee 2000, p. 130).
15. We refer to the advent of standards of the Internet Protocol Suite.
16. For a discussion of the types of information provided by users of SNSs, refer to Cutillo, Manulis, and Strufe (2010, pp. 503–506).
17. "Microblogging" refers to a blogging service specialized in short texts. Twitter, for example, has a limit on 140 characters (McFedries 2007).
18. http://www.researchgate.net/
19. For a survey on the history of SNSs, see boyd and Ellison (2007).
20. http://www.studivz.net/
21. A fact, boyd (2004) and boyd (2005) state for the symmetric SNS Friendster, while it seems to be quite common to unfollow someone on asymmetric SNSs like Twitter, as this does not imply any social or (unwanted) technical disprofits (Kwak, Chun, and Moon 2011; Kivran-Swaine, Govindan, and Naaman 2011).
22. For a model that formalizes a network which allows for cliques – in this case called the *clustering coefficient* –, see Watts and Strogatz (1998). The model says that not only the average path length between any two nodes in a small world network is small, but there is also a high clustering coefficient.
23. Ties, as discussed in Granovetter (1973), are symmetric and positive – like relations in Facebook rather than in asymmetric SNSs like Twitter.
24. Dunbar's number was an inspiration for the limitation of 150 friends of the mobile SNS *Path* (https://path.com).
25. For a discussion on the number of persons an individual knows during her lifetime, see Killworth et al. (1990).
26. Although the "favorite" function of Twitter is semantically closer to the "like" function of Facebook, it has not the same relevance for the diffusion of information, as these items are shown separated to the main stream of information.
27. This principle follows the "80/20 rule", that says, that a small proportion of offered products generate a large proportion of sales (for example, 20 % of the products generate 80 % of the sales).
28. *Amazon.com* (http://www.amazon.com/), for example, makes more than 20 % of their book sales with products below the rank of their top 100,000 products, while most traditional book stores are limited to 100,000 products at all (Brynjolfsson, Hu, and Smith 2003; Fenner, Levene, and Loizou 2010).
29. https://www.google.com/adsense/
30. For further information regarding SNA, see Wasserman and Faust (2008) and Jackson (2008).
31. http://www.youtube.com/
32. The mechanism is closely connected to the phenomenon of *Internet Memes*. Originally based on a term coined by Dawkins (1976, pp. 203–215), it references to all kinds of popular ideas and content diffused via social software and especially SNSs (Hodge 2000).
33. The frequent performance issues of Twitter led to the popularity of the infamous "fail whale", Twitter's mascot for server related problems.
34. http://www.pownce.com/
35. http://www.skype.com/
36. http://icq.com/

37. For several services, P2P systems also depend on central servers and thus are not purely distributed. For a taxonomy regarding the centralization of P2P systems, refer to Vu, Lupu, and Ooi (2010).
38. http://www.wordpress.com/
39. The value of the network is limited according to “Metcalfe’s law”, that roughly says, the value of a communication network increases more than linear as the number of nodes increase, because of the increased number of potential links for every node (Hendler and Golbeck 2008).
40. In theory, some of these companies participate in the “safe harbour” agreement between the European Union (EU) and the United States of America, that should assure citizens of the EU getting the same private data protection in the USA as in their home country regarding self-certified companies. But according to negative reviews of the EU and recent studies, the implementation is rather weak and misleading, and even representing “a new and significant privacy risk to consumers” (see Connolly 2008). In 2008 Facebook established a new headquarter in Dublin (Facebook 2008), which changed the legal base of privacy protection.
41. Due to the aforementioned recency and the ongoing debate regarding privacy in OSNs, this chapter will be limited to a shallow legal view on the topic.
42. Personalization in this way is believed to have serious impact on the diffusion of information as it prefers information from within cliques rather than weak ties. This effect, among other ways of personalization, was coined the “Filter Bubble” by Eli Pariser (2011).
43. For a survey on privacy attacks on SNSs, see Cutillo, Manulis, and Strufe (2010).
44. Michael Chisari is lead developer of the Appleseed project (see Section 5).
45. Individuals mostly prefer to have multiple identities in different contexts, that means sometimes they deliberately represent themselves with multiple accounts on different OSNs. We will refer to identity as being one of possibly many different identities of one individual. For more information on identity in the social web, refer to Maheswaran et al. (2010) and Farnham and Churchill (2011).
46. http://www.flickr.com/
47. http://www.delicious.com/
48. As OpenID is an open specification, a user can, of course, be her own provider.
49. Because most of the open formats and protocols in this workflow were invented or co-developed by Eran Hammer-Lahav, this is called the “Hammer Stack”.
50. https://openid.net/connect/
51. http://esw.w3.org/WebID
52. http://microformats.org/
53. RDF data can be accessed using the powerful query language *SPARQL* (SPARQL Protocol and RDF Query Language).
54. XHTML is a variant of HTML based on XML rather than SGML.
55. http://www.webhooks.org/
56. The hub server in PubSubHubbub can be seen as a SPOF (see Section 3.1), as there are only a few services online providing the services to choose from. However, if the hub is shut down, the availability of data is still assured.
57. To indicate a reply, Atom entries can use the *Atom Threading Extension* (Snell 2006).
58. http://info.dfrn.org/

59. http://identi.ca/
60. This approach was formerly known under the name of “OpenMicroBlogging”.
61. For an overview on OStatus compliant software, refer to Status.Net (2011).
62. https://code.google.com/p/diso/
63. http://www.movabletype.org/
64. http://www.drupal.org/
65. http://socialriver.org/
66. http://onesocialweb.org/
67. http://buddycloud.com/
68. http://appleseedproject.org/
69. http://noserub.com/
70. http://freedomboxfoundation.org/
71. For a survey on distributed SNSs on peer-to-peer basis, see Datta et al. (2010).
72. http://www.safebook.us/
73. http://www.peerson.net/
74. http://www.google.com/buzz; Google Buzz was shut down in 2011 in favor of Google+.

References

Adida, Ben, Mark Birbeck, Shane McCarron, and Steven Pemberton
2008 *RDFa in XHTML: Syntax and Processing*. Specification. → http://www.w3.org/TR/rdfa-syntax/ [Last accessed 09/08/2011].

Alexa
2011 *Top Sites. The top 500 sites on the web*. Website. September 2011. → http://www.alexa.com/topsites [Last accessed 09/08/2011].

Allen, Christopher
2004 *Tracing the Evolution of Social Software*. Blog Post. → http://www.lifewithalacrity.com/2004/10/tracing_the_evo.html [Last accessed 09/08/2011].

Anderson, Chris
2006 *The Long Tail: Why the Future of Business Is Selling Less of More*. Hyperion.

Atkins, Martin, Will Norris, Chris Messina, Monica Wilkinson, and Rob Dolin
2011 *Atom Activity Extensions 1.0*. Specification. → http://activitystrea.ms/specs/atom/1.0/ [Last accessed 09/08/2011].

Au Yeung, Chingman, Ilaria Liccardi, Kanghao Lu, Oshani Seneviratne and Tim Berners-Lee
2009 “Decentralization: The Future of Online Social Networking”. In: *W3C Workshop on the Future of Social Networking*. Barcelona, Spain.

Back, Mitja D., Juliane M. Stopfer, Simine Vazire, Sam Gaddis, Stefan C. Schmukle, Boris Egloff, and Samuel D. Gosling
2010 “Facebook Profiles Reflect Actual Personality, Not Self-Idealization”. In: *Psychological Science* 21.3, pp. 372–374.

Baran, Paul
1964 *On Distributed Communications*. Tech. rep. Memorandum, RM-3420-PR. Santa Monica, CA, USA: The RAND Corporation.
Bates, Claire
2009 *How Michael Jackson's death shut down Twitter, brought chaos to Google … and 'killed off' Jeff Goldblum*. The Daily Mail Online. 26 June 2009.
→ http://www.dailymail.co.uk/sciencetech/article-1195651/How-Michael-Jacksons-death-shut-Twitter-overwhelmed-Google--killed-Jeff-Goldblum.html [Last accessed 04/05/2012].
Beer, David
2008 "Social network(ing) sites … revisiting the story so far: A response to danah boyd & Nicole Ellison". In: *Journal of Computer-Mediated Communication* 13, pp. 516–529.
Berners-Lee, Tim
2000 *Weaving the Web. The Original Design and Ultimate Destiny of the World Wide Web*. With Mark Fischetti. HarperCollins.
2007 *Giant Global Graph*. Blog Post.
→ http://dig.csail.mit.edu/breadcrumbs/node/215 [Last accessed 09/08/2011].
2010 "Long Live the Web". December 2010. In: *Scientific American*.
boyd, danah m.
2004 "Friendster and Publicly Articulated Social Networking". In: *Proceedings of the Conference on Human Factors and Computing Systems (CHI 2004)*. ACM. Vienna, Austria.
2005 *Revenge of the Social Network: Lessons from Friendster*. Talk at Stanford's HCI Seminar. Palo Alto, California.
→ http://www.danah.org/papers/2005.0204.Stanford.txt.
boyd, danah m. and Nicole B. Ellison
2007 "Social Network Sites: Definition, History, and Scholarship". In: *Journal of Computer-Mediated Communication* 13.1.
Breslin, John G. and Stefan Decker
2007 "The Future of Social Networks on the Internet: The Need for Semantics". In: *IEEE Internet Computing* 11, pp. 86–90.
Breslin, John G., Alexandre Passant, and Stefan Decker
2009 *The Social Semantic Web*. Berlin/Heidelberg: Springer.
Brickley, Dan
2003 *Basic Geo (WGS84 lat/long) Vocabulary*. Specification.
→ http://www.w3.org/2003/01/geo/ [Last accessed 09/08/2011].
Brickley, Dan and R. V. Guha
2004 *RDF Vocabulary Description Language 1.0: RDF Schema*. Specification.
→ http://www.w3.org/TR/rdf-schema/ [Last accessed 09/08/2011].
Brickley, Dan and Libby Miller
2010 *FOAF Vocabulary Specification 0.98*. Specification.
→ http://xmlns.com/foaf/spec/ [Last accessed 09/08/2011].
Brynjolfsson, Erik, Yu (Jeffrey) Hu, and Michael D. Smith
2003 "Consumer Surplus in the Digital Economy: Estimating the Value of Increased Product Variety at Online Booksellers". In: *Management Science* 49.11, pp. 1580–1596.

Buchegger, Sonja, Doris Schiöberg, Le-Hung Vu, and Anwitaman Datta
2009 "PeerSoN: P2P Social Networking – Early Experiences and Insights". In: *Proceedings of the ACM Workshop on Social Network Systems*. Nuremberg, Germany

Burt, Ronald S.
1992 *Structural Holes. The Social Structure of Competition*. Harvard University Press.

Buys, Christian J. and Kenneth L. Larson
1979 "Human Sympathy Groups". In: *Psychological Reports* 45, pp. 547–553.

Çelik, Tantek and Brian Suda
2005 *hCard 1.0*. Specification.
→ http://microformats.org/wiki/hcard [Last accessed 09/08/2011].

Chisari, Michael
2011 *The End Of Facebook and Free Software's Quiet Revolution*. Blog Post.
→ http://developer.appleseedproject.org/profile/michael.chisari/journal/the-end-of-facebook-and-free-software's-quiet-revolution [Last accessed 09/08/2011].

Connolly, Chris
2008 *The US Safe Harbor – Fact or Fiction?*
→ http://www.galexia.com/public/research/assets/safe_harbor_fact_or_fiction_2008/safe_harbor_fact_or_fiction.pdf [Last accessed 09/08/2011].

Connolly, Dan and Libby Miller
2005 *RDF Calendar – an application of the Resource Description Framework to iCalendar Data*. Specification.
→ http://www.w3.org/TR/rdfcal/ [Last accessed 09/08/2011].

Crockford, Douglas
2006 *The application/json Media Type for JavaScript Object Notation (JSON)*. Specification.
→ http://www.ietf.org/rfc/rfc4627.txt [Last accessed 09/08/2011].

Cutillo, Leucio Antonio, Mark Manulis, and Thorsten Strufe
2010 "Security and Privacy in Online Social Networks". In: *Handbook of Social Network Technologies and Applications*. Ed. by Borko Furht. Berlin: Springer Science+Business Media. Chap. 23, pp. 497–522.

Cutillo, Leucio Antonio, Refik Molva, and Thorsten Strufe
2009 "Safebook: A Privacy Preserving Online Social Network Leveraging on Real-Life Trust". In: *Communications Magazine, IEEE* 47.12, pp. 94–101.

Datta, Anwitaman, Sonja Buchegger, Le-Hung Vu, Thorsten Strufe, and Krzysztof Rzadca
2010 "Decentralized Online Social Networks". In: *Handbook of Social Network Technologies and Applications*. Ed. by Borko Furht. Berlin: Springer Science+Business Media. Chap. 17, pp. 349–378.

Davis, Ian and David Galbraith
2002 *BIO: A vocabulary for biographical information*. Specification.
→ http://purl.org/vocab/bio/ [Last accessed 09/08/2011].

Dawkins, Richard
1976 *The Selfish Gene*. Oxford University Press.

Dodds, Peter Sheridan, Roby Muhamad, and Duncan J. Watts
2003 "An Experimental Study on Search in Global Social Networks". In: *Science* 301, pp. 827–829.
Domingos, Pedro and Matt Richardson
2001 "Mining the Network Value of Customers". In: *Proceedings of the 7th ACM SIGKDD international conference on Knowledge discovery and data mining*. New York, NY, USA, pp. 57–66. San Francisco, CA, USA.
Donath, Judith and danah m. boyd
2004 "Public Displays of Connection". In: *BT Technology Journal* 22.4, pp. 71–82.
Dunbar, Robin I. M.
1993 "Coevolution of neocortical size, group size and language in humans". In: *Behavioral and Brain Sciences* 16.4, pp. 681–735.
1995 "Neocortex size and group size in primates: a test of the hypothesis". In: *Journal of Human Evolution* 28, pp. 287–296.
1996 *Grooming, Gossip and the Evolution of Language*. 6th ed. Harvard University Press.
2010 *How Many Friends Does One Person Need? Dunbar's Number and other evolutionary quirks*. Faber and Faber Limited.
Engeström, Jyri
2005 *Why some social network services work and others don't – Or: the case for object-centered sociality.* Blog Post.
→ http://www.zengestrom.com/blog/2005/04/why-some-social-network-services-work-and-others-dont-or-the-case-for-object-centered-sociality.html [Last accessed 09/08/2011].
Facebook
2008 *Facebook to Establish International Headquarters in Dublin, Ireland.* Press Release.
→ https://newsroom.fb.com/Announcements/Facebook-to-Establish-International-Headquarters-in-Dublin-Ireland-cf.aspx [Last accessed 04/03/2012].
2011 *Statistics*. Website.
→ http://www.facebook.com/press/info.php?statistics [Last accessed 09/08/2011].
Farnham, Shelly D. and Elizabeth Churchill
2011 "Faceted Identity, Faceted Lives: Social and Technical Issues with Being Yourself Online". In: *Proceedings of the ACM 2011 conference on Computer supported cooperative work*. New York, NY, USA: ACM, pp. 359–368.
Fenner, Trevor, Mark Levene, and George Loizou
2010 "Predicting the long tail of book sales: unearthing the power-law exponent". In: *Physica A: Statistical Mechanics and its Applications* 389.12, pp. 2416–2421.
Fitzpatrick, Brad and David Recordon
2007 *Thoughts on the Social Graph*. Website.
→ http://bradfitz.com/socialgraph-problem/ [Last accessed 09/08/2011].

Fitzpatrick, Brad, Brett Slatkin, and Martin Atkins
2010 *PubSubHubbub Core 0.3 – Working Draft*. Specification.
→ http://pubsubhubbub.googlecode.com/svn/trunk/pubsubhubbub-core-0.3.html [Last accessed 09/08/2011].

Fitzpatrick, Brad, Eran Hammer-Lahav, Blaine Cook, John Panzer, and Joe Gregorio
2010 *The WebFinger protocol*. Specification.
→ http://code.google.com/p/webfinger/wiki/WebFingerProtocol [Last accessed 09/08/2011].

Friedkin, Noah E.
1982 "Information Flow Through Strong and Weak Ties in Intraorganizational Social Networks". In: *Social Networks* 3, pp. 273–285.

Global Multimedia Protocols Group
2003 *Xhtml Friends Network*. Specification.
→ http://gmpg.org/xfn/ [Last accessed 09/08/2011].

Granovetter, Mark S.
1973 "The Strength of Weak Ties". In: *American Journal of Sociology* 78.6, pp. 1360–1380.
1983 "The Strength of Weak Ties: A Network Theory Revisited". In: *Sociological Theory* 1, pp. 201–233.
1995 *Getting A Job. A Study of Contacts and Careers*. 2nd ed. Chicago/London: The University of Chicago Press.

Gregorio, Joe and Bill de hOra
2007 *The Atom Publishing Protocol*. Specification.
→ http://www.rfc-editor.org/rfc/rfc5023.txt [Last accessed 09/08/2011].

Guare, John
1990 *Six Degrees of Separation*. New York: Vintage Books.

Hammer-Lahav, Eran and Blaine Cook
2011 *Web Host Metadata*. Specification.
→ http://tools.ietf.org/html/rfc6415 [Last accessed 12/22/2011].

Hammer-Lahav, Eran and Will Norris
2010 *Extensible Resource Descriptor (XRD) Version 1.0*. Specification.
→ http://docs.oasis-open.org/xri/xrd/v1.0/xrd-1.0.html [Last accessed 09/08/2011].

Hammer-Lahav, Eran, David Recordon and Dick Hardt
2011 *The OAuth 2.0 Authorization Protocol*. Specification. This refers to Version 2.0.
→ http://tools.ietf.org/html/draft-ietf-oauth-v2 [Last accessed 09/08/2011].

Hardt, Dick, Johnny Bufu, and Josh Hoyt
2007 *OpenID Attribute Exchange 1.0 – Final*. Specification.
→ http://openid.net/specs/openid-attribute-exchange-1_0.html [Last accessed 09/08/2011].

Harrenstien, Ken
1977 *NAME/FINGER*. Specification.
→ http://tools.ietf.org/html/rfc742 [Last accessed 09/08/2011].

Hendler, James and Jennifer Golbeck
2008 "Metcalfe's law, Web 2.0, and the Semantic Web". In: *Web Semantics: Science, Services and Agents on the World Wide Web* 6.1, pp. 14–20.

Heyer, Gerhard, Florian Holz, and Sven Teresniak
2012 “P2P-based communication”. In: *Handbook of Technical Communication*. Ed. by Alexander Mehler, Laurent Romary, and Dafydd Gibbon. Handbooks of Applied Linguistics 8. In this issue. Berlin/Boston: Mouton de Gruyter.
Hodge, Karl
2000 “It’s all in the memes. Is the internet spreading a virus through our heads?” In: *guardian.co.uk*.
→ http://www.guardian.co.uk/science/2000/aug/10/technology [Last accessed 09/08/2011].
Hodge, Matthew J.
2006 “The Fourth Amendment and Privacy Issues on the “New” Internet: Facebook.com and MySpace.com”. In: *Southern Illinois University Law Journal* 31, pp. 95–123.
Internet Mail Consortium
1996 *vCard – The Electronic Business Card Version 2.1*. Specification.
→ http://www.imc.org/pdi/vcard-21.txt [Last accessed 09/08/2011].
Jabber Software Foundation
2004 *Extensible Messaging and Presence Protocol (XMPP): Core*. Specification.
→ http://tools.ietf.org/html/rfc3920 [Last accessed 09/08/2011].
Jackson, Matthew O.
2008 *Social and Economic Networks*. Princeton, New Jersey: Princeton University Press.
James, Michael L., C. Edward Wotring, and Edward J. Forrest
1995 “An Exploratory Study of the Perceived Benefits of Electronic Bulletin Board Use and Their Impact on Other Communication Activities”. In: *Journal of Broadcasting & Electronic Media* 39.1, pp. 30–50.
Johnson, Robert
2010 *More Details on Today’s Outage*. Official Facebook Blog Post.
→ http://www.facebook.com/note.php?note_id=431441338919&id=9445547199 [Last accessed 09/08/2011].
Joinson, Adam N.
2008 “‘Looking at’, ‘Looking up’ or ‘Keeping up with’ People? Motives and Uses of Facebook”. In: *Conference on Human Factors and Computing Systems (CHI 2008)*. ACM. Florence, Italy.
Jones, Paul E., Gonzalo Salgueiro, and Joseph Smarr
2011 *Webfinger (Draft)*. Specification.
→ http://datatracker.ietf.org/doc/draft-jones-appsawg-webfinger/ [Last accessed 04/05/2012].
Kempe, David, Jon Kleinberg, and Éva Tardos
2003 “Maximizing the Spread of Influence through a Social Network”. In: *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*. KDD ’03. New York, NY, USA, pp. 137–146. Washington, D.C., USA.
Kickstarter
2010 *Decentralize the web with Diaspora. An Open Software project in New York, NY by Daniel G. Maxwell S. Raphael S. Ilya Z.*. Website.

→ http://www.kickstarter.com/projects/196017994/diaspora-the-personally-controlled-do-it-all-distr [Last accessed 09/08/2011].

Killworth, Peter D., Eugene C. Johnsen, H. Russel Bernard, Gene Ann Shelley, and Christopher McCarty
1990 "Estimating the Size of Personal Networks". In: *Social Networks* 12, pp. 289–312.

Kivran-Swaine, Funda, Priya Govindan, and Mor Naaman
2011 "The Impact of Network Structure on Breaking Ties in Online Social Networks: Unfollowing on Twitter". In: *Proceedings of the 2011 annual Conference on Human Factors in Computing Systems.* New York, NY, USA: ACM, pp. 1101–1104. Vancouver, Canada.

Kleinfeld, Judith
2002 "Six Degrees of Separation: Urban Myth?" In: *Psychology Today* 35.2, p. 74.

Kneidinger, Bernadette
2010 *Facebook und Co. Eine soziologische Analyse von Interaktionsformen in Online Social Networks*. In German, Wiesbaden, Germany: VS Verlag für Sozialwissenschaften.

Krotoski, Aleks
2010 *Robin Dunbar: we can only ever have 150 friends at most …* Interview with Aleks Krotoski and Robin Dunbar.
→ http://www.guardian.co.uk/technology/2010/mar/14/my-bright-idea-robin-dunbar [Last accessed 09/08/2011].

Kubina, Petra and Andy Lücking
2012 "Barrier-free Communication". In: *Handbook of Technical Communication*. Ed. by Alexander Mehler, Laurent Romary, and Dafydd Gibbon. Handbooks of Applied Linguistics 8. In this issue. Berlin/Boston: Mouton de Gruyter.

Kwak, Haewoon, Hyunwoo Chun, and Sue Moon
2011 "Fragile Online Relationship: A First Look at Unfollow Dynamics in Twitter". In: *Proceedings of the 2011 annual Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, pp. 1091–1100. Vancouver, Canada.

Lampe, Cliff, Nicole B. Ellison, and Charles Steinfield
2006 "A Face(book) in the Crowd: Social Searching vs. Social Browsing". In: *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*. New York, NY, USA: ACM, pp. 167–170. Banff, Alberta, Canada.

Lazarsfeld, Paul F. and Robert K. Merton
1982 "Friendship as Social Process: A Substantive and Methodological Analysis". In: *The Varied Sociology of Paul F. Lazarsfeld*. Ed. by Patricia L. Kendall. Reprinted from Morroe Berger, Theodore Abel, and Charles Page, eds., Freedom and Control in Modern Society, 1954. New York: Columbia University Press, pp. 298–348.

LeBlanc, Jonathan
2011 *Programming Social Applications*. O'Reilly Media.

Lenhart, Amanda B.
2005 “Unstable Texts: An Ethnographic Look at How Bloggers and Their Audience Negotiate Self-Presentation, Authenticity and Norm Formation”. MA thesis. Graduate School of Arts and Sciences of Georgetown University.

Leskovec, Jure, Lada A. Adamic, and Bernardo A. Huberman
2006 “The Dynamics of Viral Marketing”. In: *Proceedings of the 7th ACM conference on Electronic Commerce*. Vol. 1. EC ’06. New York, NY, USA: ACM, pp. 228–237. Xi’an, China.

Leskovec, Jure and Eric Horvitz
2008 “Planetary-Scale Views on a Large Instant-Messaging Network”. In: *Proceedings of WWW 2008*. Beijing, China.

Levmore, Saul and Martha C. Nussbaum (eds.)
2011 *The Offensive Internet: Privacy, Speech, and Reputation*. Harvard University Press.

Li, Charlene
2008 *The future of social networks: Social networks will be like air*. Blog Post. A Forrester Research Blog.
→ http://forrester.typepad.com/groundswell/2008/03/the-future-of-s.html [Last accessed 09/08/2011].

Lucas, Matt
2007 *Decentralizing Digital Social Networking Applications*.
→ https://netfiles.uiuc.edu/mmlucas2/www/decentralizingDigitalSocialNetworking.pdf [Last accessed 09/08/2011].

Lugano, Giuseppe
2007 “Mobile Social Software: Definition, Scope and Applications”. In: *eChallenges 2007*, pp. 1434–1441. The Hague, Netherlands.

Maheswaran, Muthucumaru, Bader Ali, Hatice Ozguven, and Julien Lord
2010 “Online Identities and Social Networking”. In: *Handbook of Social Network Technologies and Applications*. Ed. by Borko Furht. Berlin: Springer Science+Business Media. Chap. 11, pp. 241–267.

Manago, Adriana M., Michael B. Graham, Patricia M. Greenfield, and Goldie Salimkhan
2008 “Self-Presentation and gender on MySpace”. In: *Journal of Applied Developmental Psychology* 29.6, pp. 446–458.

Marlow, Cameron
2009 *Maintained Relationships on Facebook*. Blog Post.
→ http://overstated.net/2009/03/09/maintained-relationships-on-facebook [Last accessed 09/08/2011].

Martin, Dave
2010 *Nielsen NetView (June 2009 – June 2010)*. Cited in: *What Americans Do Online: Social Media And Games Dominate Activity.*
→ http://blog.nielsen.com/nielsenwire/online_mobile/what-americans-do-online-social-media-and-games-dominate-activity/ [Last accessed 09/08/2011].

McFedries, Paul
2007 “Technically Speaking: All A-Twitter”. In: *Spectrum* 44.10, p. 84.

McKenna, Katelyn Y. A. and John A. Bargh
2000 "Plan 9 From Cyberspace: The Implications of the Internet for Personality and Social Psychology". In: *Personality and Social Psychology Review* 4.1, pp. 57–75.
McPherson, Miller, Lynn Smith-Lovin, and James M. Cook
2001 "Birds of a Feather: Homophily in Social Networks". In: *Annual Review of Sociology* 27, pp. 415–444.
Mehler, Alexander and Tilmann Sutter
2008 "Interaktive Textproduktion in Wiki-basierten Kommunikationssystemen". In: *Kommunikation, Partizipation und Wirkungen im Social Web*. Ed. by Ansgar Zerfaß, Martin Welker, and Jan Schmidt. In German, pp. 267–298.
Mika, Peter
2007 *Social Networks and the Semantic Web*. Semantic Web and Beyond: Computing for Human Experience. New York, NY, USA: Springer Science+ Business Media.
Milgram, Stanley
1967 "The Small World Problem". In: *Psychology Today* 1, pp. 61–67.
Mitchell-Wong, Juliana, Ryszard Kowalczyk, Albena Roshelova, Bruce Joy, and Henry Tsai
2007 "OpenSocial: From Social Networks to Social Ecosystem". In: *Digital EcoSystems and Technologies Conference, 2007. DEST '07. Inaugural IEEE-IE*. IEEE Computer Society, pp. 361–366. Cairnes, Australia.
Newey, Guy
2009 "Blogging guru chips away at Great Firewall of China". *AFP*.
→ http://www.google.com/hostednews/afp/article/ALeqM5igVn4hcj6ZNWlawSvzLvgEMkZmkQ [Last accessed 09/08/2011].
Nottingham, Mark and Eran Hammer-Lahav
2010 *Defining Well-Known Uniform Resource Identifiers (URIs)*. Specification.
→ http://tools.ietf.org/html/rfc5785 [Last accessed 09/08/2011].
OpenSocial and Gadgets Specification Group
2010 *OpenSocial Specification 2.0 DRAFT*. Specification.
→ http://docs.opensocial.org/display/OSD/Specs [Last accessed 09/08/2011].
Panzer, John
2010 *The Salmon Protocol*. Specification.
→ http://salmon-protocol.googlecode.com/svn/trunk/draft-panzer-salmon-00.html [Last accessed 09/08/2011].
Panzer, John, Ben Laurie, and Dirk Balfanz
2011 *Magic Signatures*. Specification.
→ http://salmon-protocol.googlecode.com/svn/trunk/draft-panzer-magicsig-01.html [Last accessed 09/08/2011].
Pariser, Eli
2011 *The Filter Bubble: What the Internet is Hiding from You*. Penguin Books.
Parks, Malcolm R.
2007 *Personal Relationships & Personal Networks*. Series on Personal Relationships. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

Pfanner, Eric
2011 "Naming Names on the Internet". In: *The International Herald Tribune*. 5 September 2011.
Prodromou, Evan
2010 *What is the federated social web?* Blog Post.
→ http://status.net/2010/07/13/what-is-the-federated-social-web [Last accessed 09/08/2011].
Prodromou, Evan, Brion Vibber, James Walker, and Zach Copley
2010 *OStatus 1.0 Draft 2*. Specification.
→ http://ostatus.org/sites/default/files/ostatus-1.0-draft-2-specification.html [Last accessed 09/08/2011].
Recordon, David
2007 *Is Zuckerberg Trying to Own the Social Graph?* Blog Post.
→ http://daveman692.livejournal.com/313178.html [Last accessed 09/08/2011].
Riley, Duncan
2007 *TC40 Keynote Conversation: Mark Zuckerberg*. Blog Post. Interview with Michael Arrington and Mark Zuckerberg.
→ http://techcrunch.com/2007/09/17/tc40-keynote-conversation-mark-zuckerberg/ [Last accessed 09/08/2011].
Riley, Shannon
2006 "Password Security: What Users Know and What They Actually Do". In: *Usability News* 8.1.
Rogers, Everett M.
2003 *Diffusion of Innovations*. 5th ed. New York: The Free Press.
Salzberg, Maxwell, Daniel Grippi, Ilya Zhitomirskiy, Sarah Mei, Yosem Companys, and Peter Schurman
2011 *Diaspora* is making a difference*. Blog Post.
→ http://blog.joindiaspora.com/2011/09/08/we-are-making-a-difference.html [Last accessed 09/15/2011].
Schneider, Roman
2008 "Web 3.0 ante portas? Integration von Social Web und Semantic Web". In: *Kommunikation, Partizipation und Wirkungen im Social Web*. Ed. by Ansgar Zerfaß, Martin Welker, and Jan Schmidt. In German, pp. 112–128.
Schonfeld, Erick
2011 *The Pros And Cons Of Facebook Comments*. Blog Post.
→ http://techcrunch.com/2011/03/01/pros-cons-facebook-comments/ [Last accessed 09/08/2011].
Singel, Ryan
2009 *Facebooks E-mail Censorship is Legally Dubious, Experts Say.* Blog Post.
→ http://www.wired.com/epicenter/2009/05/facebooks-e-mail-censorship-is-legally-dubious-experts-say/ [Last accessed 09/08/2011].
Smarr, Joseph
2008 *Portable Contacts 1.0 Draft C*. Specification.
→ http://portablecontacts.net/draft-spec.html [Last accessed 09/08/2011].
Snell, James M.
2006 *Atom Threading Extensions*. Specification.
→ http://tools.ietf.org/html/rfc4685 [Last accessed 09/08/2011].

Status.Net
2011 *OStatus/Interop*. Website/Wiki.
→ http://status.net/wiki/OStatus/Interop [Last accessed 09/15/2011].

Stoll, Clifford
1995 *Silicon Snake Oil. Second Thoughts on the Information Highway*. New York, NY, USA: Anchor Books, Doubleday.

Stone, Brad and Noam Cohen
2009 "Social Networks Spread Defiance Online". In: *The New York Times*. 16 June 2009.

Story, Henry, Bruno Harbulot, Ian Jacobi, and Mike Jones
2009 "FOAF+SSL: RESTful Authentication for the Social Web". In: *Proceedings of the ESWC2009 Workshop on Trust and Privacy on the Social and Semantic Web (SPOT2009)*. Heraklion, Greece.

The Nielsen Company
2009 *Global Advertising: Consumers Trust Real Friends and Virtual Strangers the Most*. Nielsen Global Online Consumer Survey.
→ http://blog.nielsen.com/nielsenwire/consumer/global-advertising-consumers-trust-real-friends-and-virtual-strangers-the-most/ [Last accessed 09/08/2011].

The OpenID Foundation
2007 *OpenID Authentication 2.0 – Final*. Specification. This refers to Version 2.0.
→ http://openid.net/specs/openid-authentication-2_0.html [Last accessed 09/08/2011].

Thimm, Caja
2008 "Technically-mediated interpersonal communication". In: *Handbook of Interpersonal Communication*. Ed. by Gerd Antos and Eija Ventola. Handbooks of Applied Linguistics 2. Berlin, Germany: Mouton de Gruyter. Chap. 12, pp. 331–354.

Tong, Stephanie Tom, Brandon Van der Heide, Lindsey Langwell, and Joseph B. Walther
2008 "Too Much of a Good Thing? The Relationship Between Number of Friends and Interpersonal Impressions on Facebook". In: *Journal of Computer-Mediated Communication* 13, pp. 531–549.

Tramp, Sebastian, Philipp Frischmuth, Natanael Arndt, and Sören Auer
2011 "Weaving a Distributed, Semantic Social Network for Mobile Users". Submitted to the Extended Semantic Web Conference (ESWC2011).

Travers, Jeffrey and Stanley Milgram
1969 "An Experimental Study of the Small World Problem". In: *Sociometry* 32.4, pp. 425–443.

Van Buskirk, Eliot
2009 "Denial-of-Service Attack Knocks Twitter Offline". In: *Wired.com*.
→ http://www.wired.com/epicenter/2009/08/twitter-apparently-down/ [Last accessed 09/08/2011].

Vu, Quang Hieu, Mihai Lupu, and Beng Chin Ooi
2010 "Peer-to-Peer Computing. Principles and Applications". In: *Architecture of Peer-to-Peer Systems*. Springer. Chap. 2, pp. 11–37.

W3C Social Web Incubator Group
2010 *A Standards-based, Open and Privacy-aware Social Web. W3C Incubator Group Report 6th December 2010*. Tech. rep. W3C.
→ http://www.w3.org/2005/Incubator/socialweb/XGR-socialweb-20101206/ [Last accessed 09/08/2011].

Walther, Joseph B., Brandon Van der Heide, Sang-Yeon Kim, David Westerman, and Stephanie Tom Tong
2008 "The Role of Friends' Appearance and Behavior on Evaluations of Individuals on Facebook: Are We Known by the Company We Keep?" In: *Human Communication Research* 34, pp. 28–49.

Waltinger, Ulli and Alexa Breuing
2012 "Internet-Based Communication". In: *Handbook of Technical Communication*. Ed. by Alexander Mehler, Laurent Romary, and Dafydd Gibbon. Handbooks of Applied Linguistics 8. In this issue. Berlin/Boston: Mouton de Gruyter.

Wasserman, Stanley and Katherine Faust
2008 *Social Network Analysis: Methods and Applications*. 17th printing. Cambridge University Press.

Watts, Duncan J. and Steven H. Strogatz
1998 "Collective dynamics of 'small-world' networks". In: *Nature* 393, pp. 440–442.

Wellmann, Barry and Milena Gulia
1999 "Net-Surfers Don't Ride Alone: Virtual Communities as Communities". In: *Networks in the Global Village. Life in Contemporary Communities*. Ed. by Barry Wellmann. Westview Press, pp. 331–366.

Wellmann, Barry and Stephanie Potter
1999 "The Elements of Personal Communities". In: *Networks in the Global Village. Life in Contemporary Communities*. Ed. by Barry Wellmann. Westview Press, pp. 49–81.

Werry, Christopher C.
1996 "Linguistic and Interactional Features of Internet Relay Chat". In: *Computer-Mediated Communication: Linguistic, Social and Cross-Cultural Perspectives*. Ed. by Susan C. Herring. Pragmatics and Beyond. Amsterdam, Netherlands: John Benjamins, pp. 47–63.

Winer, Dave
2009a *Implementor's guide to rssCloud*. Specification.
→ http://rsscloud.org/walkthrough.html [Last accessed 09/08/2011].
2009b *RSS 2.0 Specification*. Specification.
→ http://www.rssboard.org/rss-specification [Last accessed 09/08/2011].

Wohn, Donghee Yvette, Cliff Lampe, Rick Wash, Nicole Ellison, and Jessica Vitak
2011 "The 'S' in Social Network Games: Initiating, Maintaining, and Enhancing Relationships". In: *Proceedings of the 44th Hawaii International Conference on System Sciences*. Kauai, HI, USA.

Ziv, Nina D. and Bala Mulloth
2006 "An Exploration on Mobile Social Networking: Dodgeball as a Case in Point". In: *Proceedings of the Mobile Business Conference 2006*. IEEE. Copenhagen, Denmark.

15. Digital Curation as Communication Mediation

Christopher A. Lee

1. Introduction

Communication is a highly interactive and dynamic process. Parties to the process engage in continual efforts to revise their understanding, knowledge and assumptions.[1] Communication across space and time is often mediated through information artifacts. A journalist, for example, can communicate – through what is often called "mass communication" (Peterson 2003) – with her readers through the articles that she publishes in a newspaper. Those who read the articles can gain insights about both the subjects being discussed and the journalist herself. If the newspaper is preserved, this type of communication can occur over very long periods of time.

Humans have a long tradition of retaining information artifacts for future use. Collecting institutions – libraries, archives and museums – manage extensive collections of materials that can communicate events, insights, facts and perspectives to those who encounter them. Individuals, families, corporations, and various other types of organizations also generate and retain information artifacts across time. These activities can be seen as communication mediation.

As all sectors of contemporary societies have increasingly adopted digital technologies in order to carry out their activities, the communication mediation processes have dramatically changed. Supporting the meaningful use of digital objects over time – a set of activities that has recently come to be called "digital curation" – requires an understanding and appreciation of the various layers of representation through which meaning can be conveyed in digital systems. This chapter discusses the characteristics of those layers and strategies for ensuring perpetuation of meaning across time.

2. Digital traces as intentional and unintentional communication

Communication can be characterized as the conveyance of information from one party to one or more other parties (potentially including conveyance to him/herself). Information is "a detectable pattern on which action can be conditioned" (Cohen and Axelrod 1998: 38). Humans value information because they find it meaningful. Meaning never comes for free. Its value comes through its enactment in specific situations – and no two situations are exactly the same (Barwise and Perry 1983).

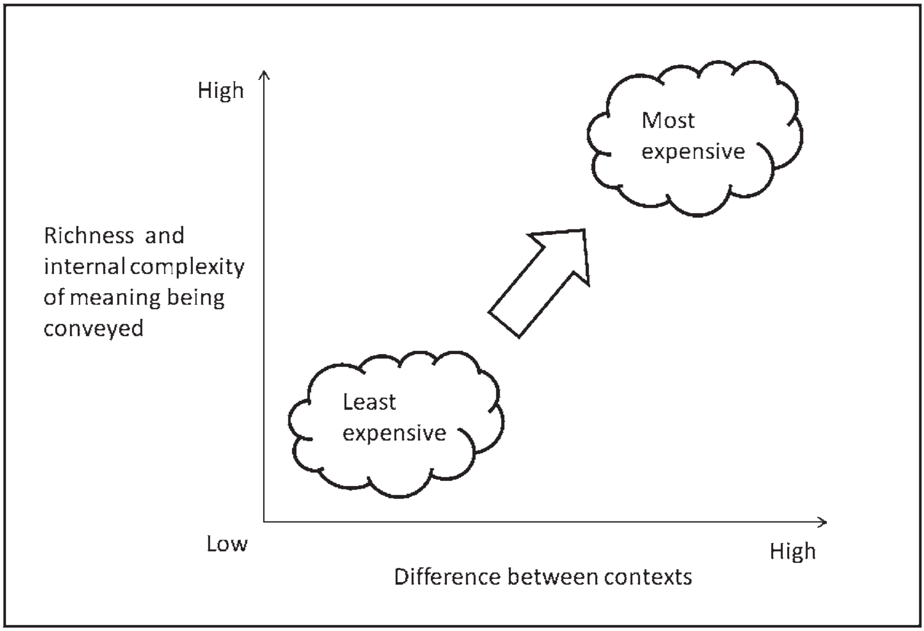


Figure 1. Factors influencing expense of conveying meaning across contexts.

Conveying meaning from one party to another – whether they are engaged in a real-time conversation or one is reading a document that another wrote hundreds of years ago – involves a "fusion of horizons" (Gadamer 1989). This requires some degree of "common ground" between them (Stalnaker 1978; Clark 1996).[2] For Alice to grasp a concept offered by Bob, Alice must devote effort and attention to the task (and Bob may also). For example, if Alice and Bob are sharing their views about a television show that they both saw the night before, they will exchange a variety of impressions and background assumptions – the process of sharing meaning is not complete just because they know that they both witnessed the same episode. As a more extreme example, developing a shared sense of political identity can take generations of interaction, deliberation, negotiation and shared experiences. Two factors that influence how difficult it is to convey meaning are: (1) the degree of richness and complexity of the meaning and (2) the degree of difference between the contexts in which the meaning is being enacted. Figure 1 provides a representation of these relationships.

Conveyance and enactment of meaning is not simply a matter of reliably transferring a discrete signal from one place to another. All entities involved in the process – conveyor, recipient and meaning of the information – are subject to transformation. If, for example, Jack knows a story, he can tell it to Nora. As a result, Jack, Nora and the meaning of the story may all be different from what they were before. If Nora then tells the story to Henry, a similar transformation can occur.

While there are some forms and aspects of communication that are completely ephemeral – such as unrecorded facial expressions or hand gestures – a great deal of human communication involves the generation of traces that have varying levels of persistence.[3] Human activities leave numerous traces, and all of the traces are potential conveyers of information. A door left open can signal that it is socially acceptable to enter the room. Footprints on a beach provide evidence of previous visitors and where they walked. Receipts convey information about transactions and the contexts in which they took place. Photographs of public officials shaking each other's hands indicate both that the two individuals met and that someone found the encounter to be noteworthy enough to warrant capture. The contents of a web browser cache reveal actions and behavior patterns of the browser's users. Extensive information about previous actions on a computer is also stored by the computer's operating system. The packets used to transfer data across the Internet include headers, which reflect the data's previous and intended path.

A great deal of the literature – in linguistics, communication and philosophy of language – related to processes for conveying meaning has focused on direct human discourse as the context of interaction. This has included definition of a speech act as a specific communicative event, whose meaning and appropriateness can depend significantly on the circumstances in which it takes place (Austin 1962). Speech acts that are carried out in face-to-face verbal conversation generally do not leave persistent documentary traces as evidence that they occurred. Many authors have attempted to formally represent the context of discourse in terms of the "body of information that is presumed to be shared by the participants in the conversation" (Stalnaker 1998; see also Bouquet, Serafini, and Thomason 2008). The given-and-take of direct human interaction is facilitated not only by shared knowledge and assumptions, but also by shared purposes. "Our talk exchanges do not normally consist of a succession of disconnected remarks, and would not be rational if they did. They are characteristically, to some degree at least, cooperative efforts; and each participant recognizes in them, to some extent, a common purpose or set of purposes, or at least a mutually accepted direction" (Grice 1975: 26). One can even characterize communication as a series of responses to "coordination problems" (Lewis 1969).

When conducted through fixed symbolic forms – such as electronic mail, telephone, video – communicative acts can yield persistent objects. The use of computer systems always involves the generation, copying and manipulation of strings of explicit symbolic representations; those representations are then potentially available for use in later interactions: a phenomenon that Shoshana Zuboff (1988) has called "informating." It is this "informating" of interactions that generates an unprecedented amount of explicit information; and facilitating long-term use of the information is the mission of digital curation.

One can intentionally generate traces or instead leave traces as unintended side effects of actions. Sara may put on a work uniform, for example, purposely to let others know that she is "on duty." If the uniform is looking worn and threadbare, that might also reveal that she has worked on the same job for many other days in the past, but that is not because she purposely distressed the fabric to convey such information. Likewise, composing and sending an email message is a deliberate act, but the creator is not usually making an intentional choice to also generate system log entries or temporary files that are associated with the message. The level of investment and conscious commitment to traces can fall along a spectrum from completely unreflective/incidental to highly crafted and constructed. When there are special moments in life, one often makes an effort to create much richer traces of them. For example, someone might take frequent low-resolution pictures of her everyday life using a cheap camera on her phone, but then arrange for a consciously orchestrated, well-lit and high-resolution photograph of her wedding. If the context in which the traces are accessed and used is significantly different from the context in which they were generated, then the interaction cannot rely on the same sorts of informal cues for understanding that would be available in a direct conversation.

Information, like money, is often given without the giver's knowing to just what use the recipient will want to put it. If someone to whom a transaction is mentioned gives it further consideration, he is likely to find himself wanting the answers to further questions that the speaker may not be able to identify in advance; if the appropriate specification will be likely to enable the hearer to answer a considerable variety of such questions for himself, then there is a presumption that the speaker should include it in his remark; if not, then there is no such presumption (Grice 1975).

The curation of information artifacts is fundamentally different from direct communication in that one cannot assume that the parties who generated the traces will be available to provide "answers to further questions." Instead, one must make educated guesses about likely uses of the traces and then pre-emptively respond to likely questions by embedding appropriate information in the mechanisms used to manage and provide access to the traces (e.g. repositories, collection descriptions, information packages).

There are three powerful strategies that one can adopt when his/her goal is to convey traces across contexts:

1. First, one can use fixity measures to increase the chances that the traces will persist across space and time. For example, low-acid paper will not deteriorate as quickly as paper with high-acid content.

2. Second, one can generate objects that are purposely self-describing. A book, for example, can include its own glossary of terms, and a business memorandum has elements to convey its context of creation such as to, from, subject and date.
3. Finally, one can embed traces into larger systems of traces that provide meaning to each other. For example, storing a whole series of correspondence together in the same folder will help to convey meaning that could be lost if letters were all separated and stored in different places.

I will discuss each of these strategies in Section 3 on digital curation as communication mediation.

In most of the world, human actions leave numerous digital traces. Curation of selected sets of those traces – by creators, users, information professionals and others – can enable numerous forms of inquiry, discovery and communication. Proper digital curation requires an understanding of the ways in which digital traces are both similar to and dramatically different from traces left in analog forms.

3. The nature of digital traces

Any use of digital resources is highly mediated. It involves the interaction of various hardware and software components. By definition, digital information is stored and processed as a series of discrete values. The values are represented through properties of physical media – usually through charged magnetic particles or tiny holes in disks. Hardware and low-level software detects the physical properties and interprets them as binary digits – i.e., digits that can take only one of two possible values – which are called *bits*. By convention, we say that the two possible binary values are 1 and 0 (Shannon 1948).

It is often useful to be able to identify and consistently reproduce a specific series of ones and zeroes, which is called a *bitstream*. The bitstream is a powerful abstraction layer, because it allows any two computer components to reliably exchange data, even if the underlying structure of the components is quite different (e.g. arrangement of sectors on a hard drive or pits and lands on an optical disk). In other words, even though the bits that make up the bitstream must be manifested through physical properties of computer hardware, the bitstreams are not inextricably tied to any specific physical manifestation. Another useful characteristic of bitstreams is that they can be reproduced with complete accuracy. By using well-established mechanisms – such as generation and comparison of SHA1 hash values[4] – one can verify that two different instances of a bitstream are exactly the same. This is more fundamental than simply saying that one has made a good copy. If the two hash values are identical, then the two instances are, by definition, the same bitstream.[5]

Humans are sometimes directly interested in bitstreams (e.g. when conducting detailed forensic analysis, security audits or debugging), but they are much more likely to care about the meaning that emerges from the dynamic interaction of hardware and software components that access and process numerous bitstreams.

3.1. Interacting components and interoperability

Users experience digital objects through human-computer interfaces. They handle and read documents that have been printed to paper or view and interact with representations of information on computer screens or hear it through computer speakers. In some cases, a user experiences a digital object as a relatively discrete entity, as when she reads an electronic journal article. In other cases, interaction with distinct digital objects is less direct, such as when a researcher is conducting an analysis through a system that pulls data from a variety of different sources. In the former case, the visual and/or auditory rendering of the digital object is likely to be important to the user. In the latter case, the user is likely to be more concerned with the accuracy and efficiency of access to and manipulation of the data contained in the digital objects (for more information on evaluation in technical communication see Menke 2012 in this volume).

Regardless of whether one is interacting with digital objects as distinct entities or interacting with representations that aggregate various sources, the process is highly mediated by hardware and software. In other words, any particular encounter with a digital object must occur within a specific technological environment. A specific stack of hardware and software that can be used to reproduce the digital object can be defined as one of the object's "view paths" (van Diessen 2002).[6] For example, a *Portable Document Format* (PDF) file could be read using the Adobe Acrobat Reader within the Windows 7 operating system running an Intel Core processor; or it could be read using Okular in combination with Poppler within the Ubuntu distribution of the Linux operating system running an AMD64 processor (on document formats see also Rahtz 2012 in this volume). The same hardware or software component can be shared by many potential view paths. For example, one could run the same Java application in either a Windows or Macintosh environment, as long as both systems are running a Java virtual machine.

For consumers, it is desirable to have numerous options for computer platforms that can be used to access any given digital object. If there are very few types of technical environments in which to use the digital object (i.e. a limited number of available view paths), then one can become overly dependent on the providers of those particular environments. Two major risks are (1) lock-in to specific vendors, which can require consumers to pay unreasonable rents to the vendors over time, or (2) vendors ceasing to provide support for the technology at all because they have gone out of business or changed their service/product offerings.

The flip side to dependency is interoperability. From an engineering perspective, interoperability is the ability for two or more systems or functional units to "exchange information and to use the information that has been exchanged" (IEEE 1990) or "communicate, execute programs, or transfer data … in a manner that requires the user to have little or no knowledge of the unique characteristics of those units" (ISO 1993). Interoperability can greatly facilitate coordination and communication across systems. It is a concept that can be applied at many different levels of granularity, from the ability to connect two physical devices to the sharing of concepts and work processes across organizational boundaries (Tolk 2003). Two factors that have made interoperability increasingly important over the past several decades are (1) "distributed computing infrastructures" based on networked access to resources and (2) "increasing specialization of work, but increasing need to reuse and analyze data" (Sheth 1999) (for more information on distributed computing infrastructures see Heyer, Holz and Teresniak 2012 in this volume).

Early work on computers tended to focus on specific combinations of hardware and software that were configured to carry out tasks. However, by the 1960s, the computer science literature reflected various approaches to promote "machine independence" (Halpern 1965). This included, for example, computer-supported translation between low-level languages, in order to mitigate the effects of platform dependency (Gaines 1965; Wilson and Moss 1965). Computer scientists and engineers increasingly recognized the desirability of being able to use data in a diversity of computer environments, and they have developed numerous methods and mechanisms for doing so, including virtualization (Parmelee et al 1972), modularity (Baldwin and Clark 2000; Langlois and Robertson 1992), and standardization (Cargill 1997).

The initiative that most decisively introduced interoperability across generations of hardware into the computer industry was the development of System 360 by *International Business Machines* (IBM) in the 1960s. One of the most widely recognized innovations of the System 360 architecture was that it allowed IBM to release and support an entire family of computers – targeted at different segments of the market – that all interoperated with each other. However, System 360 also included code that could emulate hardware and thus supported software and data files created on earlier IBM hardware. Many software and hardware vendors have followed IBM's lead, building backward compatibility into their products, in order to perpetuate use of their line of products while also supporting an easy transition to their latest offerings. At the same time that producers were applying and refining the backward compatibility concept, expertise began to develop on the user side about how to address the problems associated with data residing on legacy systems within organizations.

Starting around the 1980s, there has been an increasing emphasis in the computer industry on portable and reusable code (Krueger 1992). For some types of software, developers are often willing to pay the price of increased development time or some degradation of performance in order to increase the chances that their software can be used on a wide variety of hardware platforms. A subset of system and database administrators – particularly those working for large organizations that have maintained collections of digital objects over many years – have also developed expertise in the recovery of both data and functionality from legacy systems of all sizes.

In order to address system integration, interoperability and data exchange, computer professionals frequently describe hardware and software as a stack of independent but interacting *layers* (Messerschmitt and Szyperski 2003). For example, systems for administering interactive web sites are often characterized in terms of layers (or tiers) dedicated to specific functions such as data storage, business logic, query processing, and user interface. Perhaps the most widely recognized example of such layering is the *Open Systems Interconnection* (OSI) Reference Model for network data transfer (ISO 1994). Although the OSI Model has been widely characterized as an approach that "lost" the standards war against competing ARPANET protocols, the OSI had a significant and lasting influence on how members of the computer industry characterized and conceived of networks. While the definitions of the specific OSI layers are outside the scope of this paper, the intended purpose of the stack of layers is very relevant to the discussion.

Figure 2 presents what I contend to be the implied communication dynamics of the OSI Model. Alice is sitting at a computer in Los Angeles and Bob is sitting at a computer in New York. Alice sends an email message to Bob. Because they are not sitting at the same computer, Alice cannot simply copy the file from one directory to another. Instead she must push the data down through several layers, starting at the level of the application that she is using and ending down at the level where physical movement of the data can occur. Bob's system receives the data through the layer at the bottom of his stack, and the data is then pushed up through the set of layers that ends at the application that he is using to view the email. If Alice and Bob are now both looking at the message on their screens at the same time, then we can say that they are sharing meaning based on the contents of the message. However, they are never directly moving data across the top part of the figure – i.e. the place that they care about because the meaning resides there. All interactions are mediated through a set of layers designed to support interchange across diverse and remote systems. And those applying this model will generally assume (for good reason) that users of the network do not want to see or think about data passing through any layer but the one at the very top.[7]

A result of portability of data across systems is the range of experiences that one can have when encountering the same digital object, depending upon the

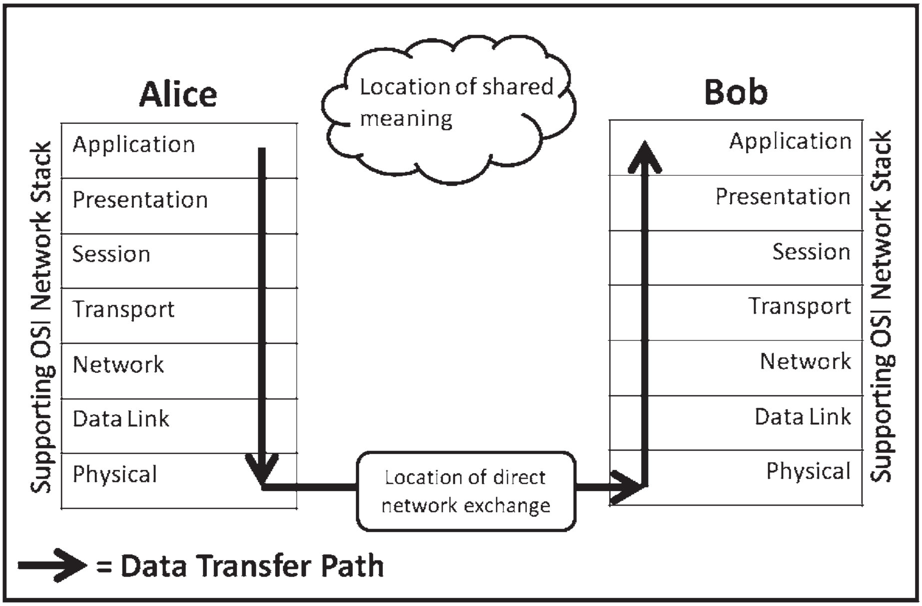


Figure 2. Implied communication dynamics of *Open Systems Interconnection* (OSI) network model.

characteristics and state of the computing environment being used to encounter it. This can be the result of fixed characteristics of the technology being used (e.g. screen size) or configuration settings. For example, when Alice and Bob view the "same" email message, they may see different fonts, lines breaking at different points, exposure of different elements of the email header, and different ways of representing embedded hyperlinks and text copied from other messages. This example demonstrates that many properties of digital objects are best understood as a range of potential values, rather than specific discrete values.

3.2. Levels of representation

Digital resources are composed of interacting components that can be considered and accessed at different levels of representation. Access to data from a storage device normally involves mounting a volume and then copying or opening files through the filesystem. There must be hardware to detect signals on the medium, hardware and software to translate the signals into bitstreams, and hardware and software to move the bitstreams into the current working computer environment. One can then interact with data as entire files or components of files. The filesystem usually plays a mediating role between the user and the underlying data, and it is designed to facilitate interaction at the file level (e.g. file naming, viewing timestamps, access controls). It serves to “hide” complicated information from the user about “where and how it stores information” (Farmer and Venema 2005). Those who are interested in the underlying data that is hidden by the filesystem can instead generate and interact with disk images, which are low-level, sector-by-sector copies of all the data that resides on the storage medium. Inspection of the disk image can reveal a significant amount of information that users of the drive did not consciously or intentionally leave there (Garfinkel and Shelat 2003). However, for most purposes, the filesystem is a very valuable abstraction mechanism, because it does not require users to understand or directly access the underlying data. Users can encounter the contents of a file as a bitstream by using a hex editor,[8] but they are more likely to render its contents within a particular application. At even higher levels of representation, one often encounters digital objects not as distinct files but as either discrete objects composed of multiple files or aggregations of such objects.

See Table 1 for a summary of several levels. One could alternatively propose a smaller set of levels that each span a larger set of properties and activities – e.g., stored data, software-readable information, and user experience – or a larger set of levels that reflect much finer distinctions – e.g., bitstreams with or without checkbits used for error correction, the work-expression-manifestation-item distinctions of the *Functional Requirements for Bibliographic Records* (FRBR), Panofsky’s natural subject, conventional and intrinsic meaning levels (Panofsky 1955). However, I have defined the levels at a degree of granularity that I believe reflects specific and important implications for digital curation as communication mediation.

The properties of information at a given level of representation are directly based upon, but are not fully reducible to, properties of information at the level immediately below it. Each level has emergent properties, which convey potential meaning that is not available through any of the other layers. This is because moving between layers always involves a process of translation that both adds and removes information.

The rendering of a web page in a browser, for example, does not reflect any comments that are within the text of the HTML file. Viewing the HTML file through a text editor would reveal the comments (and other properties of the HTML markup such as whitespace and style sheet references), but it would not reveal specifically how the page is presented to users who visit the site through a browser. The developer of a web site who is trying to fix a “broken” page will routinely shift between these two levels of representation, in order to see specifically how the HTML code is expressed and how changes to the code affect the appearance and behavior of the page in one or more web browsers. A given HTML file can be rendered quite differently in two different browsers, because the software interprets the code differently.

Table 1. Levels of Representation.

Level	Label	Explanation	Interaction Examples
7	Aggregation of objects	Set of objects that form an aggregation that is meaningful encountered as an entity	Browsing the contents of an archival collection using a finding aid
6	Object or package	Object composed of multiple files, each of which could also be encountered as individual files	Viewing a web page that contains several files, including HTML, a style sheet and several images
5	In-application rendering	As rendered and encountered within a specific application	Using Microsoft Excel to view an .xls file, watching an online video by using a Flash viewer
4	File through a filesystem	Files encountered as a discrete set of items with associate paths and file names	Viewing contents of a folder using Windows Explorer, typing "ls" at the Unix command prompt to show the contents of a directory
3	File as "raw" bitstream	Bitstream encountered as a continuous series of binary values	Opening an individual file in a hex editor
2	Sub-file data structure	Discrete "chunk" of data that is part of a larger file	Extracting a tagged data element in an XML document (see Stührenberg 2012 in this volume) or value of a field in a relational database
1	Bitstream through I/O equipment	Series of 1s and 0s as accessed from the storage medium using input/output hardware and software (e.g. controllers, drivers, ports, connectors)	Mounting a hard drive and then generating a sector-by-sector image of the disk using Unix dd command
0	Bitstream on a physical medium	Physical properties of the storage medium that are interpreted as bitstreams at Level 1	Using a high-power microscope and camera to take a picture of the patterns of magnetic charges on the surface of a hard drive or pits and lands on an optical disk

Similarly, someone using a hex editor to view a disk image bitstream – a copy of the disk that reflects the contents of every storage sector rather than just the files – will be able to see many types of information that are not visible to someone who mounts the drive and views it through Windows Explorer. Examples of such information are contents of unallocated sectors, deleted files, and “hidden” system files. However, someone who only had the hex editor view of the disk image would not be able to experience the WIMP (windows, icons, mouse and pointer) interactions with folders, files and applications to which current users are so accustomed. Someone trying to fix a corrupted file or correct a bug in a viewing application is likely to shift back and forth between the hex editor representation and in-application rendering representation, in order to determine specifically how properties in the former are reflected in the latter and vice versa.

4. Digital curation as communication mediation

As described above, use of digital resources is a process that spans multiple levels of representation. This often involves a process that is similar to the one reflected in Figure 2, if one were to replace the OSI layers with the levels of representation that are elaborated in Table 1. Let us return to the example of Alice’s email. When using her email client, Alice is operating at the top of the stack. She is interacting with aggregates of objects, in the form of folders and email discussion threads. If she views an individual message, she can experience it as a coherent object, but it may be composed of multiple files. When she makes any changes to the contents of her email account, the changes are “pushed down” into lower levels of representation, saving files to the filesystem and data onto storage media. If Alice sends a message to Bob, his encounter with the message will require a “pushing up” from storage (on his hard drive or a remote server) to a filesystem, application and an object or aggregation-level view.

The exchange between Alice and Bob leaves numerous traces at multiple levels of representation. Further use of those traces does not require that one uses the exact same access mechanism that Bob used. Someone with an interest in Alice’s digital objects – whether that is Alice herself, a family member, coworker, auditor, journalist or scholar studying her work – could potentially interact with them at any of the levels (see Figure 3).[9] One could use or imitate an email client in order to view her email within an application environment that is similar to the one that she used. Alternatively, one could investigate email messages and attachments as individual files. Yet another approach would be to treat all of the bits on her hard drive (or the hard drive of a server that held her email) as one large bitstream and use search tools to find patterns of interest – such as email addresses and headers – within the stream.

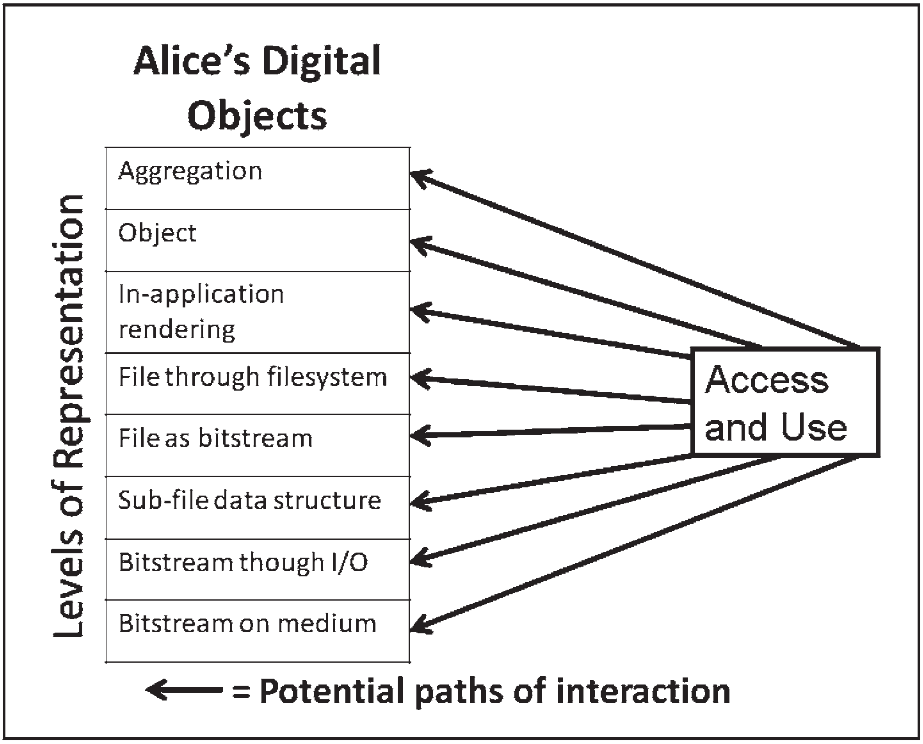


Figure 3. Deriving meaning from digital objects at multiple levels.

Digital curation is the set of activities required to ensure that digital objects can be meaningfully used over time (Lee and Tibbo 2011). Responsible care for digital traces is a highly shared and distributed endeavor. Important elements of this work are often carried out by information professionals, such as archivists, librarians, museum curators, records managers and data managers (Hedstrom and King 2006). There are also many other parties who can have an interest in and influence over the trajectory of digital traces. Those directly engaged in a communication exchange are obvious examples, because their interactions are being recorded, and they can often decide how, where and in what ways the traces of those interactions are retained. However, others can also have a major stake in how information is managed and conveyed, e.g. others who are discussed in the exchange or whose lives are significantly impacted by it (e.g. citizens of two countries whose leaders are hashing out foreign policy issues). In the following discussion, I will present several fundamental considerations for digital curation.

4.1. Reflecting context

A long-standing tenet of literature on communication is that context matters. One cannot determine the meaning or communicative effect of a message or utterance by analyzing it as a discrete entity in isolation (Dewey 1931; Rommetveit 1979). Characteristics of context are essential to the communication process. Context is inherently relational; it is always context of, about, or surrounding something, which I will call the *target entity* (TE).

There are three broad ways in which we can characterize the context of a TE (Lee 2011):

- *Context*$_1$: the set of symbolic expressions or representations that surround a TE and help one to express, make sense of, translate or otherwise act upon or within it.
- *Context*$_2$: objective or socially constructed characteristics and conditions of the situation in which a TE is, appears or occurs.
- *Context*$_3$: aspects of the mental or physical state, disposition, intentions, identity or recent experiences of an actor that bear upon how she interprets, understands, acts within, or what she notices of, the situation at hand.

Those responsible for designing and implementing information systems use symbolic representations or collections of symbolic representations (a form of context$_1$) in order to capture and maintain relevant aspects of context$_2$ and context$_3$. This process is never comprehensive or complete. There are limits to what any representation system can reflect about the environment in which it was originally embedded (Shanon 1990). "Context, in principle, is infinite. The describer selects certain layers for inclusion, and decides which of those to foreground" (Duff and Harris 2002).

The purposes and intentions of the actors associated with the resources constitute an important part of the resources' context. Identify the purposes can be essential to determining what should be done with the resources. For example, the creator of a Microsoft Word document might intend to share with future users the text that is immediately visible when opening the document in Word, but she might not intend to disclose "hidden" information embedded in the file, such as tracked changes or embedded spreadsheet data. Similarly, one might identify numerous properties that are supported by a given file format, but then realize that the creator of a set of files in that format did not purposely set any of those properties or care whether they are reproduced over time.

4.2. Capturing the state of information from dynamic environments

Digital preservation is essentially ensuring that important characteristics and values of digital objects can be consistently reproduced over time within an acceptable range of variability. As discussed above, the generation of the characteristics and values at any time requires the interaction of numerous hardware and software components. Depending on the particular set of components and settings one is using, one's experience can vary dramatically. The variability can increase over time, as both the underlying technology and user behaviors evolve. Creators and users of digital objects will value and attempt to ensure the relative fixity of some properties and values more than others. For example, Alice might care deeply about the accurate reproduction of the diacritics that appear within the text of a document that she wrote, but not particularly care about where the individual lines break, or vice versa. Bob might not care whether he is looking at the TIFF or JPG version of an image but be very concerned that they have the same persistent identifier to ensure that they are associated with the same digital object within a repository that he trusts.

There are two fundamental options for capturing and persistently reproducing information from a dynamic, distributed environment. First, one can create a snapshot of the entire state of the information at a point in time. Examples of this approach include generating a backup of a database and harvesting a snapshot of a web site. The other approach is to account for changes to information over time (i.e. capture change logs and audit trails). Examples of this approach are the revision logs available through Wikipedia and the records that are generated within a transaction processing system. Appropriate digital curation involves the identification, capture, perpetuation and reproduction of properties and values at one or more levels of representation.

4.3. Avoiding unnecessarily tight coupling to specific technologies

Digital objects are inherently dependent on technological components. However, it is possible to design, create and manage them in ways that minimize their dependence on specific technologies. As discussed above, interoperability across systems has long been a goal of computer system designers. However, there is a frequently competing goal of hardware, software and service vendors to lock consumers into their specific offerings. If Alice has created a large set of digital objects in a format that can only be read within a specific application, and that application can run on only one operating system, then Alice is very likely to purchase future releases of the application and operating system, i.e. she is locked in. If she has only one copy of the objects on her computer's hard drive or in a hosted space on the Web, she could lose the objects if a hard drive crashes. Relying completely on one hosted service provider to maintain her digital objects can also result in loss if the provider goes out of business, changes its offerings, cuts off her service, falls victim to a security breach or storage failure, or decides to delete the objects for some reason.

Digital curation is well served through “robust design” (Hargadon and Douglas 2001), which is effective in the short-term but also sufficiently flexible to remain effective in a wide range of possible future contexts. Limiting the interdependencies between subsystems can also make a design more robust against disruptions from the environment (Simon 1962). To avoid lock-in to particular combinations of hardware and software, individuals and organizations can make use of:

- redundancy (Maniatis et al. 2005);
- storing information in multiple ways (e.g., online services, formats, systems);
- diversity in technological approaches (Rosenthal et al. 2005) and business models;
- abstraction;
- virtualization (Moore 2008);
- detailed descriptive and administrative metadata beyond that which is required for immediate use and
- the development and adoption of open standards in ways that are attentive to the need for flexibility (Hanseth, Monteiro, and Hatling 1996; Egyedi 2001).

System evolution, sustainability and innovation can also be greatly facilitated through modularity (Langlois and Robertson 1992; Baldwin and Clark 2000).

4.4. Data extraction and recovery

Digital curation activities by individuals and organizations can significantly reduce the risk of information loss or corruption, but they will never entirely eliminate the need to extract or recover information from problematic hardware and software. The extraction and recovery can occur at any of the levels of representation discussed in this chapter.

Recovery of data from physical media has been a topic of discussion in the professional library and archives literature for several years (Ross and Gow 1999).There is also an active, expanding industry associated with digital forensics, which focuses on the discovery, recovery, and validation of information from computer systems that is often not immediately visible to common users. Several projects and authors have recently investigated the use of forensic tools and techniques for acquiring digital collections in libraries and archives (John 2008; Kirschenbaum, Overden and Redwine 2011; Woods, Lee and Garfinkel 2011).

4.5. Promoting discovery and sensemaking through associated metadata

Like context, metadata is a relative concept (Huc, Levoir and Nonon-Latapiem, 1997; for more information on metadata and other controlled languages see also Trippel 2012 in this volume). In relation to the body of an email message, the message header can be considered metadata. However, if one is managing each message as a distinct object, then the header might be considered part of that object's content (i.e. data rather than metadata). Similarly, one could treat the cover of a book as either metadata about the book's main content or instead part of the book's content. From the perspective of digital curation, such distinctions can be important when determining what will get retained over time and what will be exposed to users. Contemporary digital images, for example, often contain significant amounts of metadata within the headers of the files. If one were treating the files (Level 3 in Table 1) as the targets of preservation, then the header information would be preserved along with them. However, if one were treating the rendered, viewable representation of the image's content as the target of preservation, one might decide to transform the image into a new file format, which could fail to include all of the same header information.

In order for the digital objects to serve as conveyers of meaning across contexts, it can be important to generate surrogates of the objects, which can be discovered and experienced along with the digital objects themselves. For example, building an index of the contents of a set of text files will allow users to efficiently search over the content of those files. Creating thumbnails of larger images can allow users to browse through, compare and select images of interest. An archival finding aid that characterizes a whole collection of papers from an individual can provide a more holistic and richly contextualized view into the collection than one would get solely from encountering the objects on their own.

A broad category of activities by information professionals is called intellectual control (Pearce-Moses 2005). This involves mechanisms and conventions for bringing additional useful order to materials. This can involve naming conventions, authority control, and a variety of mappings across inconsistent terminologies.

4.6. Acting locally, but thinking globally

Most communication takes place within relatively well-bounded contexts. An email exchange between two individuals involves those two people and their associated expectations, background and understanding. Public broadcasts usually conform to well-established genre conventions and reach audiences that can be predicted with a fairly high degree of confidence. The exceptions – such as a private email exchange that is forwarded accidentally or "leaked" to a third party or an online story that "goes viral" and reaches an unexpectedly large and diverse audience – are noteworthy precisely because they are exceptional. However, one could argue that the reach, volume and potential persistence of digital communications makes such cases increasingly likely.

There is a more intermediate zone of interaction in which people have become quite accustomed to use of their traces beyond the original communication context. Family photos, meeting minutes and slides from conference presentations are all artifacts that serve an immediate purpose but are also often retained in order to serve further purposes in new contexts. The archives and records management literature characterize such phenomena as cases of "secondary use," while the literature on *personal information management* (PIM) investigates them as cases of re-finding and re-use. In both cases, there is an acknowledgement of the value that can be derived from perpetuating meaningful information beyond the context in which it was initially created or encountered.

No proposal for digital curation is likely to be viable if it requires individuals to devote substantial attention to secondary use of their information. People are simply too busy getting on with their lives to focus heavily on future use scenarios. However, a little bit of attention to digital curation within a local context can potentially go a long way toward more global goals. If Bob creates a collection of photos in a way that is not locked into one specific, proprietary application, this can increase the chances that both he and others will be able to make meaningful use of the photos in the future. If Alice uses simple, but consistent file naming conventions, both she and others will be able to make better sense of her files in the future. Whether one is operating in a home environment or a large bureaucracy, the questions to ask are similar: To whom might I hand off these digital traces in the future? How would that work? What are the likely motivations and needs of the recipient?

5. Conclusion

Digital traces can convey a diversity of information across space and time. They are recent additions to the repertoire of human communication mechanisms. I have argued that one of the fundamental considerations of digital traces is that they exist and can be encountered at multiple levels of representation. There is no single, canonical representation of digital resources, because the salience of a particular level of representation will vary by context. Parties responsible for digital collections should be attentive to digital traces' multiple levels of representation, making informed decisions about which levels should be encountered by future users and in what ways. Digital curation – conscious management of digital resources in order to ensure appropriate and meaningful use of the resources over time – can make the difference between accidental and ad hoc communication with the future, on the one hand, and conveyance of rich and responsible social memory, on the other.

Acknowledgements

This chapter has benefited from valuable comments and suggestions from Alexander Mehler and an anonymous reviewer.

Notes

1. For discussions of these interactions within the context of specific conversations, see Austin (1962) and Grice (1975). For a broader investigation of the relationships between the meaning of "texts" (e.g., documents, statements) and language systems, see Halliday and Matthiessen (2004).
2. As pointed out by a reviewer of this chapter, the concept of common ground is often formalized as a set of propositional assertions or attitudes. For example, A and B have common ground, because A knows something that B knows and B knows that A knows it. Digital objects cannot themselves have knowledge of propositions. However, meaningful use of a digital object requires that the user of the object has access to many of the same facts and assumptions as those who engaged in the creation of the object. Flouris and Meghini (2007) offer the concept of "Underlying Community Knowledge" as the language and theory associated with a digital object that are necessary for deriving its meaning.
3. Paul Ricoeur (2004) distinguishes between cerebral, affective and documentary traces. My use of the term is most closely aligned with, but somewhat broader than, his third category.
4. SHA-1 (NIST 2004) is one of many algorithms that can be used to generate cryptographic hashes. It can take any possible bitstream as an input and then generate a fixed-size bitstream (hash) as an output. Cryptographic hash algorithms are designed so that it is relatively easy to compute the hash of a given input but extremely difficult to 1) reconstruct the original input based on knowing the value of its hash output or 2) generate a different bitstream (including an alteration of the original bitstream) that can result in the same hash output.
5. Strictly speaking, there is always some chance that when two non-identical bitstreams (that differ by at least one of the 1 or 0 values) are run through the same hashing algorithm, they will generate the same hash value. However, if one uses a sufficiently robust hashing algorithm, the probability of this happening is so unlikely as to be effectively zero.
6. Brian Carrier (2006: 60) presents a similar idea within the context of digital forensics, which he calls an "observation tree."
7. There are many other models of communication that emphasize representation at several layers, see e.g. Strawson's three senses of meaning (1973), Levelt's "tiers" (1989) and Gumperz's "levels of signaling" (1982).

8. The hex editor itself provides a level of mediation, because it presents a *bytestream* as a set of hexadecimal values representing bytes and usually also the ASCII text values associated with the byte values. For example, the bitstream 01000001 would be seen on the left side (hexadecimal view) of a hex editor as the value "41" and on the right side (ASCII text view) of the hex editor as the value "A." Strictly speaking, humans very rarely interact with bitstreams represented as series of 1s and 0s.
9. This assumption of potential access to any level of representation is not supported by the model of "digital preservation as communication with the future" presented by Mois, Klas and Hemmje (2009), which is composed of six layers: data, representation, preservation, presentation, knowledge and content. According to Mois et al., "actual transmission" occurs only through the bottom (data) layer. The access process is assumed to always work up through the layers from the bottom to the top: "A future system is able to unpack each of the packages and gets with each of these steps the necessary information to process the information on the next higher level" (116).

References

Austin, J. L.
1962 *How To Do Things with Words*. Cambridge, MA: Harvard University Press.
Baldwin, Carliss Y. and Kim B. Clark
2000 *Design Rules. Vol. 1: The Power of Modularity*. Cambridge, MA: MIT Press.
Barwise, Jon and John Perry
1983 *Situations and Attitudes*. Cambridge, MA: MIT Press.
Bouquet, Paolo, Luciano Serafini and Richmond H. Thomason (eds.)
2008 *Perspectives on Contexts*. Stanford, CA: Center for the Study of Language and Information.
Cargill, Carl F.
1997 *Open Systems Standardization: A Business Approach*. Upper Saddle River, NJ: Prentice Hall.
Carrier, Brian D.
2006 *A Hypothesis-Based Approach to Digital Forensic Investigations*. Ph.D. dissertation, Center for Education and Research in Information Assurance and Security, Purdue University.
Clark, Herbert H.
1996 *Using Language*. Cambridge, UK: Cambridge University Press.
Cohen, Michael D. and Robert Axelrod
1998 Complexity and Adaptation in Community Information Systems: Implications for Design. In: Toru Ishida, J. G. Carbonell and J. Siekmann (eds.), *Community Computing and Support Systems: Social Interaction and Collaborative Work*, 16–42. Berlin: Springer.
Dewey, John
1931 Context and Thought. *University of California Publications in Philosophy* 12(3): 203–24.
Duff, Wendy and Verne Harris
2002 Stories and Names: Archival Description as Narrating Records and Constructing Meanings. *Archival Science* 2(3–4): 263–285.

Egyedi, Tineke
2001 Infrastructure Flexibility Created by Standardized Gateways: The Cases of XML and the ISO Container. *Knowledge, Technology & Policy* 14(3): 41–54.

Farmer, Dan and Wietse Venema
2005 *Forensic Discovery*. Upper Saddle River, NJ: Addison-Wesley.

Flouris, Giorgios and Carlo Meghini
2007 Terminology and Wish List for a Formal Theory of Preservation. In *Proceedings of PV 2007 – Ensuring Long-term Preservation and Value Adding to Scientific and Technical Data,* Weßling, Germany, October 2007.

Gadamer, Hans Georg
1989 *Truth and Method*. 2nd rev. ed. New York, NY: Crossroad.

Gaines, R. Stockton
1965 On the Translation of Machine Language Programs. *Communications of the ACM* 8(12): 736–41.

Garfinkel, Simson L. and Abhi Shelat
2003 Remembrance of Data Passed: A Study of Disk Sanitization Practices. *IEEE Security and Privacy* 1: 17–27.

Grice. H.P.
1975 Logic and Conversation. In: Peter Cole and Jerry L. Morgan (eds.), *Syntax and Semantics, vol. 3*, 41–58. New York, NY: Academic Press.

Gumperz, John Joseph
1982 *Discourse Strategies*. New York, NY: Cambridge University Press.

Halliday, M.A.K. and Christian M.I.M. Matthiessen
2004 *An Introduction to Functional Grammar*, 3rd edition. London: Arnold.

Halpern, Mark I
1965 Machine Independence: Its Technology and Economics. *Communications of the ACM* 8(12): 782–785.

Hanseth, Ole, Eric Monteiro and Morten Hatling
1996 Developing Information Infrastructure Standards: The Tension between Standardisation and Flexibility. *Science, Technology and Human Values* 21(4): 407–426.

Hargadon, Andrew B. and Yellowlees Douglas
2001 When Innovations Meet Institutions: Edison and the Design of the Electric Light. *Administrative Science Quarterly* 46(3): 476–501.

Hedstrom, Margaret and John Leslie King
2006 Epistemic Infrastructure in the Rise of the Knowledge Economy. In: Brian Kahin and Dominique Foray (eds.), *Advancing Knowledge and the Knowledge Economy*, 113–134. Cambridge, MA: MIT Press, 2006.

Heyer, Gerhard, Florian Holz, and Sven Teresniak
2012 P2P-based communication. In: Alexander Mehler, Laurent Romary and Dafydd Gibbon (Eds.): *Handbook of Technical Communication*. Berlin/ Boston: De Gruyter.

Huc, Claude, Thierry Levoir and Michel Nonon-Latapie
1997 Metadata: Models and Conceptual Limits. In: *Proceedings of the Second IEEE Metadata Conference*. Silver Spring, MD: IEEE Computer Society, 1997.

IEEE
1990 *IEEE Standard Computer Dictionary: A Compilation of IEEE Standard Computer Glossaries*. New York, NY: Institute of Electrical and Electronics Engineers.

ISO
1993 ISO/IEC 2382–01, *Information Technology Vocabulary, Fundamental Terms*.

ISO
1994 ISO/IEC 7498–1, *Information Technology – Open Systems Interconnection – Basic Reference Model: The Basic Model*.

John, Jeremy Leighton
2008 Adapting Existing Technologies for Digitally Archiving Personal Lives: Digital Forensics, Ancestral Computing, and Evolutionary Perspectives and Tools. In: *iPRES 2008: The Fifth International Conference on Preservation of Digital Objects, London, UK, September 29–30*.

Kirschenbaum, Matthew G., Richard Ovenden and Gabriela Redwine
2010 *Digital Forensics and Born-Digital Content in Cultural Heritage Collections*. Washington, DC: Council on Library and Information Resources.

Krueger, Charles W.
1992 Software Reuse. *ACM Computing Surveys* 24(2): 131–183.

Langlois, Richard N. and Paul L. Robertson
1992 Networks and Innovation in a Modular System: Lessons from the Microcomputer and Stereo Component Industries. *Research Policy* 21(4): 297–313.

Lee, Christopher A.
2011 A Framework for Contextual Information in Digital Collections. *Journal of Documentation* 67(1): 95–143.

Lee, Christopher A. and Helen R. Tibbo
2011 Where's the Archivist in Digital Curation? Exploring the Possibilities through a Matrix of Knowledge and Skills. *Archivaria* 72: 123–68.

Levelt, Willem J.M.
1989 *Speaking: From Intention to Articulation*. Cambridge, MA: MIT Press.

Lewis, David. K.
1969 *Convention: A Philosophical Study*. Cambridge, MA: Harvard University Press.

Maniatis, Petros and Mema Roussopoulos, T.J. Giuli, David S. H. Rosenthal, and Mary Baker
2005 The LOCKSS Peer-to-Peer Digital Preservation System. *ACM Transactions on Computer Systems* 23(1): 2–50.

Menke, Peter
2012 Evaluation of technical communication. In: Alexander Mehler, Laurent Romary and Dafydd Gibbon (eds.): *Handbook of Technical Communication*. Berlin/Boston: De Gruyter.

Messerschmitt, David G. and Clemens Szyperski
2003 *Software Ecosystem: Understanding an Indispensable Technology and Industry*. Cambridge, MA: MIT Press.

Mois, Martin, Claus-Peter Klas, and Matthias L. Hemmje
2009 Digital Preservation as Communication with the Future. In: *DSP'09 Proceedings of the 16th International Conference on Digital Signal Processing*, 112–119. Piscataway, NJ: IEEE Press.

Moore, Reagan
2008 Towards a Theory of Digital Preservation. *International Journal of Digital Curation* 3(1), 63–75.

NIST
2004 *Secure Hash Standard*, Federal Information Processing Standards Publication 180–182.

Parmelee, R. P., T.I. Peterson, C. C. Tillman and D.J. Hatfield
1972 Virtual Storage and Virtual Machine Concepts. *IBM Systems Journal* 2: 99–130.

Panofsky, Erwin
1955 *Meaning in the Visual Arts: Papers in and on Art History.* Garden City, NY: Doubleday.

Pearce-Moses, Richard
2005 Intellectual Control. In: *A Glossary of Archival and Records Terminology.* Chicago, IL: Society of American Archivists.

Peterson, Mark Allen
2003 Anthropology and Mass Communication: Media and Myth in the New Millennium. Oxford: Berghahn.

Rahtz, Sebastian
2012 Representation of Documents in Technical Communication. In: Alexander Mehler, Laurent Romary and Dafydd Gibbon (eds.), *Handbook of Technical Communication*. Berlin/Boston: De Gruyter.

Ricœur, Paul
2004 *Memory, History, Forgetting*. Chicago, IL: University of Chicago Press.

Rommetveit, R.
1979 Words, Contexts and Verbal Message Transmission. In: R. Rommetveit and R. M. Blakar (eds.), *Studies of Language, Thought and Verbal Communication*, 13–26. London: Academic Press.

Rosenthal, David S. H., Thomas Robertson, Tom Lipkis, Vicky Reich, and Seth Morabito
2005 Requirements for Digital Preservation Systems: A Bottom-up Approach. *D-Lib Magazine* 11(11). Online available at http://www.dlib.org/dlib/november05/rosenthal/11rosenthal.html, retrieved on May 31, 2011.

Ross, Seamus and Ann Gow
1999 *Digital Archaeology: Rescuing Neglected and Damaged Data Resources*. London: British Library.

Shannon, C.E.
1948 A Mathematical Theory of Communication, *Bell System Technical Journal* 27: 379–423, 623–656.

Shanon, Benny
1990 What Is Context? *Journal for the Theory of Social Behaviour* 20(2): 157–166.

Sheth, Amit P.
1999 Changing Focus on Interoperability in Information Systems: From System, Syntax, Structure to Semantics. In: Michael Goodchild, Max Egenhofer, Robin Fegeas and Cliff Kottman (eds.), *Interoperating Geographic Information Systems*, 5–29. Boston, MA: Kluwer Academic Publishers.
Simon, Herbert A.
1962 The Architecture of Complexity. *Proceedings of the American Philosophical Society* 106: 467–482.
Stalnaker, Robert C.
1978 Assertion. *Syntax and Semantics 9*: 315–322.
Strawson, P.F.
1973 Austin and 'Locutionary Meaning.' In: *Essays on J. L. Austin*, 46–68. Oxford: Oxford University Press.
Stührenberg, Maik
2012 Foundations of markup languages. In: Alexander Mehler, Laurent Romary and Dafydd Gibbon (eds.), *Handbook of Technical Communication*. Berlin/Boston: De Gruyter.
Tolk, Andreas
2003 Beyond Technical Interoperability: Introducing a Reference Model for Measures of Merit for Coalition Interoperability. *Paper presented at the 8th International Command and Control Research and Technology Symposium, Washington, DC*.
Trippel, Thorsten
2012 Controlled language structures in technical communication. In: Alexander Mehler, Laurent Romary and Dafydd Gibbon (Eds.): *Handbook of Technical Communication*. Berlin/Boston: De Gruyter.
van Diessen, Raymond J.
2002 *Preservation Requirements in a Deposit System*. Amsterdam: IBM Netherlands.
Wilson, Donald W. and David J. Moss
1965 CAT: A 7090–3600 Computer-Aided Translation. *Communications of the ACM 8(12)*: 777–781.
Woods, Kam, Christopher A. Lee, and Simson Garfinkel
2011 Extending Digital Repository Architectures to Support Disk Image Preservation and Access. In: *Proceedings of the 2011 Joint Conference on Digital Libraries*, 57–66. New York: Association for Computing Machinery.
Zuboff, Shoshana
1988 *In the Age of the Smart Machine*. New York, NY: Basic Books.

III. Communication by means of technology

16. Internet-Based Communication

Ulli Waltinger and Alexa Breuing

Today, internet-based communication encompasses a broad range of media services and technologies. While its architecture is still based on the traditional client-server model, its associated principles of *communication* and *information* are affected by the rising web-based communication technologies from the area of the *Web 2.0* and the *Semantic Web*. The current developments within the web have significantly stimulated many people to become involved in web-based communication. Accordingly, the web has changed from a medium for passive consumers to a network of active users who communicate via social networks, contribute to online communities, write weblogs, and publish podcasts. This survey chapter focuses on web-based technologies for human-human communication, with a special emphasis on rising technologies and applications. It reviews the principles of communication technologies within the *Web 2.0* and the *Semantic Web* and provides an overview of existing applications of recent web-based communication techniques.

1. Introduction

> *Computer Mediated Communication is a process of human communication via computers, involving people, situated in particular contexts, engaging in processes to shape media for a variety of purposes.* December (1996)

The internet has influenced and changed the way people communicate to an unprecedented degree. In modern times, online media technologies have become indispensable for the global communication process not only within the business area but also within our daily life. These developments in turn have significantly stimulated many users to become involved in web-based communication. In recent years, the World Wide Web has changed from a medium for passive consumers to a network of active users, who write weblogs (e.g. *Blogger*), participate in wikis (e.g. *Wikipedia*), publish podcasts (e.g. *iTunes*) or contribute and communicate within online communities such as using *Facebook* or *LinkedIn* (Stocker and Tochtermann, 2009). It has influenced the way we communicate (Mulvenna et al., 2000), interact (Kraut et al., 1998; Katz and Rice, 2002), shop, work, and inform ourselves (Jones, 1998). We might even argue that the internet helps to bring people closer together as it enables easy and cheap communication across vast distances (Gates, 2000). And although not everybody in the world possesses an internet access, the digital communication has become easier, faster, and, in principle, free. Altogether, the internet has obviously changed the way people collaborate and interact with each other, without regard to their geographic location.

The current web is characterized by its flexibility. Recent online publishing tools have simplified matters so that anyone can write, publish, and contribute data to the web. By now, internet *users* have become internet *editors* by taking a more active role in the collective development of a global knowledge (Surowiecki, 2004; Hammwöhner, 2007a). Thus, the internet has turned into a network of creative knowledge, where different cultures and disciplines have linked up to make something that is sometimes called the *collective intelligence* (O'Reilly, 2005). It establishes a mixture of very general and very specific interests and is thus a very precise, *self-limiting* and *self-organizing form* (e.g. in the domain of social software) (de Kerckhove et al., 2008). Hence, despite the development of web-based technologies, we have also witnessed a tremendous growth of data published on the web. However, as the available information increases, the inability of users to assimilate and profitably utilize such large amounts of information becomes increasingly apparent. For computer scientists, this body of information has opened up a new area of research, to develop and test new computational methods and to benefit from the collectively organized knowledge within the online medium (Mills, 2009).

> *Internet-based, computer-mediated communication involves information exchange that takes place on the global, cooperative collection of networks using the TCP/IP protocol suite and the client-server model for data communication. Messages may undergo a range of time and distribution manipulations and encode a variety of media types. The resulting information content exchanged can involve a wide range of symbols people use for communication.* December (1997)

People consider the WWW and the internet to be synonyms. From a technical standpoint, this is wrong. Initially, the internet refers to a system composed of computers connected by physical objects (e.g. cables). It is a physical network which facilitates the transfer and exchange of data. In this respect, the *ARPANET* (Advanced Research Projects Agency Network) was one of the first prototypes of such a system of communication (in 1969) (Abbate, 2000). It realized the first version of the *Transmission Control Protocol* (TCP) and the *Internet Protocol* (IP), used to exchange data and files and to connect to other computers (Caldarelli, 2007).

The WWW can be seen as a service layered over the internet. It refers to a hypertext system which is primarily based on three components: the *Hypertext Transfer Protocol* (HTTP) as the protocol to transfer data over a network; the *(eXtensible) HyperText Markup Language* (X)HTML allowing the definition of how the accessible information (web document) is structured, presented and connected to each other via hyperlinks; and finally, a *Uniform Resource Locator*

(URL) which serves as a unique identifier (web address) of a given resource (web document). Thereby, hyperlinks are not only used to interlink web documents but to interconnect other resources within the WWW as well. One of the first global hypertext projects was proposed by Tim Berners-Lee at the nuclear research center CERN in Geneva in 1989. Initially brought into being to share research findings and to improve collaboration with colleagues, in 1990 Berners-Lee presented the first web viewer that enabled users to view (browse) the first published texts. Subsequently, the web browser technology improved, starting from *Viola* (Pei Wei in 1992)[1], which added the ability to display graphics, to *Mosaic* (Marc Andreessen in 1993)[2] which turned into the later *Netscape Navigator* (1994)[3] and quickly became the leader in the rapid growth stage of the WWW (Abbate, 2000). In the 1990s, the WWW consisted of only a few static web pages which were manually listed in a first web directory/catalogue. But gradually, with the development of the first web search engines (e.g. *AltaVista*[4] and *Yahoo!*[5] in 1995, *Google*[6] in 1998) and decades after the first idea of a knowledge identification and retrieval system (*Memex*) (Bush, 1945), it soon led to the phenomenon that not only companies but also private persons tend to have individual *home pages* – accessible and findable on the web. Since that time, we have been able to observe a rising popularity in participation, contribution, and interaction with web-based technologies. This progress transformed the web into what nowadays is called *Web 2.0* (O'Reilly, 2005), *Semantic Web* (Berners-Lee et al., 2001), or *Social Semantic Web* (Blumauer and Pellegrini, 2009) – the web as we know it today.

2. Technologies of web-based communication

Technical developments in terms of an increase in broadband connections, a spread of internet access, and a decrease in prices of storage media (which are needed to archive such an enormous amount of digital data) was attended by significant developments regarding web-based communication technologies. While at the beginning of the WWW, say *Web 1.0*, it was all about getting people to publish and to find and connect information via hyperlinks, the WWW gradually transformed into a *web of social participation* (Mills, 2009). As a result, not only documents but also users are connected with each other. Since 2005, the term "Web 2.0" has become a popular buzzword for a number of interactive and collaborative elements within the WWW and we may argue that a so-called *Web 3.0* is currently emerging (Berners-Lee et al., 2001); a web where the focus is on the *automatic* identification and representation of the *semantic meaning* of information published on the web. More precisely, where the focus is on *connecting knowledge* rather than on documents and on putting them into a context.

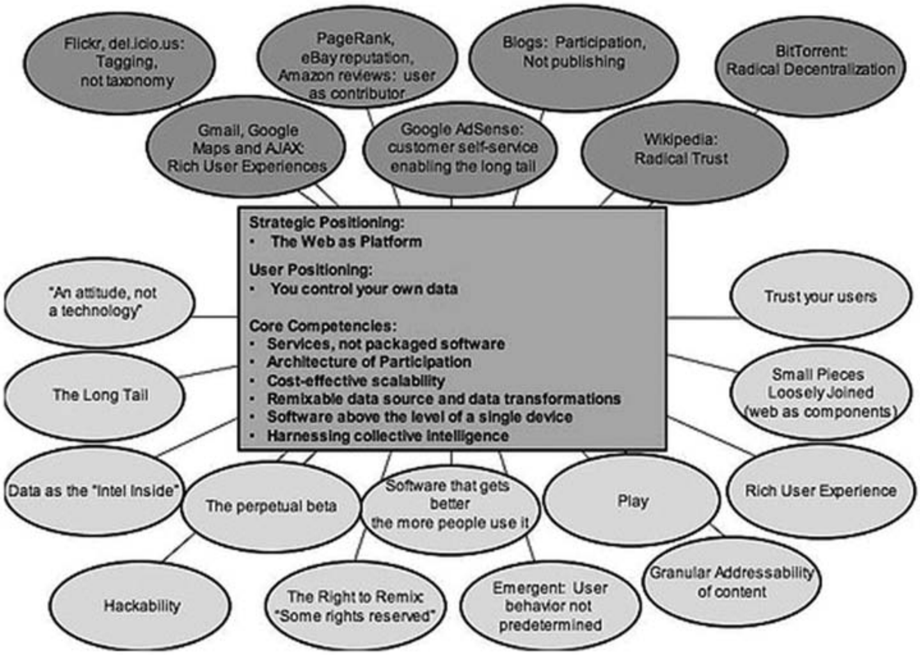


Figure 1. Ideas that radiate out from the Web 2.0 core by O'Reilly (2005).

2.1. Communication within the Web 2.0

From a technical perspective, the phenomenon of Web 2.0 actually refers to techniques and services mostly developed in the late 90s. These include *Rich Site Summary* or *Really Simple Syndication* (RSS) (e.g. Netscape RSS in 1997) and web services such as *Application Programming Interfaces* (APIs), *Asynchronous JavaScript and XML* (Ajax in 1998), *Google* (1998), *Wikipedia* (2001) or *Flickr* (2002). Today, the term *Social Software* (referring to, for example, *wikis*, *weblogs*, and *microblogging services*) is mentioned as a special characteristic of the Web 2.0 (Blumauer and Pellegrini, 2009). So what is actually meant by "Web 2.0" and how does it facilitate web-based communication? In 2005, Tim O'Reilly (2005) presented the following initial sketch of the main principles of the Web 2.0 (see Figure 1):

1. *The Web as a platform:* In the beginning of the WWW, we could identify a clear separation between offline software applications and online web services. More precisely, in the past, software platforms and tools were merely installed locally on a single computer whereas web-based applications were mostly provided as *read-only* web services. Thus, even tools which were used to access and browse the WWW were based on offline software appli-

cations (e.g. Netscape Browser or AOL client). Thereby, all information and data gathered on the web had to be stored additionally on the user's hardware. Online bookmarks (in terms of favorite website hyperlinks), for example, could only be saved within the browser toolbar of the offline application. A true reciprocity or inter-connection between the local installed software tools and the different web resources was rather limited.

With the emergence of new online application services associated to the Web 2.0 such as the online photo-service *Flickr* or so-called social bookmarking tools such as *del.icio.us*, this paradigm changed. Users were now enabled to use online applications just like offline software. That is, they were enabled to contribute, store, access and supplementarily share their own created data and information within different online web services, always accessible from different interfaces (e.g. pc or cell phone) and independent from their current location. Accordingly, the users started to use the web as a new platform, just like their own computer. Meanwhile, it is common usage not to store individual documents, photos, or bookmarks on a single computer but additionally or even only on certain online service platforms and hence to share, categorize, and exchange this data with other internet users. A currently much discussed approach utilizing this trend is *cloud computing*. Cloud computing refers to both the applications delivered as web services and the hardware and system software that provide those services (Armbrust et al., 2009). By means of so-called *clouds*, people are able not only to provide data but also to access other computational resources such as software or calculating capacity by means of a computer network. Cloud computing especially has commercial relevance, as the flexible allocation and utilization of services as well as extensive capacities allow a great economy of scale (Baun et al., 2011). However, applications for private use receive substantial boost as well. Thus, for example, *Apple* already provides its customers with several cloud computing services (as *iCloud*). Altogether, we may argue that the vision of O'Reilly (2005), who stated in 2005 that the web is *ubiquitous and serves as a platform without hard boundaries*, is currently at the point of origin.

2. *Harnessing collective intelligence:* Any internet user may write and contribute content to the web and any other user is allowed to edit this content. Accordingly, using the web as a platform allows the harnessing of collective intelligence in terms of *Weblogs*, user-created structures such as *Folksonomies* or collaborative work by means of *Social Software* (see Section 3). This principle of harnessing *the wisdom of crowds* (Surowiecki, 2004) also forms the basic idea behind the collaborative approach for software development (see Principle 4).

   One aspect of collective intelligence becoming popular and possible by means of Web 2.0 technologies is *crowdsourcing*. It describes the act of out-sourcing tasks through an open call to an undefined group of people and gathering those who are most fit to perform the announced task (Howe, 2008). Common examples are the online encyclopedia *Wikipedia* and *OpenStreetMap* where users can upload and edit vector data.

3. *Data as the next Intel Inside:* Database management is a core competence of most web companies as data present the main component of Web 2.0 applications. Every internet user accesses, utilizes, and often even extends the data on the web. Thus, the question of accessibility, usability, and ownership of this data needs to be clarified and put into the focus. Think of *Facebook*, where the data is made up of personal information provided by the numerous Facebook users. The more people participate and eventually assemble the data, the more is its increase in value. As a result of its currently more than 750 million active users[7], Facebook's value has, in the meantime, swelled to $50bn[8].
   In future, data suppliers and application vendors will battle for certain data classes that will become highly important for Web 2.0 applications. Companies already begin to realize that control over data may be their chief source of competitive advantage.
4. *End of the software release cycle:* Software is no longer distributed as a product/software release but as a service within the web. The implementation and testing of software become a public business turning internet users into co-developers. Thus, software development turns into a practice extending the open source philosophy *"Release early. Release often."* (Raymond, 2001) in that the product is additionally developed openly. Only by this, Web 2.0 applications are capable of meeting the challenges of daily operations and constantly changing user needs. Consequently, many open source initiatives (for example *Linux*) have obtained high quality results from collaborative work (Vivacqua and Borges, 2010). In addition, web developers have the opportunity to try out new features by providing them on their websites and monitoring the corresponding user behavior in real-time.
5. *Lightweight programming models:* Web services are more accepted and more prevalent the simpler they are. Thus, they have to be realized consistent with the principle *"innovation in assembly"* (O'Reilly, 2005). Therefore, lightweight programming models have to be supported in a way that allows an easy combining of content and features of different, existing web services to provide new services. The focus has to be on syndication rather than on coordination and on lowering the barrier to the public to (re-)use data and services. Furthermore, the service design needs to provide for "hackability" and "remixability" that enable and almost invite the harnessing of collective intelligence and thus to improve the web service collaboratively (O'Reilly, 2005).

Based on the principle of simplicity, many web services achieved great success. *RSS*, for example, has become perhaps the most widely deployed web service as it is simple, open, and easy to extend. Another example illustrating the great popularity of simple web services compared to complex ones is *Amazon* which provides its web services in two ways: On the one hand strictly adhering to the technical formalisms of *SOAP* (Simple Object Access Protocol) and, on the other hand, utilizing the simple distribution of *XML* data over HTTP based on a lightweight approach known as *REST* (Representational State Transfer). The majority of the Amazon usage (even 95 percent) is of the service building on REST (O'Reilly, 2005). This clear user preference is observable within other web services as well and it is not at all absurd to state that *Google* became the most successful search engine due to the simplicity of its application.

6. *Software above the level of single devices:* Web applications and services may no longer be limited to the PC platform but should also be suitable for mobile terminals and other devices. *iTunes* is the best example to date of the accomplishing of a seamless transition from several mobile devices to a massive web back-end from scratch. Other examples are *Skype* and *Facebook* which, amongst others, are available as single apps for smart phones. As a result, this Web 2.0 domain strongly drives the development of new, widely varying mobile devices such as iTV and traffic and activity monitoring platforms.
7. *Rich user experience:* The focus is on creating web-based software that works in a similar fashion and is perceived by the user also to be similar to a device-based software application. That is, on combining features of graphical user interfaces (e.g. drag-and-drop) and multimedia content on the web by means of client-side programmability and asynchronous data retrieval (e.g. *XMLHttpRequest* via Ajax (Garrett, 2005)). This enormous innovation in the field of user interfaces finally enables the construction of web applications which are as rich as local PC-based applications. Furthermore, we can expect to see many newly conceived applications as well as re-implementations of PC applications in the near future. Actually, *Gmail* has already provided innovations in e-mail which combine the strengths of the web with a browser-based user interface that is conform with independent e-mail software such as *Outlook Express* or *Thunderbird*.

Obviously, the majority of these principles can be positively attributed to current web interaction and communication techniques. The future address book will utilize web-based synchronization processes similar to Gmail that remember every mail, e-mail address, and phone number and which composes a separate social network allowing for a fast contacting and communication. Apple's *iCloud* already veers towards this evolution as synchronizing contacts, calendars, and e-mails as well as music, apps, etc. between Macs, PCs, and iOS-devices via up- and downloads of data within a cloud. Furthermore, collaborative writing will reach a higher level as it will not be limited to editing only but will be supported by word processing applications similar to those for stand-alone documents (e.g. *Microsoft Word*).

The actual implementation of the Web 2.0 led to the development of so-called *personal publics* referring to online networks of communicative expressions regarding topics of mostly personal relevance (Schmidt, 2009). Such networks mainly include individual-related information such as one's date of birth, hobbies, or current situation and generally obtain rather small publics often consisting of already known people (see Section 3 for examples). Personal publics differ from other communication situations in terms of four attributes (cf. (Boyd, 2008)):

1. *Persistence:* Online expressions are recorded and available permanently. In contrast, information exchanged during everyday communication situations might be forgotten shortly after.
2. *Replicability:* As content is available in digital form, they can be copied without degradation and possibly even unnoticeably.
3. *Scalability:* Although personal publics might reach only a few audiences, their potential visibility is great. Many people, for instance, became prominent just by means of private videoclips provided online (e.g. on *Youtube*).
4. *Searchability:* Information can be accessed through search, uncoupled from their context.

In particular, Web 2.0 applications such as weblogs and social networks (see Section 3.4, 3.7) provide personal publics and thus generate another type of human communication based on technical web platforms.

2.2. Communication within the Semantic Web

The so-called "Semantic Web" can be seen as an extension of the current web (Web 2.0). It is sometimes referred to as a component of an upcoming Web 3.0 (Lassila and Hendler, 2007). In general, it is a vision that web resources should not only be readable and understandable for humans but also for machines, thus offering the crucial possibility of automation (Berners-Lee et al., 2001). The Semantic Web initiative has the objective to realize this vision by focusing on semantic annotations of web content and the usage of a shared terminology (see Figure 2). Thereby, the objective is not to establish a new web but rather to extend the existing WWW by a semantic layer.

> *The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation.* Berners-Lee et al. (2001, pp. 2)

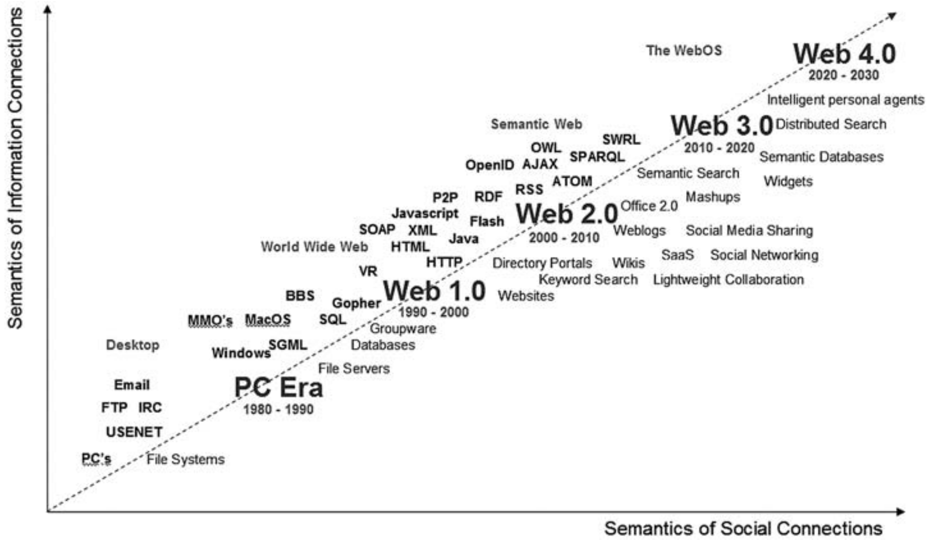


Figure 2. Ideas that radiate out from the Web 3.0 by Spivack (2007).

The semantic layer can be achieved by a more precise description of web documents by means of *metadata* (data about data) assigning semantics to the document content (Antoniou and van Harmelen, 2004). Hence, the Semantic Web vision follows the idea of connected intelligences in an ubiquitous web, where both humans and machines communicate with each other (*content intelligence*) (Mills, 2009). More precisely, in contrast to the Web 2.0, the Semantic Web is more about a web of *relations between resources* that denote real world objects such as people, countries, or events rather than a web of *accessible documents* (Guha et al., 2003).

> *The Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries. […] It is about common formats for integration and combination of data drawn from diverse sources, where on the original Web mainly concentrated on the interchange of documents.* Herman (2001, pp. 1)

Currently, the primary approach to achieve the annotation of documents is to define an ontology (e.g. *OpenCyc*[9], *GOLD*[10]) (see also Sharoff and Hartley (2012) in this volume). According to Gruber, an ontology is "*an explicit specification of a conceptualization*" (Gruber, 1993, pp. 199). The essential aspect of ontologies is a shared understanding of a specific domain among different web agents and applications and to subsequently use references to the concepts in the ontology as semantic annotations on top of web resources (top-down/bottom-up approach) (Berners-Lee et al., 2001).

> *Ontologies are the vocabulary and the formal specification of the vocabulary only, which can be used for expressing a knowledge base […] It should be stressed that one initial motivation for ontologies was achieving interoperability between multiple knowledge bases.* Hepp et al. (2008, pp. 6)

That is, users may create applications that collect web content from diverse sources, process, analyze, and annotate the information, and exchange their results with other applications by means of semantic markup specifications (Berners-Lee and Fischetti, 2000). Correspondingly, future software agents will carry out tasks for users rather than searching for information. Thus, completion of jobs such as finding a medical specialist nearby or setting up an appointment no longer consists of accomplishing numerous key-word-based search queries on the web and browsing the resulting text documents for important information manually. Instead, individual web agents communicate with other web agents, take into account one's circumstances such as the living place or time preferences, access one's personal calendar, and even reschedule appointments to present a solution. As much information exists only implicitly, the agents need a set of inference rules that enables them to conduct automatic reasoning (Berners-Lee et al., 2001). Hence, for Berners-Lee and colleagues, the challenge of the Semantic Web is to provide a representation formalism (ontology) that allows for reasoning about data to infer information which is not available explicitly (e.g. the distance from one place to another or connections between two persons). Currently, the most widely accepted languages for the description of ontologies are based on the *Resource Description Framework* (RDF) (Manola and Miller, 2004; Hayes, 2004; Klyne and Carroll, 2004). RDF is a markup language recommendation proposed by the *World Wide Web Consortium* (W3C) for the description of resources and their properties in a machine-processable manner and it provides the basis for most ontology languages (May, 2006). The identification of resources is ensured based on so-called *Uniform Resource Identifiers* (URIs) assigning a clear description to any resource on the web. The most familiar type of an URI is the *Uniform Resource Locator* (URL) which specifies where a specific resource is available and how to retrieve it. In contrast, URIs are more general and can refer to everything, thus additionally to objects of the real world such as persons, institutions, or words (Walton, 2006). RDF possesses a vocabulary containing a particular set of URIs with predefined meanings (Hayes, 2004; Hitzler et al., 2009). Additional information on URIs is available within the corresponding RFC (Berners-Lee, 1994).

RDF enables both the allocation of information in a human and machine-readable format and the realization of efficient data exchange. Thereby the resource description may be available as a data model in terms of a graph or serialized in a particular syntax. The former allows the relation of resources to other resources by means of triple-based *statements* each consisting of a subject (node), a predicate (link), and an object (node) (see Figure 3). As the object may

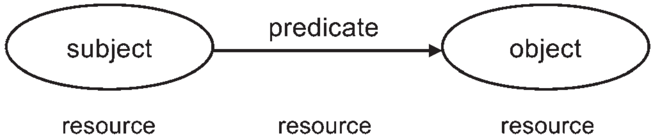


Figure 3. Components of a RDF statement.

be the subject of further statements, several statements can be concatenated and thus build up a directed graph (Walton, 2006). The serialized representation of RDF is based on *RDF/XML* which benefits from the advantages provided by *XML* (simplicity, generality and usability) but additionally enables the description of machine-understandable semantics (Beckett, 2004) (see also Stuhrenberg (2012) in this volume). The data is often stored in a relational database or in a database optimized for the storage of statements called triplestore. Altogether, RDF is a general-purpose language for the representation of semantic information on the web (Brickley and Guha, 2004). However, to build up ontologies, resources need to be arranged in categories enabling the definition of statements about classes of resources (Birkenbihl, 2006).

RDF Schema (RDFS) (Brickley and Guha, 2004) is an extension of RDF in that it provides basic elements for the definition of ontologies. In RDFS, resources with the same properties are considered to be instances within one class. Classes in turn are resources which can be referenced via URIs and which may have own properties. RDFS supports both the definition of sub-class relations and the definition of completely new relations (Aitken et al., 2004). The latter can be limited by domain and range specifications. Thus, RDFS constitutes an ontology language that allows for the definition of vocabularies, the specification of properties and their values in connection with different classes, and the definition of relations between resources (Stuckenschmidt and van Harmelen, 2005). However, the inability to define equalities between resources and additional limitations regarding properties present a substantial deficit of RDFS. Then again, exactly this was precisely a consciously made design decision to possess a simple and plain language for ontologies (Walton, 2006).

The mentioned inadequateness of RDFS is remedied by the allocation of a more expressive ontology language. The *Web Ontology Language* (OWL) (McGuinness and van Harmelen, 2004; Patel-Schneider et al., 2001, 2004) (see also Stührenberg and Sasaki and Lommel both in this volume) can be used to represent ontologies just like RDFS, but additionally provides more possibilities to express content and meanings. Thus, OWL extend RDFS by classes and properties allowing the definition of new classes based on enumerations of instances, calculations of intersections and unifications as well as specifications of cardinalities and complementary operations (Birkenbihl, 2006). However, a higher expressiveness always goes hand in hand with a reduced support for ef-

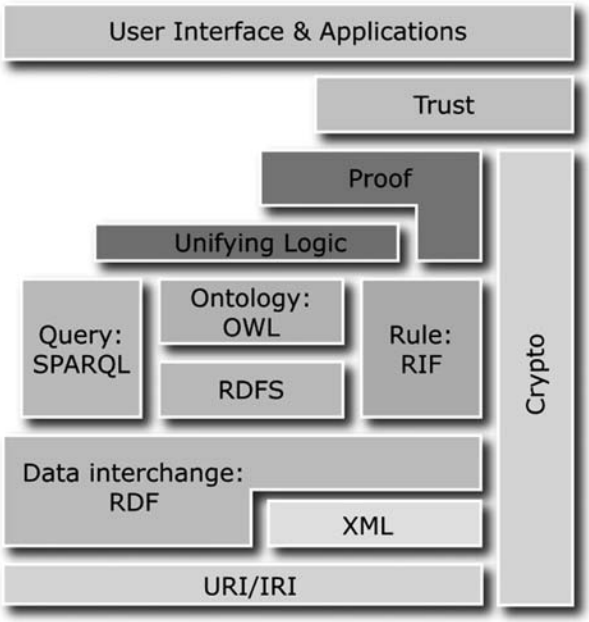


Figure 4. The Semantic Web Layer Cake proposed by (W3C, 2009).

ficient deduction. For this reason, three different levels of languages were defined: *OWL Lite*, *OWL DL*, and *OWL Full*. The expressiveness increases with each level, so that OWL Full is the most expressive one (Walton, 2006). Stuckenschmidt and van Harmelen (2005) define the three levels as follows:

1. *OWL Lite:* supports classification hierarchies and allows the formulation of equalities, inequalities, and simple constraints (e.g. cardinality values of 0 or 1).
2. *OWL DL:* allows both maximum expressiveness and computational completeness as well as decidability at the same time. “DL” refers to the correspondence with description logic, a knowledge representation language used for formal reasoning on terminological knowledge.
3. *OWL Full:* allows maximum expressiveness and syntactical liberty but is not determinable.

The manual creation of semantic annotation is both an expensive and time-consuming task that is extensively studied in the area of learning and managing ontologies (Staab and Studer, 2004). In this context, a lot of work has been done on developing tools for ontology engineering in order to support users in the process of creating ontologies (see, e.g., the *Protégé Ontology Editor and Knowledge Acquisition System* (Knublauch et al., 2004) or the *SWOOP* project

(Kalyanpur et al., 2004)). Miller (2001) from the W3C (W3C, 2009) describes the guiding architectural principles (see Figure 4) of the Semantic Web as follows:

1. *Everything Identifiable is on the Semantic Web:* Referring to the Semantic Web, all resources on the web need to have an identity. That is, people, countries, and other data objects referred to need to have an online representation in the form of an RDF-URI. URIs should support the effective integration, participation, and contextualization of data within the Semantic Web (Manola and Miller, 2004).
2. *Partial Information:* In accordance with the idea of the *web as a platform* (see Section 2.1.1), the web is ubiquitous in that it lacks hard boundaries. Semantic Web technologies need to focus on the aspect to *enable web access for anyone, anywhere, anytime, using any device* (Hoschka, 2011). In addition, the Semantic Web should tolerate partial information, such as missing links (e.g. non-existing targets) and/or incomplete or inconsistent information. Every user may contribute data and information to the web by creating different types of links to (web) resources (e.g., RDF-URI), which can be discovered and/or differently typed by others.
3. *Web of Trust:* The web is a collaborative environment (not read-only) that should offer confidentiality and confidence but simultaneously allows users to take responsibility for what they publish on the web (e.g. via signatures, group authoring, or versioning). In addition, obviously, all data, information, and statements accessible on the web occur in a specific context. It is the job of the respective applications to check the reliability and to evaluate the trustworthiness of the information found on the web.
4. *Evolution:* The Semantic Web should support the independence of different communities. According to their own specifications, these communities should be able to add and distribute information independently of other standards. Thereby it should be based on some *descriptive conventions* which can also be understood by the users.
5. *Minimalist Design:* In the context of the Semantic Web the *principle of minimalist design* refers to the aspect that every formulation of a *semantic assertion* should be based on a common and universally valid model (Berners-Lee, 1998). The idea behind is that it is the general validity and simplicity of a model which enables other applications to use, link or map to the formulated assertions.
6. *Building on Proven Ideas:* The Semantic Web community provides already existing technologies, formats, and frameworks to annotate and interlink data objects (e.g. RDF, OWL). Emergent applications and technologies should agree to the general principles of web automation, data aggregation, and support the establishment of a common syntax and vocabulary (Stumme et al.,

2011). For instance, to provide simple applications that do not ignore the possibility of additional complexity in the future (e.g. *Dublin Core* (Dublin Core Metadata Initiative, 2010) – see also Trippel (2012) in this volume).

The goal of the Semantic Web is to convert the web into a distributed, machine-understandable data resource by means of semantic markup specifications and metadata. This metadata should be defined, formulated, and structured in a way that, at best, enables machines and applications not only to process but also to understand it correctly (in terms of reasoning). Furthermore, it should allow for drawing conclusions based on existing data as well as the generation of new information in this way. Data objects within the Semantic Web do not target its direct usage by people but by machines. However, this approach implies several challenges for the computer science domain: Due to the enormous amount of data and information available on the Web, *it is impossible to manually enrich all of these resources* (Stumme et al., 2011). Furthermore, one faces questions of the following sort: How to establish such a semantic backbone for such a large number of distributed web resources? How to query, process, and analyze such an amount of ontology and *RDF* data efficiently on a decentralized network (see approaches like *SPARQL* (Prud'hommeaux and Seaborne, 2008))? How to ensure that two different web agents share the same semantics for a given annotated concept (in terms of standardizing the information distribution across the network)? In essence, there is the need to automatically learn, interpret, organize, and retrieve relevant information and resources so as to meet the heterogeneous requirements of users and agents (Chishti et al., 2009; Stumme et al., 2011).

3. Applications of web-based communication

A variety of web tools, applications, and (social) resource-sharing systems and networks, such as *wikis*[11], *weblogs*[12], *Twitter*[13], *Facebook*[14], *del.icio.us*[15] or *Flickr*[16]. are associated with the current web (Cattuto et al., 2007). Obviously, these mainly user-centric publishing and knowledge-management platforms have acquired a large number of users in recent years. One of the main reasons for the success of these platforms is that there is no specific competence needed in order to participate. Built-in tools enable even non-technicians to contribute to the web and to engage in web-based social communities.

At first, *e-mail* and *newsgroup communication* had taken a major role in internet-based communication. Subsequently, the first online-based *bulletin boards* and *guest books* provided an additional platform where users could communicate via comment functions and forum posts. However, until then, the actual content production and publishing process was left to commercial internet portals or online newspaper publishers. With the emergence of *weblogs*, the first independent content publishers became present on the web. At that time, the term *user-generated content* was coined. Nowadays, telephoning, video messaging, and other services are also attributed to a *web-based communication* which, as we may argue, formed a second functional layer of interpersonal communication. Finally, with the rising popularity of social networks such as *Facebook* and *Twitter*, users not only became able to publish content on the web but also to aggregate and conjoin different web-based communication techniques with each other. More precisely, we can identify an amalgamation of different communication channels and online services now. Text messages sent by cell phones via the online service *Twitter* are indexed and searchable by *Google*; videos published by *YouTube* are interlinked to *weblogs*, rated and commented on *Facebook* and shown on *television* at home. Newly emerging web services within the Semantic Web support the connectivity of data by enabling the exchange of addresses, events, or social relationships (Brickley and Miller, 2010). That is, information (e.g. comments, images, or videos) which is published on one web site by a single user which may uses different technical devices (e.g. telephone, TV, or computer) may promptly appear in many other places (and hence in databases) on the web as well. This phenomenon relates to some crucial aspects of the Semantic Web (e.g. partial information – the web of data) but also raises privacy issues and the role of remembering in a *web that never forgets* (Mayer-Schonberger, 2009).

3.1. E-mail

By far the most common and oldest form of internet-based communication is *electronic mail* (e-mail).[17] In general, e-mail refers to the transmission of messages over communication networks. Messages and attachments (e.g. files and documents) can be sent using an offline e-mail client (e.g. *Thunderbird*) or an online web service (e.g. *Gmail*) through an internet gateway as provided by an *Internet Service Provider*. Today, sending e-mails is still a fast and effective way to synchronize real-time communications (Schuff et al., 2007). However, with the increase of e-mail users (more users sending more e-mails), information overload has become a significant problem in recent years. For example, it led to an increase of so-called spam e-mails (Youn and McLeod, 2007). Spam refers to an electronic advertising mail (junk) which is sent (automatically) to thousands of people. Researchers interested in e-mail have addressed these concerns with classification systems focusing on spam protection (Hayati and Potdar, 2008), relevance rating, and summarization (Aery and Chakravarthy, 2005; Xiang et al., 2007) as well as information extraction and disambiguation (Minkov et al., 2006) (see also Paaß (2012) in this volume). Finally, these systems also address e-mail-act categorization (e.g. request, commitment) (Carvalho and Cohen, 2005). Most recently, the search engine company *Google* presented a web service called *Google Wave*[18] which aims to amalgamate the traditional (asynchronous) e-mail communication with collaborative elements of the Web 2.0 in real-time (e.g. instant messaging, wikis, and weblogs) within one enhanced e-mail application. This communication and collaboration tool enables users to exchange messages (so-called *waves*) as they would do it in e-mail or instant messaging conversations. The *wave* can contain text, comments, pictures, or embedded videos (e.g. *YouTube*, *Google Maps*). All *wave* collaborators work from a single real time copy. All replies and edits are visible in real-time for the collaborators (e.g. *letter-by-letter*). The *wave* represents in this context both a conversation and a document, it is equitable treated. This service was one of the first attempts that combined e-mail with instant messaging, documents with conversations, and allowed the web-based collaboration and communication within one *wave* object.

3.2. Instant messaging and video conferencing

Instant Messaging (IM) services are still one of the most common and fastest-growing forms of web-based communication. IM enables users to communicate (chat) with other users (chat groups) via text messages in real-time. Messages are sent via the network using a (standardized) messaging protocol such as *XMPP* (Extensible Messaging and Presence Protocol) or *OSCAR* (Open System for Communication in Real-time). In general, users need to have a chat client installed on a device (e.g. computer or cell phone) which is connected to a server via a network. Most of the IM-clients offer a function which alerts the user whenever a recognized contact within the user's buddy list (see Section 3.3.7) is online. One distinguishes between three main forms of instant/chat messaging forms (Storrer, 2001):

1. Traditional IM services are usually not listed in a public domain (e.g. chat rooms) and use the client-server or a peer-to-peer model for the communication (in terms of data transmission). That is, participants need to have the client software installed on their computer.
2. *Internet Relay Chat* (IRC) also requires its own chat servers, which are usually connected through a network. For the communication within the IRC client specific IRC commands are used.
3. Recently, web-based chat services have gained popularity, which enable communication/chat directly using the web browser as client. In contrast to IM or IRC, these services do not require any additional software but are mostly limited to a particular website (e.g. *GoogleTalk*, *Facebook*).

In this context, IM and chat services effectively combine features of the telephone (one-to-one near-synchronous interaction), e-mail (text-based communication), and chat-rooms (one-to-many communication). Altogether, these features contribute in making IM inexpensive compared to other forms of media such as the telephone. Due to its convenience and usefulness and as a result to an increase in internet connection time, IM has also become popular in the business world (Glass and Li, 2010). Popular IM clients and services include *QQ*, *ICQ*, *AOL Instant Messenger*, *Microsoft MSN Messenger*, and IBM's *Lotus Instant Messaging*.

While IM enables users to communicate by means of text-messages, *Video Conferencing* (VC) enables two or more participants to communicate via audio and video data at the same time. That is, VC provides a visual interaction that cannot be achieved by the traditional IM or e-mail communication. In addition, most of the current VC clients integrate services for instant messaging (such as, for example, *Windows Live Messenger*). Research on IM and VC focuses on usability issues by means of message composition (by analyzing, for example, keystroke data – see Campbell 2004), the usage of IM for personal and business purposes with respect to conversational characteristics and functions of IM (Isaacs et al., 2002), on VC-based traffic classification (Angevine and Zincir-Heywood, 2008) and on linguistic characteristics of chat communication (Storrer, 2001; Beißwenger, 2007).

3.3. Internet telephony and VoIP

Another internet-based communication technique that has gained popularity recently is internet telephony, also known as *Voice-over-IP* (VoIP). VoIP refers to a method of transmitting voice data in packets over the internet (IP network) rather than by traditional circuit-switched telephone networks (Public Switched Telephone Network). Thereby, the voice data is converted into data packets, sent through the network, and converted back into voice data at the corresponding destination. If a telephone is used as a communication device, it needs to be connected to a VoIP phone adapter that in turn is connected to the network. Similar to an internet service provider, this technique is enabled by a VoIP provider such as, for instance, *Skype* or *Sipgate*[19]. In general, VoIP is less expensive than traditional voice communication. Hence, gratis VoIP services are already available in terms of device-independent lightweight applications (so-called *apps*, (see Section 2.1)). According to this and by means of an appropriate device, it is possible to chat with a relative living 1,000 miles away at no cost while being on the playground with one's children. In connection with VC, one can simultaneously present some livestreams from the playground to the relative to show how the children have grown up. As a result of the development of latest Web 2.0 techniques, such scenarios are commonplace today. However, the quality of the conversation via VoIP still depends on the available internet bandwidth (Bhanu et al., 2010). Compared to the conventional telephony, VoIP shows several advantages for both providers and users. In terms of the provider's part, internet telephony offers more developmental freedom as it is provided as a service, independent from old protocols or hardware. Thus, VoIP providers have more opportunities to exploit the new technology to fulfill the users' needs (Tapio, 2005). In terms of the users, the search for an adequate provider is not longer limited to those which provide their services locally and to those having roaming agreements with local operators in other areas. Another advantage for the users of quite a different nature is the increased control regarding their availability for conversations. In *Skype*, for example, users are able to specify their availability by means of icons appearing in the friend's buddy list (see Section 3.3.7). These icons symbolize if the corresponding contact is working, absent, etc., or if he/she is available. Despite of this information, many users additionally check for the availability of their contacts by sending them a message via IM as most VoIP applications involve this function as well. As a consequence, unannounced calls are increasingly regarded as annoyances (Joisten, 2007). Hence, interaction rules of IM and telephony have converged and generated a new form of multimedia-based communication application.

3.4. Weblogs

In recent years, weblogs have become rapidly popular as a new and easily accessible tool for *Personal Publishing* that is suitable even for non-technical users. In general, weblogs or *blogs* refer to (personal) websites where users publish (short) articles (sometimes referred to as *microcontent*) about their lives, opinions, or about technical issues (Blood, 2002, 6). It can be seen as a form of a web-based diary that is publicly available. The authors of weblogs (so-called *bloggers*) primarily use weblog-publishing-systems (e.g. *WordPress*[20]) or *Content Management Systems* (CMS)[21] to publish their content (*user-generated content*) (Vakali and Pallis, 2003; Stocker and Tochtermann, 2009). These systems provide the technical basis for the publishing process. At the beginning of the Web 2.0, weblogs gained much popularity by the online community, since these systems were free of charge, interactive, and easy-to-use. That is, there is no technical or web-design knowledge needed to setup a weblog system. Most weblog systems allow other users to post comments to the respective articles. Related weblogs, online friends, or bookmarks are linked together through the so-called *blogroll* section within the website. From a technical point of view, weblogs introduced several interactive elements which quickly helped them to become popular:

1. *Permalinks* refer to a permanent accessible URL linked to an item of microcontent (e.g. a single article or video) within the weblog. That is, single contributions are directly and permanently accessible regardless of where the content is located within the weblog.

2. *Trackbacks* or *Pingbacks* (SixApart, 2002) enable bloggers to determine whether another weblog refers to one of their own weblog entries. This feature sends (mostly automatically) a network signal (a so-called *ping*) to a specific URL of the target weblog, which in turn creates a backlink to the initial website. In this way, the weblog software is notified of new referenced weblog entries.
3. *RSS-Feeds* provide (possibly summarized) text including metadata such as publishing date, authorship, and backlink of the most recent entries of a weblog in a machine-readable XML format (e.g. based on *Really Simple Syndication 2.0* or *Atom Syndication Format*). The majority of modern weblog software systems automatically provide RSS for content syndication.

In the past, all these characteristics contributed to the creation of a virtual and interactive *weblog network*, which is sometimes called the *blogosphere* (Stocker and Tochtermann, 2009). In this context, weblogs have influenced the nature of web-based communication. They can be regarded as less objective due to their emotional communication style (Back et al., 2008, 18), have a greater diversity of topics covered, and enable users to engage in conversations on most recent issues almost in real-time. Consequently, the emergence of such user-generated content offers a wide range of applications not only to computer science (e.g. sentiment analysis (Melville et al., 2009), web genre classification (Santini et al., 2010) and topic detection (Sun et al., 2007)) but also for political science (Drezner and Farrell, 2008; Farrell and Drezner, 2008) and social psychology (Herring et al., 2004; Miura and Yamashita, 2007). It allows researchers to analyze what people think of and why (Kolari et al., 2006).

3.5. Folksonomy

Composed of the terms “folks” and “taxonomy”, *folksonomies* (Wal, 2005) provide systems for the collaborative categorization of web content by allowing anyone, especially consumers, to annotate this content with arbitrary tags (Alby, 2008). They are used for tagging web pages (e.g. *del.icio.us*), annotating pictures (e.g. *Flickr*) or classifying scholarly publications (e.g. *bibsonomy*[22]) (Damme et al., 2007). Thereby, different kinds of tags can be specified, for instance tags that identify what (or who) the content is about, what is presented or who owns the content (Golder and Huberman, 2006). Thus, tagging features reflect subjective assignments between words and categories of objects, intersubjective patterns in these associations and implicit information on social networks (Damme et al., 2007). The advantage of folksonomies, compared with traditional taxonomies, is an intuitive, user-based annotation of data without predetermined concepts which allows a faster and simpler usage (provided by, e.g. *del.icio.us*). Typically, these sites enable users to publicly tag and share content, that is, they do not only categorize information for themselves and they can also browse the content categorized by other users (Golder and Huberman, 2006).

Besides web sites that pursue this 'social' aspect of collaborative tagging, approaches aiming for commercial purposes only exist. *Google*, for instance, published a tagging game (*Google Image Labeler*[23]) where two people see the same image and label this image with as many tags as possible. To receive points (the goal of the game) both participants need to enter the same label. The more specific this label is, the more points can be awarded. By this, the quality of Google's image search results is improved by means of collective intelligence (see Section 2.1.2).

3.6. WikiWikiWeb

The *WikiWikiWeb* or *wikis* refer to the domain of *open collaboration hypertext systems*. Even though the first version of a wiki was published in 1995 (Leuf and Cunningham, 2001), wikis became first popular as a part of the Web 2.0 in 2001, namely by means of the rise of the online encyclopedia *Wikipedia*. The basic concept of a wiki is that users are allowed to not only read, write, publish, and contribute articles, but also to edit and (re-)structure existing documents of other contributors within the same wiki. That is, the content of a wiki is created and revised by a multitude of users in a collaborative manner. This principle follows also the notion of what Berners-Lee described as the *read-and-write web* (Berners-Lee and Fischetti, 2000). A well-known and widely used open-source wiki software is *MediaWiki* (e.g. used by the *Wikipedia* project[24]). Like traditional content management systems, the *MediaWiki* software allows even technically non-experienced users to publish content (e.g. texts and pictures) and to contribute to the corresponding community of *wikilocutors* in this way. That is, users do not need to have a sophisticated knowledge about markup languages (e.g. HTML) but only about the simple syntax of the wiki in order to publish data. Every article within a wiki corresponds to a separate page (assigned by an URL). Furthermore, users can attach cross-references to other pages via hyperlinks to express some kind of relationship (e.g. specification or definition). In this context, wikis establish a mix of very general and very specific interests and hence integrate a highly collaborative but also self-organizational property within the community (de Kerckhove et al., 2008). The enormous amount of documents available of the *Wikipedia* project has also drawn much attention within the scientific research community in the last years. As for instance, with regards to the quality of Wikipedia articles (Hammwöhner, 2007b), the relationship of Wikipedia categories (Chernov et al., 2006; Gleim et al., 2007; Waltinger and Mehler, 2009), its graph and hyperlink topology (Voss, 2005; Krötzsch et al., 2005; Nakayama et al., 2008; Mehler, 2006; Milne and Witten, 2008; Mehler, 2011), in the context of domain specific thesauri (Milne et al., 2006; Zesch et al., 2008) or semantic relatedness (Strube and Paolo-Ponzetto, 2006; Gabrilovich and Markovitch, 2007; Waltinger et al., 2009) In recent years, *Semantic Wikis* (Schaffert et al., 2008) – as an extension of the wiki idea – have drawn some attention in the research community. These software tools aim to combine the strengths of the Semantic Web (data integration and the realization of machine processing and complex search queries) and the *MediaWiki* technologies (easy-to-use, interactive, and emphasizing collaboration). That is, users can not only contribute web content but are also able to define and annotate entities and concepts within contributed articles. These annotations (possibly based on RDF – see Section 2.2) may refer to predefined or newly created types of relations in ontologies. For instance, the relation “has birthplace” may be used to connect a proper name with an article about a geographic location. The collaboratively organized annotation of metadata facilitates the automatic content analysis by enabling, for example, complex search queries using the *SPARQL* query language for RDF (as done, for example, in the *DBpedia* project – see Auer et al. 2007). In this context, *Semantic Wikis* can not only be used to improve the browsing and search experience (e.g. faceted browsing), but may also enable non-technical users to contribute structured information to the Semantic Web.

3.7. Social Networks

Today, the web is characterized by a variety of social networks such as *Facebook*[25], *Twitter*[26], *MySpace*[27], or *LinkedIn*[28] (see also Diewald (2012) in this volume). Unlike wikis or weblogs where the main objectives are the publishing process and the linkage of documents, social networks primarily focus on the self-portrayal of users and the establishment of social relations among them. That is, social networks allow people to not only present themselves on the web via user profiles, but also to explore, interact, communicate, and connect to other profiles (people) within the shared community. The members of a social network even interact and collaborate with people they did not know before, just by following acquaintance links starting from the profile of a neighboring acquaintance. However, most friends and communication partners are already known persons. Hence, social networks support the practice of *friendship-driven participation* (Ito et al., 2009). The basic principles within social networks are the following:

1. *User profiles* represent the nodes of a social network. Most network sites enable users to represent themselves by means of an online profile or portfolio that allows them to add and upload textual information (e.g. profession, gender or avocation), images, audio, and video files or documents.

2. *Buddy lists* represent the edges between the user profiles. That is, approved friends are added to a list of contacts or to a so-called buddy list. People finding themselves in these lists are allowed to post comments, view the user profile, and to forward the status of a friend to other users.
3. *Private messaging* (in terms of web-based e-mail messaging) is mostly enabled between approved online relationships only. Sophisticated online networks also provide communication (restricted to friends only) via instant messaging (chat) services.
4. *Bulletin boards* allow friends as well as the profile owner herself to leave marks in terms of textual messages, images, or videos that are readable and accessible for any other friend.
5. *Notifications* indicate whether a user has changed or added profile-related information (e.g. text messages, images or comments). Friends are often encouraged to post comments on the latest status update of an user. These status update features are sometimes referred to as *microblogging* features.
6. *Virtual groups* can be built by any member of the social network. That is, any user may create a group around a (common or shared) topic of interest and invite other members of the community to participate.

Communicating via bulletin boards has become very popular as it supports easy and fast spread of news, notifications, expressions of opinions, and public discussions. As a result, these boards obtain communicative power and social networks such as *Facebook* more and more become a counterpublics instrument (see the “facebook revolution” in Tunisia and Egypt (Hauslohner, 2011)). How easy and especially rapid publications are spread over social networks can be exemplified by the news of an unintentional invitation of 1,600 people to a party published worldwide: A student from Hamburg posted an invitation to her birthday party, which included her address, on *Facebook*, unfortunately without marking the event as “private”. As each user who responds to a public event automatically invites his friends as well, thousands of Facebook users felt invited. In the end, 1,600 of them crashed the girl’s birthday party (Bird, 2011). Most recently, a new form of social network has gained popularity within the online community. *Microblogging platforms* such as *Twitter* or *Tumblr*[29] differ from traditional weblogs and social networks in that they primarily focus on *status update features*. Here, published content (e.g. posts or tweets) is typically smaller than weblog articles as the published information only consists of a short sentence, comment, quotation, image, or video (e.g. the number of characters allowed at *Twitter* is 140). Addressed topics may range from small comments about recent issues (e.g. political statements, reviews of movies, or forwarded tweet messages, so called *retweets*) to what the user is currently doing or where he/she is now (e.g. geographical information) (Lenhart and Fox, 2009). However, studies of communication behaviors on *Twitter* show that the exchange of reports, episodes, and anecdotes about the everyday life, that is, the exchange of personal information, predominate (Java et al., 2007b). These platforms often serve as modern news media spreading information rapidly, easily, and globally. Thus, the news about the first nuclear meltdown in Fukushima and subsequent news about evacuations, explosions, etc. were notified promptly on *Twitter* and other microblogging platforms, while this information was published in the daily newspaper two days later. The data provided by social networks has opened up new areas for research including political science (Mustafaraj and Metaxas, 2010), usage and intention (Bohringer and Barnes, 2009), user influence (Cha et al., 2010), microblogging spam (Lee et al., 2010), and opinion mining (O'Connor et al., 2010) (for text mining see Paaß (2012) in this volume). This also includes research on topological and geographical properties of social networks (Java et al., 2007a; Backstrom et al., 2010).

4. Challenges of web-based communication

With the sustained technical advances of web-based communication technologies and the rising number of online communities and participants, the WWW is gaining a growing importance for studies on computer-mediated communication (e.g. user-to-user) and human-computer interaction (e.g. computer (agents)-to-user) (Thurlow et al., 2004). Due to the convergence of emergent communication and interaction technologies (in terms of web-based communication via conventional technologies), a wide range of (linguistically motivated) heterogeneous data is associated with the web (Hancock et al., 2005). Modern communication technologies have enhanced the quantity of written or spoken data that manifests rich sources of collaboratively organized knowledge (of a widespread thematic diversity regarding subject-specific as well as common-sense knowledge). This data can be explored to improve information retrieval systems (Waltinger and Mehler, 2009; Mehler and Waltinger, 2009), but also to speed up information processing among users (Wilson et al., 2009; Canali et al., 2011). That is, newly emerging issues and events are published, spread, communicated, and also critically rated almost in real-time (e.g. think of presidential elections or news of natural disasters communicated via *Twitter*).

Today, the web can not only be seen as a mirror of language use within different linguistic and social communities, but also as a communication medium in which language use is shaped (Erbach, 2004). It enables researchers to explore the social cues on how people use and interact with their computers, with each other (in terms of network relations which are characterized by content and direction), or within groups (in terms of network ties between networked users) (Garton et al., 1997). Within this *Web of Data* (Berners-Lee et al., 2001), hyperlinks are more reflective of communication pathways between communities than of traditional communication channels. In this sense, a hyperlink serves as an *indicator* (Park and Thelwall, 2003) of a *communication relationship* (Erbach, 2004). In addition, the web is currently changing from a web of *hyperlinked documents* to a web of *typed hyperlinked data* (Ding et al., 2008). Web-based communication as well as interaction services and applications such as *semantic wikis* (Schaffert et al., 2008), *social bookmarking* (Hotho et al., 2006) (in terms of *social tagging* (Mika, 2004; Staab et al., 2005)), and traditional wikis (in terms of *social ontologies* (Cattuto et al., 2007; Mehler, 2008, 2011)) contributed significantly to this phenomenon. However, due to the convergence of different computer-mediated techniques, the domain of web-based communication now also incorporates a broad spectrum of heterogeneous sources and data formats, which pose new challenges for the research community. That is, tools and methods developed for the automatic annotation of linguistic data can not be directly applied to computer-mediated data (e.g. chat annotations comprise non-standard spellings, conceptually oral forms, or written dialects) without being adapted (Beißwenger and Storrer, 2010). Therefore, the linguistic analysis of web-based communication must consider that language use within the web (say *netspeak*) differs from traditional linguistic data (e.g. *TIGER Treebank* (Brants et al., 2002), *Reuters Corpus* (Rose et al., 2002)) so that the tools and linguistic services must be extended to new tasks (Beißwenger and Storrer, 2010).

Another challenge is the control of privacy within a social web that makes topics and information of personal relevance publicly available (Schmidt, 2009). For instance, the access to information about persons situated in different social contexts causes role conflicts as several self-portrayals arise from a role-specific context but are accessible for persons possibly belonging to other reference groups (*unintended public* (Schmidt, 2009)). According to this, many employees do not want to be associated to their private self-portrayals within their job-related context. In reality, many recruiter search social networks for profiles of job candidates as expecting to figure out their true personality. Furthermore, internet users have to face the problem that other persons provide photos or information on the web regardless of whether they agree to this or not. Social psychological studies have shown inconsistencies regarding the relation between attitudes towards privacy and the behavior within the web. Thus, people having little interest in disclosing information do often not attach great importance to the privacy protection of others – also known as *voyeurism* (Schmidt, 2009). One way to retain the control of privacy is to present oneself variously for different publics (e.g. by creating self-portrayals on occupational networks like *XING*[30] and on a platform designed for personal, recreational interactions). Additionally, many platforms offer further mechanisms for privacy control by means of password protection or restricted access. However, especially due to the mentioned attributes of personal publics (persistence, replicability, scalability, searchability; see Section 2.1) versant borders between the public and privacy perish and need to be renegotiated in terms of the current web.

5. Conclusion

Today, the area of internet-based communication encompasses a broad range of media services and technologies. Unlike previously, we can now identify an amalgamation of different communication systems and applications. Communication technologies have been transformed from being highly technical and specialized to being easy-to-use and personal – an integral part of our daily life. The developed technologies from the area of the Web 2.0 and the Semantic Web have made a significant contribution to these advances. With the tremendous growth of published data, exchanged and communicated via the web, new challenges for computer scientists arise. While the content within the internet-based communication can be encoded and decoded using a variety of types (in terms of media compilation and annotation), special methods and services need to be developed or adapted to be applied. Internet-based communication provides a challenging area to explore computational models of natural language texts and human communication on networked communication systems, promoting new research in computational linguistics and computer science.

Notes

1. http://www.viola.org/
2. http://mosaic.mcom.com/
3. http://browser.netscape.com/
4. http://altavista.digital.com
5. http://yahoo.com
6. http://www.google.de/
7. http://www.facebook.com/press/info.php?statistics
8. http://www.guardian.co.uk/technology/2011/jan/03/facebook-value-50bn-goldman-sachs-investment
9. http://www.cyc.com/opencyc/
10. http://www.linguistics-ontology.org/gold.html
11. http://www.wikimedia.org/
12. https://www.blogger.com/
13. https://www.twitter.com/
14. https://www.facebook.com/
15. http://delicious.com/
16. http://www.flickr.com/
17. The first text message was sent by Ray Tomlinson in 1971 via *ARPANET.*
18. https://wave.google.com/wave/

19. http://www.sipgate.de
20. http://wordpress.org/
21. A content management system is a system used to manage and edit content from a web-site without needing the expertise of the Hypertext Markup Language (HTML).
22. http://www.bibsonomy.org/
23. http://images.google.com/imagelabeler/
24. http://www.mediawiki.org/wiki/MediaWiki
25. http://www.facebook.com
26. http://www.twitter.com
27. http://www.myspace.com
28. http://www.linkedin.com
29. http://www.tumblr.com
30. http://www.xing.com/

References

Abbate, J.
2000 *Inventing the Internet*. MIT Press, Cambridge, Massachusetts.

Aery, Manu and Sharma Chakravarthy
2005 eMailSift: Email classification based on structure and content. In *ICDM '05: Proceedings of the Fifth IEEE International Conference on Data Mining*, pages 18–25, Washington, DC, USA. IEEE Computer Society.

Aitken, J. S., B. L. Webber and J. B. L. Bard
2004 Part-Of Relations in Anatomy Ontologies: A Proposal for RDFS and OWL Formalisations. *Pacific Symposium on Biocomputing*, 8:166–177.

Alby, Tom (ed.)
2008 *Web 2.0: Konzepte, Anwendungen, Technologien*. Carl Hanser, 3. edition.

Angevine, Duffy and A. Nur Zincir-Heywood
2008 A preliminary investigation of skype traffic classification using a minimalist feature set. In *Proceedings of the The Third International Conference on Availability, Reliability and Security, ARES 2008, March 4–7, 2008, Technical University of Catalonia, Barcelona*, pages 1075–1079.

Antoniou, Grigoris and Frank van Harmelen
2004 *A Semantic Web Primer*. MIT Press.

Armbrust, Michael, Armando Fox, Rean Griffith, Anthony D. Joseph, Randy H. Katz, Andrew Konwinski, Gunho Lee, David A. Patterson, Ariel Rabkin, Ion Stoica and Matei Zaharia
2009 Above the clouds: A berkeley view of cloud computing. Technical Report UCB/EECS-2009-28, Electrical Engineering and Computer Sciences, University of California at Berkeley. URL http://www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-28.html.

Auer, Soren, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak and Zachary Ives
2007 DBpedia: a nucleus for a web of open data. In *Proceedings of the 6th International and the 2nd Asian Semantic Web Conference*, ISWC'07/ASWC'07, pages 722–735, Berlin, Heidelberg. Springer.

Back, Andrea, Norbert Gronau and Klaus Tochtermann
2008 *Web 2.0 in der Unternehmenspraxis: Grundlagen, Fallstudien und Trends zum Einsatz von Social Software*. Oldenbourg, München.
Backstrom, Lars, Eric Sun and Cameron Marlow
2010 Find me if you can: improving geographical prediction with social and spatial proximity. In *Proceedings of the 19th international conference on World wide web*, WWW '10, pages 61–70, New York, NY, USA. ACM.
Baun, Christian, Marcel Kunze, Jens Nimis and Stefan Tai (eds.)
2011 *Cloud Computing – Web-basierte dynamische IT-Services*. Informatik im Fokus. Springer.
Beckett, Dave
2004 RDF/XML Syntax Specification (Revised). W3C Recommendation, W3C. URL http://www.w3.org/TR/rdf-syntax-grammar/.
Beißwenger, Michael
2007 *Sprachhandlungskoordination in der Chat-Kommunikation*, volume 26 of *Linguistik – Impulse & Tendenzen*. De Gruyter, Berlin/New York.
Beißwenger, Michael and Angelika Storrer
2010 Corpora of computer-mediated communication. In Anke Lüdeling and Merja Kytö, editors, *Corpus Linguistics. An International Handbook of the Science of Language and Society*, pages 292–308, Berlin/Boston. De Gruyter.
Berners-Lee, Tim
1994 Universal Resource Identifiers in WWW. Request for Comments 1630, Network Working Group. URL http://www.ietf.org/rfc/rfc1630.txt.
Berners-Lee, Tim
2011 Semantic web road map. http://www.w3.org/DesignIssues/Semantic.html, 1998. Retrieved on August 29.
Berners-Lee, Tim and Mark Fischetti
2000 *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by Its Inventor*. Harper-Collins, San Francisco.
Berners-Lee, Tim, James Hendler and Ora Lassila
2001 The semantic web. *Scientific American Magazine*, 284(5):34–43.
Bhanu, S. Vijay, R. M. Chandrasekaran and V. Balakrishnan
2010 Effective bandwidth utilization in IEEE802.11 for VOIP. *International Journal of Computer Science and Information Security*, 8(1):68–75.
Bird, Daniel
2011 Teenager's remorse over facebook house party. http://www.telegraph.co.uk/technology/facebook/8561253/Teenagers-remorse-over-Facebook-house-party.html, 2011. Retrieved on August 28.
Birkenbihl, Klaus
2006 Standards für das Semantic Web. In Tassilo Pellegrini and Andreas Blumauer, editors, *Semantic Web – Wege zur vernetzten Wissensgesellschaft*, pages 73–88. Springer-Verlag.
Blood, Rebecca
2002 *The Weblog Handbook: Practical Advice on Creating and Maintaining Your Blog*. Perseus Books, Cambridge MA.
Blumauer, Andreas and Tassilo Pellegrini (eds.)
2009 *Social Semantic Web: Web 2.0 – Was nun?* X.media.press. Springer, Berlin/New York.

Böhringer, Martin and Stuart J. Barnes
2009 Continuance usage intention in microblogging services: The case of twitter. In *Proceedings of the 17th European Conference on Information Systems (ECIS), Verona, Italy, June 08–10, 2009*, page in press.

Boyd, Danah
2008 *Taken out of Context. American Teen Sociality in Networked Publics.* PhD thesis, University of California. URL http://www.danah.org/papers/TakenOutOfContext.pdf.

Brants, Sabine, Stefanie Dipper, Silvia Hansen, Wolfgang Lezius and George Smith
2002 The TIGER treebank. In *Proceedings of the Workshop on Treebanks and Linguistic Theories*, pages 24–41.

Brickley, Dan and R.V. Guha
2004 RDF Vocabulary Description Language 1.0: RDF Schema. W3C Recommendation, W3C. URL http://www.w3.org/TR/rdf-schema/.

Brickley, Dan and Libby Miller
2010 FOAF vocabulary specification 0.98. http://xmlns.com/foaf/spec/, 2010. Retrieved on September 18.

Bush, Vannevar
1945 As we may think. *Atlantic Monthly*, 176(1):641–649.

Caldarelli, Guido
2007 *Scale-Free Networks: Complex Webs in Nature and Technology.* Oxford University Press, Oxford.

Campbell, Jeffrey D.
2004 Instant messages: a framework for reading between the lines. In *CSCW '04: Proceedings of the 2004 ACM conference on Computer supported cooperative work*, pages 519–522, New York, NY, USA. ACM.

Canali, Claudia, Michele Colajanni and Riccardo Lancellotti
2011 Data acquisition in social networks: Issues and proposals. In *Proceedings of the International Workshop on Services and Open Sources (SOS'11), Brest, France*.

Carvalho, Vitor R. and William W. Cohen
2005 On the collective classification of email "speech acts". In *SIGIR '05: Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 345–352, New York, NY, USA. ACM.

Cattuto, Ciro, Christoph Schmitz, Andrea Baldassarri, Vito Servedio, Vittorio Loreto, Andreas Hotho, Miranda Grahl and Gerd Stumme
2007 Network properties of folksonomies. *AI Communications*, 20(4):245–262.

Cha, Meeyoung, Hamed Haddadi, Fabricio Benevenuto and Krishna Gummadi
2010 Measuring user influence in twitter: The million follower fallacy. In *Proceedings of the International AAAI Conference on Weblogs and Social Media*.

Chernov, Sergey, Tereza Iofciu, Wolfgang Nejdl and Xuan Zhou
2006 Extracting semantic relationships between Wikipedia categories. In Max Völkel and Sebastian Schaffert, editors, *Proceedings of the First Workshop on Semantic Wikis – From Wiki To Semantics*, Workshop on Semantic Wikis. ESWC2006. URL http://semwiki.org/semwiki2006.

Chishti, M. Ahsan, Shaima Qureshi, A. H. Mir, Shariq Haseeb and Iftikhar Ahmad
2009 Utilizing semantic web as communication protocol in faded information field (FIF) architecture for information retrieval and dissemination. *The International Journal of Computing and ICT*, 3(2):41–46.

Damme, Céline Van, Martin Hepp and Katharina Siorpaes
2007 FolksOntology: An Integrated Approach for Turning Folksonomies into Ontologies. In *Proceedings of the ESWC 2007 Workshop Bridging the Gap between Semantic Web and Web 2.0*, pages 71–84, Innsbruck, Austria.

de Kerckhove, Derrick, Martina Leeker and Kerstin Schmidt
2008 *McLuhan neu lesen: Kritische Analysen zu Medien und Kultur im 21. Jahrhundert*. Transcript, Bielefeld.

December, John
1996 Units of analysis for internet communication. *Journal of Computer-mediated Communication*, 1(4).

December, John
2010 Notes on defining of computer-mediated communication. http://www.december.com/cmc/mag/1997/jan/december.html, 1997. Retrieved on September 18.

Diewald, Niels
2012 Decentralized social network sites. In Alexander Mehler, Laurent Romary, and Dafydd Gibbon, editors, *Handbook of Technical Communication*. De Gruyter, Berlin/Boston.

Ding, Ying, Ioan Toma, Sin-Jae Kang, Zhixiong Zhang and Michael Fried
2008 Mediating and analyzing social data. In *OTM '08: Proceedings of the OTM 2008 Confederated International Conferences, CoopIS, DOA, GADA, IS, and ODBASE 2008. Part II on On the Move to Meaningful Internet Systems*, pages 1355–1366, Berlin/Heidelberg. Springer.

Drezner, Daniel and Henry Farrell
2008 Introduction: Blogs, politics and power: a special issue of public choice. *Public Choice*, 134:1–13.

Dublin Core Metadata Initiative
2010 Dublin core metadata element set, version 1.1. Technical report, DCMI. URL url{http://dublincore.org/documents/2010/10/11/dces/}.

Erbach, Gregor
2004 Mapping, measuring, and modelling the diffusion of linguistic material on the internet. In *Proceedings of the First International Workshop on the Emergence and Evolution of Linguistic Communication*, Kanazawa.

Farrell, Henry and Daniel Drezner
2008 The power and politics of blogs. *Public Choice*, 134:15–30.

Gabrilovich, Evgeniy and Shaul Markovitch
2007 Computing semantic relatedness using Wikipedia-based explicit semantic analysis. *Proceedings of the 20th International Joint Conference on Artificial Intelligence*, pages 6–12.

Garrett, Jesse James
2005 Ajax: A new approach to web applications. http://adaptivepath.com/ideas/essays/archives/000385.php.

Garton, Laura, Caroline Haythornthwaite and Barry Wellman
1997 Studying online social networks. *Journal of Computer-Mediated Communications*, 3 (1).

Gates, Bill
2000 Shaping the internet age. http://www.microsoft.com/presspass/exec/billg/writing/shapingtheinternet.mspx. Retrieved on September 18, 2010.

Glass, Richard and Suhong Li
2010 Social influence and instant messaging adoption. *Journal of Computer Information Systems*, W2010:24–30.

Gleim, Rüdiger, Alexander Mehler, Matthias Dehmer and Olga Pustylnikov
2007 Aisles through the category forest – utilising the Wikipedia category system for corpus building in machine learning. In *Proceedings of the 3rd International Conference on Web Information Systems and Technologies (WEBIST '07), March 3–6, 2007, Barcelona*, pages 142–149.

Golder, Scott A. and Bernardo A. Huberman
2006 Usage patterns of collaborative tagging systems. *Journal of Information Science*, 32(2):198208.

Gruber, Thomas R.
1993 A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5:199–220.

Guha, R., Rob McCool and Eric Miller
2003 Semantic search. In *Proceedings of the 12th international conference on the World Wide Web (WWW '03)*, pages 700–709, New York, NY, USA. ACM.

Hammwöhner, Rainer
2007a Wikipedia – ein Medium der Ignoranz? In Achim Geisenhanslüke, editor, *Ignoranz: Nichtwissen, Vergessen und Missverstehen in Prozessen kultureller Transformationen*, volume 3 of *Literalität und Liminalität*, pages 229–258. transcript, Bielefeld.

Hammwöhner, Rainer
2007b Interlingual aspects of wikipedias quality. In *Proceedings of the 12th International Conference on Information Quality*, volume 1, page 3949.

Hancock, Jeffrey T., Lauren Curry, Saurabh Goorha and Michael Woodworth
2005 Automated linguistic analysis of deceptive and truthful synchronous computer-mediated communication. In *Proceedings of the Proceedings of the 38th Annual Hawaii International Conference on System Sciences (HICSS'05) – Track 1 – Volume 01*, HICSS '05, pages 22.3–, Washington, DC, USA. IEEE Computer Society.

Hauslohner, Abigail
2011 Is Egypt about to have a facebook revolution? http://www.nytimes.com/2010/07/25/magazine/25privacy-t2.html. Retrieved on August 28, 2011.

Hayati, Pedram and Vidyasagar Potdar
2008 Evaluation of spam detection and prevention frameworks for email and image spam: a state of art. In *Proceedings of the 10th International Conference on Information Integration and Web-based Applications and Services*, pages 520–527, New York, NY, USA. ACM.

Hayes, Patrick
2004 RDF semantics. http://www.w3.org/TR/2004/REC-rdf-mt-20040210/. Retrieved on September 18, 2010.

Hepp, Martin, Pieter De Leenheer, Aldo de Moor and York Sure (eds.)
2008 *Ontology Management, Semantic Web, Semantic Web Services, and Business Applications*, volume 7 of *Semantic Web And Beyond Computing for Human Experience*. Springer.

Herman, Ivan
2001 W3C semantic web activity. http://www.w3.org/2001/sw/. Retrieved on September 18, 2010.

Herring, Susan C., Inna Kouper, Lois Ann Scheidt and Elijah L. Wright
2004 Women and children last: The discursive construction of weblogs. In *Into the Blogosphere: Rhetoric, Community, and Culture of Weblogs*, page 2004.

Hitzler, Pascal, Markus Krötzsch and Sebastian Rudolph
Foundations of Semantic Web Technologies. Chapman & Hall/CRC, 2009.

Philipp Hoschka
2011 Ubiquitous web domain. http://www.w3.org/UbiWeb/. Retrieved on August 29, 2011.

Hotho, Andreas, Robert Jäschke, Christoph Schmitz and Gerd Stumme
2006 Bibsonomy: A social bookmark and publication sharing system. In *Proceedings of the Conceptual Structures Tool Interoperability Workshop at the 14th International Conference on Conceptual Structures*, pages 87–102, Aalborg. Aalborg University Press.

Howe, Jeff
2008 *Crowdsourcing: Why the Power of the Crowd Is Driving the Future of Business*. Crown Business, 1. edition.

Isaacs, Ellen, Alan Walendowski, Steve Whittaker, Diane J. Schiano and Candace Kamm
2002 The character, functions, and styles of instant messaging in the workplace. In *CSCW '02: Proceedings of the 2002 ACM conference on Computer supported cooperative work*, pages 11–20, New York, NY, USA. ACM.

Ito, Mizuko, Sonja Baumer and Matteo Bittanti
2009 *Hanging Out, Messing Around, and Geeking Out: Kids Living and Learning with New Media*. MIT Press.

Java, Akshay, Xiaodan Song, Tim Finin and Belle Tseng
2007a Why we twitter: Understanding microblogging usage and communities. In *Procedings of the Joint 9th WEBKDD and 1st SNA-KDD Workshop 2007*, College Park, Maryland. Computer Science and Electrical Engineering, University of Maryland.

Java, Akshay, Xiaodan Song, Tim Finin and Belle Tseng
2007b Why we twitter: Understanding microblogging usage and communities. In *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis*, pages 56–65.

Joisten, Martina
2007 Multimediale Gespräche in Skype: Hybridisierung von Gebrauchsweisen in der interpersonalen Kommunikation. In Simone Kimpeler, Michael Mangold, and Wolfgang Schweiger, editors, *Die digitale Herausforderung – Zehn Jahre Forschung zur computervermittelten Kommunikation*. Springer.

Jones, Steven G.
1998 Information, internet, and community in the information age: notes toward an understanding of community. In Steven G. Jones, editor, *Cybersociety 2.0*, pages 1–34. Sage Publications, Thousand Oaks, CA, USA.

Kalyanpur, Aditya, Evren Sirin, Bijan Parsia and James Hendler
2004 Hypermedia inspired ontology engineering environment: Swoop. In *Proc. International Semantic Web Conference (ISWC)*, pages 7–11.

Katz, James E. and Ronald E. Rice
2002 *Social Consequences of Internet Use: Access, Involvement, and Interaction*. The MIT Press, Cambridge.

Klyne, Graham and Jeremy Carroll
2004 Resource Description Framework (RDF): Concepts and abstract syntax. http://www.w3.org/TR/rdf-concepts/. Retrieved on September 18, 2010.

Knublauch, Holger, Ray W. Fergerson, Natalya F. Noy and Mark A. Musen
2004 The Protégé OWL plugin: An open development environment for semantic web applications. In *3rd International Semantic Web Conference (ISWC 2004)*, Hiroshima, Japan.

Kolari, Pranam, Tim Finin, Akshay Java, Anupam Joshi and Justin Martineau
2006 Blog track open task: Spam blog classification. In *TREC 2006 Blog Track Notebook*.

Kraut,R., M. Patterson, V. Lundmark, S. Kiesler, T. Mukopadhyay and W. Scherlis
1998 Internet paradox. A social technology that reduces social involvement and psychological well-being? *American Psychologist*, 53(9): 1017–1031.

Krötzsch, Markus, Denny Vrandecic and Max Völkel
2005 Wikipedia and the semantic web – the missing links. In *Proceedings of the WikiMania2005*.

Lassila, Ora and James A. Hendler
2007 Embracing "web 3.0". *IEEE Internet Computing*, 11(3):90–93.

Lee, Kyumin, James Caverlee and Steve Webb
2010 Uncovering social spammers: social honeypots + machine learning. In *Proceeding of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '10, pages 435–442, New York, NY, USA. ACM. URL http://doi.acm.org/10.1145/1835449.1835522.

Lenhart, Amanda and Susannah Fox
2009 Twitter and status updating. Technical report, Pew Internet and American Life Project. URL http://www.pewinternet.org/Reports/2009/Twitter-and-status-updating.aspx.

Leuf, B. and W. Cunningham
2001 *The Wiki Way: Quick Collaboration on the Web*. Addison-Wesley, Boston, pap/cdr edition.

Manola, Frank and Eric Miller
2004 RDF Primer, W3C Recommendation. http://www.w3.org/TR/rdf-primer/. Retrieved on September 18, 2010.

May, Wolfgang
2006 Reasoning im und für das Semantic Web. In Tassilo Pellegrini and Andreas Blumauer, editors, *Semantic Web – Wege zur vernetzten Wissensgesellschaft*, pages 485–503. Springer-Verlag.

Mayer-Schonberger, Viktor
2009 *Delete: The Virtue of Forgetting in the Digital Age*. Princeton University Press.

McGuinness, Deborah L. and Frank van Harmelen
2004 OWL Web Ontology Language overview. W3C recommendation, World Wide Web Consortium. URL url{http://www.w3.org/TR/owl-features}. Retrieved on September 18, 2010.

Mehler, Alexander
2006 Text linkage in the wiki medium – a comparative study. In Jussi Karlgren, editor, Proceedings of the EACL Workshop on New Text – Wikis and blogs and other dynamic text sources, April 3–7, 2006, Trento, Italy, pages 1–8.

Mehler, Alexander
2008 Structural similarities of complex networks: A computational model by example of wiki graphs. *Applied Artificial Intelligence*, 22 (7–8):619–683.

Mehler, Alexander
2011 A quantitative graph model of social ontologies by example of Wikipedia. In Matthias Dehmer, Frank Emmert-Streib, and Alexander Mehler, editors, *Towards an Information Theory of Complex Networks: Statistical Methods and Applications*. Birkhäuser, Boston/Basel.

Mehler, Alexander and Ulli Waltinger
2009 Enhancing document modeling by means of open topic models: Crossing the frontier of classification schemes in digital libraries by example of the DDC. *Library Hi Tech*, 27(4):520–539.

Melville, Prem, Wojciech Gryc and Richard D. Lawrence
2009 Sentiment analysis of blogs by combining lexical knowledge with text classification. In *KDD '09: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1275–1284, New York, NY, USA. ACM.

Mika, Peter
2004 Social networks and the semantic web. In *WI '04: Proceedings of the 2004 IEEE/WIC/ACM International Conference on Web Intelligence*, pages 285–291, Washington, DC, USA. IEEE Computer Society.

Miller, Eric
2001 Digital libraries and the semantic web. http://www.w3.org/2001/09/06-ecdl/. Retrieved on September 18, 2010.

Mills, David
2009 What is the role of cloud computing, web 2.0, and web 3.0 semantic technologies in an era of connected governance? In *Proceedings of the Semantic Community Workshop and Web Conference on Semantic Exchange*, Fairfax, VA.

Milne, David and Ian H. Witten
2008 Learning to link with Wikipedia. In *CIKM '08: Proceeding of the 17th ACM conference on Information and knowledge mining*, pages 509–518, New York, NY, USA. ACM.

Milne, David, Olena Medelyan and Ian H. Witten
2006 Mining domain-specific thesauri from Wikipedia: A case study. In *ACM International Conference on Web Intelligence (WI 2006 Main Conference Proceedings)(WI'06)*, pages 442–448.

Minkov, Einat, William W. Cohen and Andrew Y. Ng
2006 Contextual search and name disambiguation in email using graphs. In *SIGIR '06: Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 27–34, New York, NY, USA. ACM.

Miura, Asako and Kiyomi Yamashita
2007 Psychological and social influences on blog writing: An online survey of blog authors in Japan. *Journal of Computer-Mediated Communication*, 12(4):1452–1471.

Mulvenna, Maurice D., Sarabjot S. Anand and Alex G. Büchner
2000 Personalization on the net using web mining: introduction. *Communication of the ACM*, 43(8):122–125.

Mustafaraj, Eni and Panagiotis Metaxas
2010 From obscurity to prominence in minutes: Political speech and real-time search. In *WebSci10: Extending the Frontiers of Society On-Line*.

Nakayama, Kotaro, Takahiro Hara and Shojiro Nishio
2008 Wikipedia link structure and text mining for semantic relation extraction. In Stephan Bloehdorn, Marko Grobelnik, Peter Mika, and Duc Thanh Tran, editors, *Proceedings of the Workshop on Semantic Search (SemSearch 2008) at the 5th European Semantic Web Conference (ESWC 2008)*, volume 334 of *CEUR Workshop Proceedings*. CEUR-WS.org.

O'Connor, Brendan, Ramnath Balasubramanyan, Bryan R. Routledge and Noah A. Smith
2010 From tweets to polls: Linking text sentiment to public opinion time series. In *Proceedings of the International AAAI Conference on Weblogs and Social Media*.

O'Reilly, Tim
2005 O'reilly – what is web 2.0. http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html. Retrieved on September 18, 2010.

Paaß, Gerhard
2012 Document classification, information retrieval, text and web mining. In Alexander Mehler, Laurent Romary, and Dafydd Gibbon, editors, *Handbook of Technical Communication*. De Gruyter, Berlin/Boston.

Park, Han Woo and Mike Thelwall
2003 Hyperlink analyses of the World Wide Web: A review. *Journal of Computer-mediated Communication*, 8(4).

Patel-Schneider, Peter, Patrick Hayes and Ian Horrocks
2001 DAML+OIL reference description. http://www.w3.org/TR/2001/NOTE-daml+oil-reference-20011218. Retrieved on September 18, 2010.

Patel-Schneider, Peter, Patrick Hayes and Ian Horrocks
2004 OWL Web Ontology Language semantics and abstract syntax. http://www.w3.org/TR/owl-semantics/. Retrieved on September 18, 2010.

Prud'hommeaux, Eric and Andy Seaborne
2008 SPARQL query language for RDF. http://www.w3.org/TR/rdf-sparql-query/. Retrieved on September 18, 2010.

Raymond, Eric S.
2001 *The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary.* O'Reilly & Associates, Inc., Sebastopol, CA, USA.
Rose, T., M. Stevenson and M. Whitehead
2002 The Reuters Corpus Volume 1 – from yesterday's news to tomorrow's language resources. In *3rd International Conference on Language Resources and Evaluation*, pages 827–832.
Santini, Marina, Alexander Mehler and Serge Sharoff
2010 Riding the rough waves of genre on the web: Concepts and research questions. In Alexander Mehler, Serge Sharoff, and Marina Santini, editors, *Genres on the Web: Computational Models and Empirical Studies*, pages 3–32. Springer, Berlin/New York.
Sasaki, Felix and Arle Lommel
2012 Multilingual computing. In Alexander Mehler, Laurent Romary, and Dafydd Gibbon, editors, *Handbook of Technical Communication*. De Gruyter, Berlin/Boston.
Schaffert, Sebastian, François Bry, Joachim Baumeister and Malte Kiesel
2008 Semantic wikis. *IEEE Software*, 25(4):8–11.
Schmidt, Jan
2009 *Das neue Netz – Merkmale, Praktiken und Folgen des Web 2.0*. UVK Verlagsgesellschaft.
Schuff, David, Ozgur Turetken, John D'Arcy and David Croson
2007 Managing e-mail overload: Solutions and future challenges. *Computer*, 40:31–36.
Sharoff, Serge and Anthony Hartley
2012 Lexicography, terminology and ontologies. In Alexander Mehler, Laurent Romary, and Dafydd Gibbon, editors, *Handbook of Technical Communication*. De Gruyter, Berlin/Boston.
SixApart
2002 Trackback technical specification. http://www.sixapart.com/pronet/docs/trackback _spec. Retrieved on September 18, 2010.
Spivack, Nova
2007 Radar networks towards a webos. http://novaspivack.typepad.com/RadarNetworksTowardsAWebOS.jpg. Retrieved on June 1, 2011.
Staab, Steffen and Rudi Studer (eds.)
2004 *Handbook on Ontologies*. International Handbooks on Information Systems. Springer, Berlin.
Staab, Steffen, Pedro Domingos, Peter Mika, Jennifer Golbeck, Li Ding, Tim Finin, Anupam Joshi, Andrzej Nowak and Robin R. Vallacher
2005 Social networks applied. *IEEE Intelligent Systems*, 20(1):80–93.
Stocker, Alexander and Klaus Tochtermann
2009 Anwendungen und Technologien des Web 2.0: Ein Überblick. In Andreas Blumauer and Tassilo Pellegrini, editors, *Social Semantic Web: Web 2.0 – Was nun?*, X.media.press, pages 63–82. Springer, Berlin/New York.
Storrer, Angelika
2001 Sprachliche Besonderheiten getippter Gespräche. Sprecherwechsel und sprachliches Zeigen in der Chat-Kommunikation. In Michael Beißwenger, editor, *Chat-Kommunikation. Sprache, Interaktion, Sozialität & Identität in*

synchroner computervermittelter Kommunikation. Perspektiven auf ein interdisziplinäres Forschungsfeld, pages 3–24. Ibidem Verlag, Hannover.

Strube, M. and S. Paolo-Ponzetto
2006 WikiRelate! Computing semantic relatedness using Wikipedia. In *Proceedings of the 21st National Conference on Artificial Intelligence*, Boston, USA.

Stuckenschmidt, Heiner and Frank van Harmelen
2005 *Information Sharing on the Semantic Web*. Springer.

Stührenberg, Maik
2012 Foundations of markup languages. In Alexander Mehler, Laurent Romary, and Dafydd Gibbon, editors, *Handbook of Technical Communication*. De Gruyter, Berlin/Boston.

Stumme, Gerd, Andreas Hotho and Bettina Berendt
2011 Semantic web mining. State of the art and future directions. *Web Semantics: Science, Services and Agents on the World Wide Web*, 4(2):124–143.

Sun, Aixin, Maggy Anastasia Suryanto and Ying Liu
2007 Blog classification using tags: an empirical study. In *ICADL'07: Proceedings of the 10th international conference on Asian digital libraries*, pages 307–316, Berlin, Heidelberg. Springer.

Surowiecki, James
2004 *The Wisdom of Crowds*. Random House, New York.

Tapio, Antti
2005 Future of telecommunication – internet telephone operator skype. HUT T-110.551 Seminar on Internetworking.

Thurlow, Crispin, Laura Lengel and Alice Tomic
2004 *Computer Mediated Communication*. Sage Publications, London.

Trippel, Thorsten
2012 Controlled language structures in technical communication. In Alexander Mehler, Laurent Romary, and Dafydd Gibbon, editors, *Handbook of Technical Communication*. De Gruyter, Berlin/Boston.

Vakali, A. and G. Pallis
2003 Content delivery networks: status and trends. *Internet Computing, IEEE*, 7(6):68–74.

Vivacqua, Adriana S. and Marcos R.S. Borges
2010 Collective intelligence for the design of emergency response. In *Proceedings of the 2010 14th International Conference on Computer Supported Cooperative Work in Design*, pages 623–628.

Voss, Jakob
2005 Measuring Wikipedia. In *International Conference of the International Society for Scientometrics and Informetrics: 10th*. ISSI.

W3C
2009 W3C semantic web activity. http://www.w3.org/2001/sw/. Retrieved on September 18, 2010.

Wal, Thomas Vander
2005 Folksonomy. http://www.vanderwal.net/essays/051130/folksonomy.pdf. Retrieved on August 9, 2011.

Waltinger, Ulli and Alexander Mehler
2009 Social semantics and its evaluation by means of semantic relatedness and open topic models. In *Proceedings of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology – Volume 01*, WI-IAT '09, pages 42–49, Washington, DC, USA. IEEE Computer Society.

Waltinger, Ulli, Irene Cramer and Tonio Wandmacher
2009 From social networks to distributional properties: A comparative study on computing semantic relatedness. In *Proceedings of the Annual Meeting of the Cognitive Science Society – CogSci 2009, Amsterdam (NL), 2009*.

Walton, Christopher D.
2006 *Agency and the Semantic Web*. Oxford University Press.

Wilson, Christo, Bryce Boe, Alessandra Sala, Krishna P.N. Puttaswamy and Ben Y. Zhao
2009 User interactions in social networks and their implications. In *Proceedings of the 4th ACM European conference on Computer systems*, EuroSys '09, pages 205–218, New York, NY, USA. ACM. URL http://doi.acm.org/10.1145/1519065.1519089.

Xiang, Yang, Wanlei Zhou and Jinjun Chen
2007 Managing email overload with an automatic nonparametric clustering approach. In *Proceedings of IFIP International Conference on Network and Parallel Computing (NPC 2007)*, pages 81–90.

Youn, S. and D. McLeod
2007 A comparative study for email classification. In Khaled Elleithy, editor, *Advances and Innovations in Systems, Computing Sciences and Software Engineering*, pages 387–391. Springer, Dordrecht.

Zesch, Torsten, Christof Müller and Iryna Gurevych
2008 Extracting lexical semantic knowledge from Wikipedia and Wiktionary. In *Proceedings of the 6th International Conference on Language Resources and Evaluation*, Marrakech, Morocco.

17. Tutoring Systems and Computer-Assisted Language Learning (CALL)

Henning Lobin and Dietmar Rösler

1. Introduction[1]

The steady increase in access to computers in educational settings has enhanced the possibilities for learners and teachers to engage in meaningful interaction in foreign language classrooms and has changed the shape of linguistic material used for learning purposes in ways that are both productive and problematic. Given the relatively recent emergence of digital media, it is fascinating to observe the speed of the development of the use of digital media in foreign language learning both on the level of the interaction between users and computers (from early stand alone CALL exercises to sophisticated online learning environments and data driven learning) and on the level of computer mediated interaction between humans (from early email exchanges to sophisticated cooperative projects which make use of the diversity of available channels). The availability of digital media in foreign language learning has already altered fundamental aspects of foreign language teaching and learning (cf. Legutke and Rösler 2005) in the following ways:

- transcending the boundaries of the classroom: teachers and learners can now overcome traditional constraints of classroom learning because digital media provide them with various channels for interacting with speakers of the target language through e-mail, web-conferencing, chat etc.,
- augmenting the traditional textbook: learners, even if they live and learn in places far away from areas in which the target language is spoken, now for the first time ever have unlimited access to a wide range of written, audio or audiovisual target language resources. Above and beyond this they can even play an important role in co-creating a rich and meaningful learning environment by contributing texts and topics according to their needs and interests. Interactive material provides learners with immediate feedback, huge collections of linguistic data provide the opportunity for data-driven learning in a way not previously conceivable,
- empowering the learners: the chance for learners to co-create their learning environment greatly expands their role and shifts responsibility for successful learning to the so-called autonomous learners,
- re-defining teachers' roles: the shift of responsibility to individual learners and small cooperative groups and the use of project formats require a high

degree of flexibility on the part of the teacher because the learning process is far less predictable. It is therefore not surprising that the discussion of the role of digital media in foreign language learning has led to an intensification of discussions on the role of the teacher generally, and
– offering a greater variety of learning formats: both the new opportunities for learners and teachers to engage in meaningful interaction in foreign language classrooms and the changes in the shape of linguistic material used there has led to a new discussion of learning formats. New ways of focusing on individualized learning have been the result, as well as a questioning of teacher-centered methods and discussions about cooperative learning and project work in which learners take over teaching functions.

These changes will be discussed in the following pages, which will first offer a historical perspective, then go on to sketch the use of media in foreign language learning and teaching over the last century, and finally systematically describe current approaches to computer mediated communication, traditional and intelligent computer assisted language learning, and data-driven learning.

2. Historical development

Since the invention of audio recording, foreign language learners could be exposed to spoken language independently of the time when it was spoken; first recordings for foreign language teaching were produced as early as the beginning of the 20th century. With the availability of spoken data on sound storage media it was possible to add an acoustic dimension to textbooks and to provide learners with a linguistic model. For the first time, in theory at least, spoken language could also be learned in self-study with the help of recorded material, and it was also possible to introduce model native speakers and different dialects and sociolects into the classroom to counteract a one-sided orientation of learners towards the linguistic model of the teacher. Learners could imitate the native models and it was also possible, with recording devices, to record learners in their attempts to speak the target language, which enabled a comparison between the model's speech and the learners' realization (cf. Rösler 2010)

With the introduction of language laboratories at the beginning of the 1960s and the associated audio-lingual method, exercises were systematically realized in medial form (cf. Nübold 2006). The main exercises in the language labs were imitation exercises in the form of pattern drills. The undoubted advantages of language labs such as the individualization of practicing and the huge increase of speaking time per learner was accompanied by disadvantages such as the excessive demands on the learners in terms of self-correction, the rigidity of the teaching and learning arrangement and the exclusive focus on form.

From the 1970s on, parallel to the so-called communicative turn in foreign language teaching and learning, analogue media networks found their way into the classroom. Different analogue media were used simultaneously (cf. Legutke and Müller-Hartmann and Ulrich 2000, 53ff): cassettes, slides, overhead transparencies, films, television. This enabled access to different types of authentic texts, but the disadvantage of these analogue media networks lay, as with the language labs, in the rigidity of the teaching and learning arrangement; teachers felt demoted to media operators who simply had to ensure that predetermined sequences were technically executed (cf. Rösler 2007, 209).

The development of digital media has at least dispelled the technical problems of the analogue media networks. The availability of digital forms of communication and of multimedial material is, however, only a precondition for a new form of learning; they themselves don't actually generate it. Neither an unreflected concept of learner autonomy enabled by the use of digital media (cf. Rösler 1998; Hess 2006 and Schmenk 2008) nor a concept of media use which does not consider the actual interests of the learners (cf. Hess 2003) actually help learners of foreign languages.

The development of foreign language learning with digital media can be sketched as a line of development from simple exercises to complex multimedia learning environments and can be divided into two main areas: *Computer Mediated Communication* (CMC), in which digital media support and encourage interaction between learners, and *Computer Assisted Language Learning* (CALL), in which learning material in digitalized form is featured. Terminological confusion is caused by the fact that in some discursive contexts CALL is used as a superordinate for both and in others on the same level as and to demarcate it from CMC. Attempts to systematically define the scope of CMC and CALL have been undertaken by Harrington & Levy (2001) and others, the development and demarcation of different themes in CALL by Chapelle (2000) and Bax (2003) and others, and there are several descriptions of the historical development of computer-mediated learning (cf. e.g. Chapelle 2001: 1–26; Last 1987; or Levy 1997).

3. Computer mediated communication

Plenty of research has been done on CMC in the past decade (cf. e.g. the contributions in O'Dowd 2007 as well as Belz and Thorne 2006). Collaborative work in foreign language learning was not invented with the advent of digital media (cf. the survey in Legutke and Rösler 2005) but the issues discussed before their involvement in foreign language learning have now been enhanced by interesting new perspectives. There are cooperation projects between learners of the same foreign language from different countries, foreign language learners and native speakers of the target language, and also between foreign language learners and future foreign language teachers who work as tutors in cooperation projects (cf. Tamme 2001). In addition, the classical project of learning in tandem has become more widespread since digitalization.

3.1. Transcending the classroom: cooperative projects

In the same way as in classical cooperation projects where letters, audio or video cassettes were exchanged, questions of contents, general conditions of the institutions involved etc. also determine the success of digital cooperation projects; the availability of particular communication channels alone is not sufficient.

Difficulties related to the composition of groups and the development of social skills, challenging enough in face-to-face learning environments and in pre-digital cooperative cross-border projects in which both groups basically worked in their own educational contexts and 'merely' exchanged products, can multiply in digital cooperations which combine synchronous and asynchronous cooperation, because intercultural interferences as well as superficially banal constraints such as time-tables can be an additional hindrance to collaboration (on cultural aspects of technical communication see also Sasaki and Lommel 2012 in this volume). Issues such as group assessment and accountability can hit a dead end if the specific interpretations of these concepts differ significantly in the respective cultures of the learners (cf. Müller-Hartmann 2000; Belz and Müller-Hartmann 2003 as well as O'Dowd and Ritter 2006). The discussion of procedural and negotiatory language is traditionally not an essential element in textbooks and frequently not given enough focus in face-to-face cooperative work due to the implicit understanding that the shared frames of the face-to-face group make an elaborate analysis superfluous. It is given heightened attention in CMC collaboration, as complex negotiations through CMC demand a meta-communicative clearing of procedures and modes of negotiation, especially if these are predominantly in written form.

Collaboration via the internet offers a form of intercultural learning which, in the long term, can liberate cultural studies from the circular debate between exemplification and broad survey. Through personal interaction with a partner in one of the countries of the target language, it is possible for learners to experience, in a narrative context, the complexity of another part of the world (cf. Tamme 2001: 128–130).

3.2. One-to-one cooperation: learning in tandem

That it has become easier to initiate cooperation since the advent of digital media is shown by correspondence between classes, but especially by the tandem concept (cf. Brammerts and Kleppin 2001) which has developed an extensive exchange forum in the internet (see http://www.slf.ruhr-uni-bochum.de/). The basic constellation of the tandem has not been changed since its beginnings: people still communicate in mutually agreed forms, each of whom is an expert for their own language, there is still no classical teacher-learner relationship. But the form of exchange has changed, it now takes place through e-mail or via Skype, and possibly even in virtual worlds.

In classical tandem the participants were either in the same place or their exchanges via letters or cassettes took as long as the postal service decreed, the interaction was either synchronous or asynchronous. This ‘either-or’ has now been eliminated so that partners in different places can choose to communicate synchronously and/or asynchronously. In contrast to learning in the same place, however, there is a loss of the holistic dimension, as neither of the partner’s experiences the target language as a space in which they are staying. On the other hand, asynchronic communication such as tandems per e-mail has considerable advantages compared with synchronous communication especially for beginner learners: the partners can take as much time as they need to decode and produce the texts.

4. Computer-assisted language learning

CALL augments traditional textbooks by providing learners with access to a wide range of written, audio or audiovisual target language resources. This material is either authentic material not written for didactic purposes or else material specifically produced for learners making use of available technology. Material specifically produced for learners includes podcasts, photos and digital versions of printed material. Two parallel developments exist at the moment, the use of additional digital material to accompany teaching material which is still text book led, and the production of purely digital material. The main attraction of digital material specifically produced for learners lies in its potential interactivity: it can provide them with immediate feedback. The area of CALL most focused upon therefore is the area of tasks and exercises.

4.1. Digital textbooks

As far as the provision of additional digital material to textbooks is concerned, there are some obvious improvements:

– there is more room for individualization,
– certain elements of books traditionally not profitable and hence often neglected by publishers (teachers’ manuals, glossaries etc), can now be distributed in a more cost-effective and differentiated manner,

– up-to-date digital add-ons can compensate for the loss of topicality of printed textbooks,
– the space constraints of the traditional textbook which often lead to a shortage of visual and audio-visual information, differentiated sequences of exercises and so on, become less relevant when textbooks can be digitalized.

In the long-term this development can lead to the production of 'textbooks on demand', a new type of textbook which contains a common core element as well as a multitude of digital material tailor-made for the needs of specific learners. Such an approach would recast the traditional textbook as a type of database for which most material could be produced in a decentralized manner around a common core (cf. Rösler 2006).

Soon after the stand-alone exercises at the initial stages of CALL, digital components of textbooks came into being; there is hardly a foreign language textbook which appears today without digital material either in the form of CDs with exercises and tasks or with the provision of online activities. Beyond the closed exercises which focus on form, digital material as a component of textbooks can provide suggestions for projects, material for area studies etc. which are superior to the classical textbooks in terms of potential quantity (extensive collections of visual material, audio and video files) and in terms of topicality; but for this to be an advantage, they have to be produced at the same quality level as the classical printed textbooks.

4.2. Interactive exercises

With the advent of the computer came attempts to use it for exercises focusing on form. It was, to some degree, a repetition of the language lab concept (cf. Davies 1997); the learner could address a learning problem more efficiently individualized than in a group of learners. Due to the technology of programming and the authoring software these exercises were initially very limited: fill-in exercises, transformational exercises, drag-and-drop exercises, all of these closed exercises dominated the first round of computer-assisted language learning (cf. Rösler 2007, 104–109). From the point of view of research into teaching foreign language learning, this was a step back behind the diversity of form focused exercise already developed; one could speak, using the name of the most widespread authoring programme, of a *Hot-Potatoising* of the world of digital exercises.

4.3. Types and limitations of automatically generated feedback

A special problem of these exercises is the feedback provided for learners. In contrast to the classical language laboratory which gave the learner a correct answer which he or she could check and repeat, one of the greatest advantages of the computer is its interactivity: feedback, response to the learner's input, is possible. The programme has to respond to the learner's input in pre-programmed ways. This feedback is organised within the programme by matching patterns: the learner's input is compared to predetermined patterns, according to the pattern recognized a designated answer follows (cf. Mitschian 1998, 606). In the easiest case the programme generates the message "wrong answer" when the learner does not provide exactly the input anticipated. Whether he or she made a typing error, confused upper with lower case letters or whether it was a serious violation of the rules of the morphology of the target language would not be registered in this case. This type of programming is obviously not didactically satisfactory.

How good or bad automatically generated feedback is depends on how extensive and careful the programmers anticipation is (cf. Nagata 1993, Bangs 2003, Rösler 2007: 177–194). CALL exercises do not have to be as bad as many of those which can be found in the internet: at least with closed exercises the learner's input can be anticipated up to a certain point, careless mistakes can be distinguished from errors in competence and they can receive different feedbacks. The feedbacks can either stimulate self-correction or can provide the correction by referring to sources etc. The most important aspect for ensuring the quality of feedback is the question of how much time and energy was invested in this part of the software which is usually invisible for its users.

With open exercises and tasks, by which the learners have to type their own texts, the programmed analysis fails; all that can be given here as automatically generated feedback are sample solutions, or the texts are sent on to a human corrector who is available online. A new level of quality of feedback can only be achieved if artificial intelligence is introduced into the game and CALL becomes ICALL. This will be discussed later in chapter 6.

5. Data-driven learning and teaching

5.1. Learning with corpora

The advent of corpus linguistics introduced new possibilities into foreign language learning (cf. Mukherjee 2008, Beißwenger and Storrer 2009, Granger 2009, Chambers 2010). Its influence appears on two levels. a) Data generated and sorted by corpus linguists enable new insights into the target language for textbook authors and provides them with authentic material for their textbooks. b) Corpus linguistics can be used in the learning process by allowing data-driven learning to play a greater part. While a) obviously increases the potential of foreign language learning, b) is only a useful addition if it is used for the appropriate target groups and learning objects. In the early 1990s Johns and King developed corpus-linguistic learning material to motivate learners to contemplate and describe language independently with the help of software and computers. They defined this type of learning as *Data-Driven Learning* (DDL): "[Data-driven learning is] the use in the classroom of computer-generated concordances to get students to explore regularities of patterning in the target language, and the development of activities and exercises based on concordance output." (Johns and Kings 1991: iii).

The DDL approach aims to introduce learners to the new language they learn by means of corpora so that they are able to discover independently rules and structures of language. This type of learning is opposed to the traditional concept of rule-driven learning, which means that a set of rules is given for a language and variations are not tolerated. Examples are only used to show how grammatical rules work. Strict following of the rules is always the focus, and as a consequence exceptions to the rules are not explained. DDL, in contrast, encourages learners to explore the rules of language themselves. Exceptions to these rules help learners to see that language is not only based on strict rules but that its use always also involves making exceptions. Grammatical rules can always only serve as a benchmark to be followed but they can also be flexed or even broken and sometimes even have to be broken. This is what learners learn from DDL. Some studies, as for example those of Tognini-Bonelli (2001) have shown that this type of learning has a great influence on a learner's motivation.

Several corpus linguists have recognized the impact of DDL and have developed software tools which can be used for integrating corpora into teaching in a rather straightforward manner. A typical example is the *Compleat Lexical Tutor* (CLT, http://www.lextutor.ca/). CLT offers various functionalities such as vocabulary tests, analyses of mistakes and examinations of concordances which help learners of English, French and German to improve their language skills. Besides such web tools concordance programmes have been developed, e.g. the WordSmith Tools (http://www.lexically.net/wordsmith/), and have been designed in a more user-friendly way during the last years. As a rule of thumb it can be said that the more advanced learners are the more useful the application of DDL becomes. Advanced learners, for instance, can profit considerably from corpora while trying to engage with complex lexical phenomena such as collocations (cf. Ludewig 2005).

5.2. Learner corpora

A learner corpus is a huge computer-generated collection of written or oral texts produced by learners of a foreign language which can be automatically analyzed just like any other corpus. In contrast to first language corpora (cf. Lüdeling and Kytö 2008) learner corpora are set up in a more systematic way: mostly external criteria like first language background, foreign language competence, level of competence, learning setting, sex, age etc. determine the choice of texts and participants. Each of these external criteria can be analyzed individually or as a group, the analysis can thus yield results on the specific characteristics in foreign language learning. In the field of corpus linguistics, learner corpora represent a relatively new development. The first learner corpora were compiled in the late eighties. The most and the biggest learner corpora exist for English as a target language. In comparison, the number of freely accessible learner corpora for the German language is small (cf. Lüdeling et al. 2008).

Learner corpora give insights into non-native speakers' characteristics of language production. Thus, they are relevant for theories of second language acquisition and learning as well as for the development of teaching materials and instruction materials. Based on findings from learner corpora, exercises for individual groups of learner can be developed (cf. Seidlhofer 2002: 214).

In its initial stages learner corpora analysis was dominated by a descriptive linguistic view comparing language production of learners with corpus data of native speakers to identify characteristics of learner language. Typical questions asked by this approach are the following: Which kind of structures does the learner use too often (overuse), too rarely (underuse) or wrongly (misuse)? Are there any structures in the learner's language which do not find a similar use either in the structure of the target language or in the structure of the native language (learner-idiosyncratic structures)? In which area do learners tend to use an avoidance strategy, where do they not exploit the full potential of the target language?

However, new types of corpus projects make it possible to define core areas where learner with a specific first language (= L1) and learners with varying first languages (Lx) do not behave appropriately in the target language. The first one gives insights into the interferences caused by features of the learner's native language whereas the latter, by comparing interlanguage use of learners with different native languages, gives insights into the complexity of learning certain structures in the target language (cf. Granger 2002). Examples of this kind of corpora are the *International Corpus of Learner English* (ICLE: 50 Subcorpora, learner language with different native language, target language English, written text, cf. Nesselhauf 2004) and the *Louvain International Database of Spoken English Interlanguage* (LINDSEI: 50 Subcorpora, Learner language with different native languages, target language English, spoken texts).

The use of findings from learner corpora analysis is only slowly integrated into the development of teaching material. When comparing the learner language with different native language backgrounds it can be shown that a great number of mistakes is based on interference from L1 or other acquired languages (cf. Hufeisen 1990) and has no universal character. Thus one will find the same grammatical/lexical phenomenon of a foreign language strongly depending on the learner's native or previously acquired language(s). Based on this, individual teaching materials and instructional materials should be developed which focus on problems and move away from a prototypical notion of the learner (cf. Granger 1998, Seidlhofer 2002).

6. Intelligent computer assisted language learning

6.1. CALL and ICALL

The term *Computer-assisted Language Learning* (CALL) is not always used consistently throughout literature (see the end of chapter 2). It is often used as a generic term for traditional CALL-systems as well as for *Intelligent CALL* (ICALL) systems. According to Levy (1997) CALL could, in general, be defined as "the search for and the study of applications of the computer in language teaching and learning. […] The subject is interdisciplinary in nature and it has evolved out of early efforts to find ways of using the computer for teaching or for instructional purposes across a wide variety of subject areas, with the weight of knowledge and breadth of application in language learning ultimately resulting in a more specialized field of study." (Levy 1997: 1). Thus CALL can be described as a special concept of language learning.

Warschauer (1996) distinguishes three phases of CALL: Behaviour-related CALL, based on behaviour-related learning theory, Communicative CALL, based on communication-related learning theory, and Integrative CALL, which is based on two technological developments, namely multimedia and internet. Traditional applications are based on a behaviouristic learning theory. They involve only closed exercises, e.g. Multiple Choice Exercises, click exercises and texts with gaps. Those systems consist mainly of the following steps: First there is a short introduction to the topic at the beginning of each exercise where factual knowledge is given. Then the factual knowledge is supposed to be learned with drill-and-practice exercises. Here, the learner is given positive feedback when the answer is correct, but there, no feedback at all is given to wrong answers. Such drilling programs can be developed with little programming effort. While this is an advantage to the person developing such a program, it turns out to be monotonous and boring for the learner working with it. Besides CALL applications there are many ICALL-systems which have been designed according to behaviour-related principles and use only closed exercises.

In contrast to behaviourist exercises, detailed feedback plays an important role in cognitive theory which is supposed to help the learner to deduce rules independently, notice learning problems and to reflect on learning processes. Furthermore cognitive teaching material is supposed to support independent learning, for instance concerning the choice of the level of difficulty, the way of looking at a problem, and the learning speed. Cognitive theories are implemented by *Intelligent Tutoring Systems*. Here, the computer takes the role of a tutor. The systems are supposed to be intelligent because they arrange the learning material in a flexible way and because they adapt to the level of knowledge and cognitive performance of the learner (cf. Schulmeister 2002: 187). However, it has to be pointed out that the implementation of such application is rather complex, so that it often did not develop beyond a prototype. Furthermore most systems are a long way from adequately imitating cognitive characteristics of information processing.

In the constructive approach, the learning material is embedded in a simulated real-life interaction to enable learners to generate relevant strategies of dealing with a communicative problem by themselves. Here the teacher functions as a coach, who – according to the situation – explains a process or demonstrates a pattern of behaviour, introduces the learners to the problem, but finally fades out and accompanies the learner during the process.

6.2. Approaches to ICALL systems

The development of CALL has been strongly influenced by upcoming new technologies during the past few years. This has lead to a general criticism of ICALL projects being too technology driven at the expense of didactics. This criticism has been reinforced by classifications which reflect those technical aspects. Wolff (1993: 21) distinguished five groups of applications:

- Traditional CALL applications: programs of the initial phase of CALL based on behaviour-related learning theory.
- Multimedia-based applications: combining text with visual and audio devices.
- Tools and backup systems (utilities): applications which have not especially been developed for foreign language learning such as monolingual or bilingual dictionaries, spell checkers, word processing programs etc.
- Artificial Intelligence applications: consisting of formalized knowledge on the subject or on the user (see Kopp and Wachsmuth 2012 in this volume on artificial intelligence in artificial interactivity).
- Communicative applications: These can be summarized by the term "computer-mediated communication" (see above, chapter 3).

It is important to mention that most available systems show different characteristics. In the following only *Artificial Intelligence* (AI, cf. Luger 2008) applications will be considered because only they use *Natural Language Processing* (NLP, cf. Clark et al. 2010) for approaching the learner's usage of language. Such a CALL application implies for example grammar knowledge about the target language, domain knowledge or knowledge about the user to adapt individually to the necessities of each single user. Even though these systems could be more communicatively complex, they often invoke the same learner activities as traditional CALL applications. An advantage of this type of system is its intelligent behaviour with regard to the learner input, which offers sophisticated feedback.

The term "parser-based CALL" does not only better match the special field than the expression "intelligent CALL". It also indicates the application of language technology in CALL as a possible approach beside others (cf. Heift and Schulze 2007: 2). Numerous applications for different languages have been developed during the now 40-year history of ICALL – unfortunately very often not going beyond the experimental stage (for an overview see Amaral and Meurers 2006). As it would be impossible to enumerate all these systems, three applications for the German language will be considered below.

Bridge German Tutor

The *Bridge German Tutor* (Holland, Kaplan, and Sams 1995; Kreyer and Criswell 1995; Sams 1995; Weinberg et al. 1995) was developed at the Army Research Institute (ARI). The system utilizes a robust shift-reduce-parser (cf. Weinberg et al. 1995) that can be upgraded by the system administrator. It offers detailed feedback to the learner input and manages simultaneously a student model with the learner's weak and strong sides which can be used for supervising the progress (cf. Sams 1995). The Bridge Tutor realizes different types of exercises. In order to give feedback to the learner, gap-filling exercises and free text production exercises are analyzed by the natural language processing modules. Grammar mistakes are identified and classified as primary and secondary mistakes. This categorization can be influenced by the authors of the exercises. The exercises are tailored in such a way that they take the parser's limits into consideration but take advantage of its full potential. The exercises can be composed by means of an authoring system, and no specific programming skills are needed for its use. The exercises can then either be solved one after the other according to the progress or be chosen from an exercise pool.

ALLES

ALLES (*Advanced Long-distance Learning Education System*, EU project 2002–2005, http://alles.atosorigin.es) is a multimedia-based language learning system that was developed for the advanced independent learner of level B2 to C1 of the Common European Framework of Reference for Languages. The following skills can be trained for English, German, Spanish and Catalan:

- listening comprehension,
- reading comprehension and
- writing skills.

The exercises imply a monotonous series of MC-exercises and gap-filling texts and partly open types of exercises which expect a free text input. Here, short questions have to be answered, sentences to be paraphrased and free texts such as e-mails and memos to be written under strong constraints. This kind of control is supposed to provide for a content-related feedback. It is checked whether important concepts have been kept in the right order in the answer. This feedback, however, is based on given phrases which have to be applied exactly in the same way in the text. If a phrase cannot be found, it will be suggested that the given content-related item is to be mentioned in the text. All gap-filling exercises (free text input counts as gap-filling, too) are first checked orthographically during the analysis. For spelling mistakes proposals of correction are presented, if possible. Finally, the input is analyzed with the morphological parser MPRO ("Morphology PROgram"). This natural-language analysis offers appropriate feedback for declination, conjugation and for wrong commas in infinitive constructions. The feedback for syntax errors is based on given pattern solutions, so missing or useless words or those in a wrong position are found with the help of a simple pattern-matching device. If a single mistake causes several mistakes, messages about those are not collected and summarized but are passed on directly to the learner – this manner of correction can be very confusing for users.

Deutsch-Uni online (DUO)

DUO, the "Deutsch-Uni online" (http://www.iai.uni-sb.de/iaide/de/uni_deutsch.htm) (cf. Roche 2008) is a complex communicative learning setting. It offers online courses for all levels with different focuses. The courses imply interactive instruction material which can be accompanied by tutor-oriented assistance (assisted learning), if desired. Furthermore, DUO offers possibilities for working in groups together in virtual classes and enables contacts to other learners in the whole world. Exercises imply the following forms of interaction: Marking of text parts, drag-and-drop exercises, MC-exercises (simple choice, multi-choice) yes/no-questions with input of right solutions) gap-filling text (traditional with drop-down-list, paraphrase), crossword puzzle, word guessing (European mandrake), matching exercises, and audio tapes. The program implies an "intelligent electronic assistant" who executes grammatical analysis on the syntactical level and gives hints to the learner for correction. The analysis is based on the MPRO system of the IAI Saarbrücken. Furthermore additional helping tools (as for example a learning dictionary and a concordance module) and media communication are integrated in the learning setting.

7. Conclusion and prospect of future developments

Given the relatively short time span which has been available for the development so far, it is impressive to see to what extent digital media has already been integrated into foreign language learning. From the early stand-alone exercise emerged the multimedia learning environment accompanying and sometimes replacing traditional textbooks. Early computer mediated communication via email developed into sophisticated types of interaction via conference software, communal learning environments and the whole range of social software (cf. Chaudhuri and Puskas 2011). First studies on the integration of virtual environments like Second Life, in which learners meet in the form of their avatars, show how the various means of communication can be made functional for different learning purposes, e.g. while voice chat is used for direct 'face to face' communication, the accompanying text chat is used simultaneously for meta-communicative messages (cf. Biebighäuser and Marques-Schäfer 2009).

Looking into the future, one could imagine an integration of CALL and CMC as sketched in Rösler 2000: virtual worlds becoming so sophisticated that learners don't just meet other learners in the form of avatars, but encounter other figures which/who, having passed the Touring test, can neither be identified as other people's avatars nor as artificial scripted figures. Learners would then simply interact naturally with other figures who adapt, in the interaction, to the learners' level of linguistic and cultural knowledge: the learner's perspective would be that of one in the process of natural foreign language acquisition, but it would take place in a very complex artificial learning environment which anticipates learners' moves far beyond what is currently available as simulations for foreign language learning.

Note

1. Some parts of this paper are based on state-of-the-art reports provided by Susanne Kämmerer, Csilla Puskas and Sven Saage. The authors would like to thank them for this valuable support.

References

Amaral, Luiz and Detmar Meurers
2006 Where does ICALL Fit into Foreign Language Teaching? 23rd Annual Conference of the Computer Assisted Language Instruction Consortium (CALICO), May 19, Honolulu, University of Hawaii [URL http://purl.org/net/icall/handouts/calico06-amaral-meurers.pdf].

Bangs, Paul
2003 Engaging the learner – how to author for best feedback. In: Uschi Felix (ed.), *Language learning online: Towards best practice*, 81–96. Lisse: Sweets & Zeitlinger.

Bax, Stephen
2003 CALL – past, present and future. *System* 31, 1/03: 13–28.

Beißwenger, Michael and Angelika Storrer
2008 Corpora of Computer-Mediated Communication. In: Anke Lüdeling and Merja Kytö (eds.), *Corpus Linguistics. An International Handbook*. Volume 1, 292–308. (Handbooks of Linguistics and Communication Science 29.1.) Berlin/New York: De Gruyter.

Belz, Julie A. and Andreas Müller-Hartmann
2003 Teachers as Intercultural Learners. Negotiating German-American Telecollaboration along the Institutional Fault Line. *The Modern Language Journal* 87, 1: 71–89.

Belz, Julie A. and Steven L. Thorne (eds.)
2005 *Internet-Mediated Intercultural Foreign Language Education.* Boston: Thomson Heinle.

Biebighäuser, Karin and Gabriela Marques-Schäfer
2009 Text-Chat und Voice-Chat beim DaF-Lernen online: Eine empirische Analyse anhand der Chatangebote des Goethe-Instituts in JETZT Deutsch lernen und in Second Life. *Info DaF* 36, 5: 411–428.

Brammerts, Helmut and Karin Kleppin (eds.)
2001 *Selbstgesteuertes Sprachenlernen im Tandem. Ein Handbuch*. Tübingen: Stauffenburg.

Chambers, Angela
2010 What is data-driven learning? In: Anne O'Keffee and Michael McCarthy (eds.), *The Routledge Handbook of Corpus Linguistics*, 345–358. London: Routledge.

Chapelle, Carol A.
2000 Is network-based learning CALL? In: Mark Warschauer and Richard Kern (eds.), *Network-based language teaching: Concepts and practice*, 204–228. Cambridge: Cambridge University Press.

Chapelle, Carol A.
2001 *Computer applications in second language acquisition*. Cambridge: Cambridge University Press.

Chaudhuri, Tushar and Csilla Puskás
2011 Interkulturelle Lernaktivitäten im Zeitalter des Web 2.0. Erkenntnisse eines telekollaborativen Projektes zwischen der Hong Kong Baptist University und der Justus-Liebig-Universität Gießen. In: *Info DaF* 38, 1: 3–25.

Clark, Alexander, Chris Fox and Shalom Lappin (eds.)
2010 *The Handbook of Computational Linguistics and Natural Language Processing* (Blackwell Handbooks in Linguistics 57.) Oxford: Wiley-Blackwell.

Davies, Graham
1997 Lessons from the past, lessons for the future. 20 years of CALL. In: Dieter Kranz, Lienhard Legenhausen and Bernd Lücking (eds.), *Multimedia Internet-Lernsoftware*, 111–130. Münster: agenda Verlag.

Granger, Sylvaine
1998 The computer learner corpus: a versatile new source of data for SLA research. In: Sylvaine Granger (eds.), *Learner English on Computer*, 3–8. London: Longman.

Granger, Sylvaine
2002 A bird's-eye view of computer learner corpus research. In: Sylvaine Granger, Joseph Hung and Stephanie Petch-Tyson (eds.), *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*, 3–33. Amsterdam/Philadelphia: John Benjamins.

Granger, Sylviane
2009 Learner Corpora. In: Anke Lüdeling and Merja Kytö (eds.), *Corpus Linguistics. An International Handbook.* Vol. 1, 259–275. (Handbooks of Linguistics and Communication Science.) Berlin/New York: De Gruyter.

Harrington, Michael and Mike Levy
2001 CALL begins with a 'C': interaction in computer-mediated language learning. *System* 29: 15–26.

Heift, Trude and Mathias Schulze
2007 *Errors and Intelligence in Computer-Assisted Language Learning. Parsers and Pedagogues*. New York/London: Routledge.

Hess, Hans-Werner
2003 Lerner als Kunden. Informationstechnologie im Alltagseinsatz [Learners as Customers. Applied Information Technology]. *Deutsch als Fremdsprache* 40, 1: 14–23.

Hess, Hans-Werner
2006 Beyond the Rhetoric of the 'Autonomous Learner': Combining E-Learning and the Classroom. *Electronic Journal of Foreign Language Teaching* 3, 1: 102–120.

Holland, Melissa V., Jonathan D. Kaplan and Michelle R. Sams (eds.)
1995 *Intelligent Language Tutors: Theory Shaping Technology.* New Jersey: Lawrence Erlbaum Associates.

Johns, Tim and Philip King
1991 *Classroom Concordancing*. Birmingham: University of Birmingham.

Kopp, Stefan and Ipke Wachsmuth
2012 Artificial interactivity. In: Alexander Mehler, Laurent Romary and Dafydd Gibbon (eds.): *Handbook of Technical Communication*. Berlin/Boston: De Gruyter.

Kreyer, Steve and Eleanor Criswell
1995 Instructor as Author in an Adaptive, Multimedia, Foreign Language Tutor. In: Melissa V. Holland, Jonathan D. Kaplan and Michelle R. Sams (eds.), *Intelligent Language Tutors: Theory Shaping Technology*, 45–53. Mahwah/New Jersey: Lawrence Erlbaum.

Last, Rex W.
1989 *Artificial intelligence techniques in language learning*. Chichester: Horwood.

Legutke, Michael and Dietmar Rösler
2005 Enhancing collaborative work by integrating digital media into foreign language teacher education. *Fremdsprachen lehren und lernen* 34: 174–191.

Legutke, Michael, Andreas Müller-Hartmann and Stefan Ulrich
2000 Neue Kommunikationsformen im fremdsprachlichen Unterricht. In: Gerd Fritz and Anderas Jucker (eds.), *Kommunikationsformen im Wandel der Zeit*, 51–73. Tübingen: Niemeyer.

Levy, Michael
1997 *Computer-Assisted Language Learning. Context and Conceptualisation*. Oxford: Clarendon Press.

Lüdeling, Anke, Seanna Doolittle, Hagen Hirschmann, Karin Schmidt and Maik Walter
2008 Das Lernerkorpus Falko. *Deutsch als Fremdsprache* 2, 45: 67–73.

Lüdeling, Anke and Merja Kytö (eds.)
2008/2009 *Corpus Linguistics. An International Handbook*. Volume 1/2. (Handbooks of Linguistics and Communication Science.) Berlin/New York: De Gruyter.

Ludewig, Petra
2005 *Korpusbasiertes Kollokationslernen: Computer-Assisted Language Learning als prototypisches Anwendungsszenario der Computerlinguistik.* (Computer Studies in Language and Speech 9.) Frankfurt-Main/Berlin: Peter Lang.

Mitschian, Haymo
1998 Von Mitteln und Mittlern. Zur Rolle des Computers beim Fremdsprachenlernen. *Info DaF* 25, 5: 590–613.

Mukherjee, Joybrato
2008 Anglistische Korpuslinguistik und Fremdsprachenforschung: Entwicklungslinien und Perspektiven. *Zeitschrift für Fremdsprachenforschung* 19, 1: 31–60.

Müller-Hartmann, Andreas
2000 "Wenn sich die Lehrenden nicht verstehen, wie sollen sich dann die Lernenden verstehen?" Fragen nach der Rolle der Lehrenden in global vernetzten Klassenräumen. In: Lothar Bredella, Herbert Christ and Michael K. Legutke (eds.), *Fremdverstehen zwischen Theorie und Praxis, Arbeiten aus dem Graduierten-Kolleg "Didaktik des Fremdverstehens"*, 275–301. Tübingen: Narr.

Nagata, Noriko
1993 Intelligent computer feedback for second language instruction. *The Modern Language Journal* 77, iii: 330–339.

Nesselhauf, Nadja
2004 Learner Corpora and their potential for language teaching. In: John Sinclair (ed.), *How to Use Corpora in Language Teaching*, 125–152. Amsterdam/Philadelphia: John Benjamins.

Nübold, Peter
2006 Vom Sprachlabor zum Multimedia-Sprachlabor. In: Udo O.H. Jung (ed.), *Praktische Handreichungen für Fremdsprachenlehrer*, 299–304. Frankfurt am Main: Peter Lang, 4. edition.

O'Dowd, Robert (ed.)
2007 *Online Intercultural Exchange. An Introduction for Foreign Language Teachers*. Clevedon/Buffalo/Toronto: Multilingual Matters.

O'Dowd, Robert and Markus Ritter
2006 Understanding and Working with 'Failed Communication' in Telecollaborative Exchanges. *CALICO Journal* 23, 3: 1–20.
Roche, Jörg
2008 *Handbuch Mediendidaktik*. Ismaning: Hueber.
Rösler, Dietmar
1998 Autonomes Lernen? Neue Medien und 'altes' Fremdsprachenlernen. *Info DaF* 25, 1: 3–20.
Rösler, Dietmar
2000 Fremdsprachenlernen außerhalb des zielsprachigen Raums per virtueller Realität. In: Gerd Fritz and Andreas Jucker (eds.), *Kommunikationsformen im Wandel der Zeit*, 121–135. Tübingen: Niemeyer.
Rösler, Dietmar
2006 Zwischen Lehrwerk und Online-Sprachbad. Zur Neubestimmung von Funktion und Gestalt von Lehrmaterialien im Anfängerunterricht. In: Hans-Jürger Krumm and Paul Portmann-Tselikas (eds.), *Theorie und Praxis. Österreichische Beiträge zu Deutsch als Fremdsprache*, 143–154. Innsbruck: Studien-Verlag.
Rösler, Dietmar
2007 *E-Learning Fremdsprachen – eine kritische Einführung*. Tübingen: Stauffenburg, 2nd edition.
Rösler, Dietmar
2010 Die Funktion von Medien im Deutsch als Fremd- und Deutsch als Zweitsprache-Unterricht. In: Hans-Jürgen Krumm, Christian Fandrych, Britta Hufeisen and Claudia Riemer (eds.), *Deutsch als Fremd- und Zweitsprache. Ein internationales Handbuch*, 1199–1214 (Handbooks of Linguistics and Communication Science 35/2). Berlin/New York: de Gruyter Mouton.
Sams, Michelle R.
1995 Advanced Technologies for Language Learning: The BRIDGE Project Within the ARI Language Tutor Programm. In: Melissa V. Holland, Jonathan D. Kaplan and Michelle R. Sams (eds.), *Intelligent Language Tutors: Theory Shaping Technology,* 7–21. Mahwah/New Jersey: Lawrence Erlbaum.
Sasaki, Felix and Arle Lommel
2012 Multilingual computing. In: Alexander Mehler, Laurent Romary and Dafydd Gibbon (Eds.): Handbook of Technical Communication. Berlin/Boston: de Gruyter.
Schmenk, Barbara
2008 *Lernerautonomie: Karriere und Sloganisierung des Autonomiebegriffs*. Tübingen: Narr.
Schulmeister, Rolf
2002 *Grundlagen hypermedialer Lernsysteme: Theorie – Didaktik – Design. München:* Oldenbourg, 3. edition.
Seidlhofer, Barbara
2002 Pedagogy and local learner corpora: working with learning-driven data. In: Sylviane Granger, Joseph Hung and Stephanie Petch-Tyson (eds.), *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*, 214–234. Amsterdam/Philadelphia: John Benjamins.

Tamme, Claudia
2001 *E-Mail-Tutorien: eine empirische Untersuchung E-Mail-vermittelter Kommunikationen von Deutschstudierenden und Deutsch-als-Fremdsprache-Lehrenden in der Ausbildung*. Online: http://geb.uni-iessen.de/geb/volltexte/2003/1009/ (16. 6. 10).

Tognini-Bonelli, Elena
2001 *Corpus Linguistics at Work*. (Studies in Corpus Linguistics 6.) Amsterdam: John Benjamins B.V.

Warschauer, Mark
1996 Computer-Assisted Language Learning: An Introduction. In: Sandra Fotos (eds.*), Multimedia Language Teaching*, 3–20. Tokyo: Logos International.

Weinberg, Amy, Jeffery Garman, Joe Martin and Paola Merlo
1995 A Principle-Based Parser for Foreign Language Tutoring in German and Arabic. In: Melissa V. Holland, Jonathan D. Kaplan and Michelle R. Sams (ed.), *Intelligent Language Tutors: Theory Shaping Technology,* 23–43. Mahwah/New Jersey: Lawrence Erlbaum.

Wolff, Dieter
1993 New Technologies for Foreign Language Teaching. In: *Foreign Language Learning and the Use of New Technologies. Conference Proceedings. London 1993*. Brussels: Bureau Lingua.

18. Framing Multimodal Technical Communication

With Focal Points in Speech-Gesture-Integration and Gaze Recognition

Andy Lücking and Thies Pfeiffer

1. Underused semiotic means

In *face-to-face* conversation, the whole human body is used as a "semiotic display".[1] The distance that interlocutors maintain between each other signals, under normal conditions,[2] their social relationship (Hall, 1966). There is a culturally varying proximal personal space that only familiar people are allowed to enter. Gaze and facial expressions are used to convey attitudes and emotions (Hager and Ekman, 1983); gaze is also an important means to establish and signal attention (Argyle and Cook, 1976). Gestures, that are hand and arm movements that contribute to the interlocutors' dialog turns, are a further frequent semiotic means in everyday talking (Kendon, 1980). Try to describe how to change flat tires without moving your hands! According to the truism "It's not what you say, but how you say it", a compliment may become an insult simply by intonation; the paralinguistic or suprasegmental dimension of articulation modifies or amplifies what is said. These highlights shall just bring to mind the well-known fact that communication is a multifaceted affair that is performed simultaneously on various levels (see for instance the ethnomethodological communication analysis of Goodwin, 1981). While each type of verbal or non-verbal communication means has a semiotic autonomy to a certain degree, that is, obeys inherent rules and defaults, in a communication situation they are interrelated to one another. The types of multimodal interaction form an intricate and complex communication network (Condon and Ogsten, 1967). At the bottom line, human communication draws heavily on various non-verbal resources, resources that wait to be exploited for enhancing technical communication: *Semiotic means are still underused. How can they be made productive for technical communication?* There are a couple of applications around as a reaction to this question. We try to frame the field of technical multimodal communication along the lines already made salient by the composite structure of this chapter's title, namely, by means of *technical*, of *multimodal*, and of *communication* sections. Section 2 introduces multimodality as well as pertinent related notions and provides a terminological framework.

The technical side is taken up in Sections 4 to 6. Technical systems are to be distinguished according to their processing direction into *production* systems and *reception* ones. The technical challenge for multimodal reception systems is the *fusion* of information coming in via two or more modalities and is dealt with in Section 5. Production systems face the reversed problem: how to *fission* a unified information and output it via various media? Production issues are taken up in Section 6. A precondition for a technical mastering of multimodal communication is capturing pertaining non-verbal input/output behavior by some processing device. The coverage of technical applications will be spread out in terms of key tools in sections 4 and 6.2, respectively.

Throughout this chapter we use boldface typesetting in order to highlight **benchmarks** for truly multimodal communication. As part of the conclusions in Section 7, we assess current applications in light of these theoretical and descriptive benchmarks, finally framing the perspectives of technical multimodal communication.

2. Modality, medium, code

In order to clarify what 'multimodality' means, one has to explicate the notion of 'modality'. However, "[i]t is not easy to define the notion of 'modality', or of 'multimodality'", as Gibbon, Mertins, and Moore (2000, p. 102) complain. Part of the diffculty is that there are a two closely related terms, namely 'medium' and 'code'. Since we regard them as a *sine qua non* in the present context, a contextual explication is a particularly suitable method for the desired terminological clarification. In order to set the stage, let us simplistically assume that communication is the exchange of information between interlocutors. The general scheme is the following: *Sender A transmits via M information X by means of sign Z to recipient B.* This rather general and technical characterization needs further clarification and refinement. We confine ourselves by pointing to some pertinent seminal frameworks. Please see Israel and Perry (1990) for an account of the notion of information; see Shannon (1948) and Shannon and Weaver (1949) for a telecommunication approach to communication that introduces the popular sender-receiver model where the information exchanged is coded in the transmitted signal; see Grice (1975) and Sperber and Wilson (1986) for large-scale communication theories that also account for information not explicitly encoded in the signal. Of course, many more parameters could be and have been added to the scheme, like the social status of the interlocutors, their mental household (beliefs, desires, intentions), the spatio-temporal context of the conversation, the kind of conversation, some notion of lifeworld, and also biological facilities (see e.g. Anderson and Keenan, 1985; Recanati, 2004; Garfinkel, 1967; Sonesson, 2009; Clark, 1992; Levinson, 2006; Cohen, Morgan, and Pollack, 1990; Austin, 1962; Tomasello, 2008). However, the general scheme already provides the reference points needed for distinguishing the domains relevant for multimodal technical communication.[3]

The first thing to note is that information *X* (or meaning, if you prefer a more semantic term)[4] is an abstract entity that cannot be accessed directly by the human senses. Whatever information is to be conveyed, it has to be conveyed by means of a perceptible form that is open to interpretation – that is, by means of a *sign* (*Z* in the general scheme above). Information has to be packaged into a physical container in order to accomplish accessibility, that is, information has to be transmitted via a *medium M* that in turn is perceivable by human cognition – the *modality* aspect of the information (see Subsection 2.2 below). A sign can be interpreted or decoded via interpretation guidelines. We collectively refer to aspects that relate to the interpretation of a sign as the *code* of the sign (Chandler, 2002, chap. 9). Codes may range from conventional semantic rules to wide cultural backgrounds (cf. Eco, 1976; Searle, 1978; Chandler, 2002). Furthermore, signs are organized into *sign systems*, which give rise to syntactic relations and provide a means for typing signs according to the sign system they are a part of (Goodman, 1976; Posner, Robering, and Seboek, 1997). At the bottom line, there is the following three-fold systematics: a sign is packaged into a "sign container", and this is so in both a concrete and a virtual sense. In a concrete sense the container is the sign's medium, that is, a material device like ink of paper. Virtual containers are the kinds of information type or the sign systems the sign belongs to, that are, for instance, written or spoken language, figures, tables, paintings, gestures, music, or film. The latter container facilitates as well as constrains the interpretation of a sign, the former one warrants its perceptibility by the sign's modality.

Against this background we are able to distinguish multimodal, multimedia and multicodal communication (see Weidenmann (1997) and Posner, Robering, and Seboek (1997) for closely related explications):

– *Multimodality:* communication is *multimodal* iff the sign(s) *Z* exchanged between interlocutors *A* and *B* are perceived by the recipient via more than one sensory interface.
– *Multimediality:* communication is *multimedia* iff interlocutors *A* and *B* use more than one means *M* for transmitting *Z*.
– *Multicodality:* communication is *multicodal* iff the information *X* transmitted between interlocutors *A* and *B* is encoded in signs *Z* that belong to more than one sign system (i.e. are of more than one information type).

Otherwise, communication is *unimodal*, *-media*, or *-codal*.

We call a communication instance *technical*, iff

- either one of the interlocutors *A* or *B* is an artificial device (the other being human), or both are artificial devices (say, robots or avatars); or
- the transmission means *M* is an artificial device (see Kittler (2004) for a short history on technical communication means).

According to the first distinction, we have to distinguish *human-human* (HHI) from *human-computer* (HCI) and *computer-computer* (CCI) interaction (Bellotti, Back, Edwards, Grinter, Henderson, and Lopes, 2002; Poslad, 2009). The second distinction separates *technically mediated* from *technically unmediated* communication (Walther, 1996; Thurlow, Lengel, and Tomic, 2004). Both conditions are not entirely independent from one another, meaning that technical communication instances can be cross-classified only partially. The reason behind the dependence is that communication involving non-human avatars necessarily draws on technical devices – HCI and CCI are *per definitionem* technically mediated. HHI on the other hand, which basically is technically unmediated, can be augmented by artificial devices. The twofold distinction is also reflected in two prevalent uses of the term 'medium' (at least in current English and German), namely medium as "a substance regarded as the means of transmission of a force or effect" vs. medium as "something (as a magnetic disk) on which information may be stored".[5] In order to avoid confusion, both senses of *medium* have to be kept apart. We distinguish them terminologically by calling the former one *medium as substance* (see Subsection 2.2) and the latter one *medium as device* (see Subsection 2.3). For a more fine-grained terminological classification see Hess-Lüttich and Schmauks (2004) and the literature given there.

Out of these distinctions arise as a result $2^3 \times 3 \times 2 = 48$ combinatorically possible manifestations of communication: {multi-/unimodal, multi-/unicodal, multi-/unimedia} × {HHI, HCI, CCI} × {technically mediated/unmediated}, Substracting the impossible combinations of {HCI, CCI}/technically unmediated interactions, there are 32 ($2^3 \times 4$) communication settings that actually can be manifested, from which the half (i.e. 16) is multimodal. Seen from an artificial intelligence perspective, the most advanced technical system would be capable of *multimodal, multimedia, multicodal, computer-computer communication*.

Multimodality, multimediality and multicodality appear conversely to sender *A* and recipient *B*, be they human or artificial, as displayed in Figure 1. The figure is partitioned according to the comprehension and the production side of processing, namely into *multimodal fusion*, the integration of heterogeneously accessed information on the one hand and into *multimedia fission*, the "extegration" of unified information onto (possibly) different media. The in-

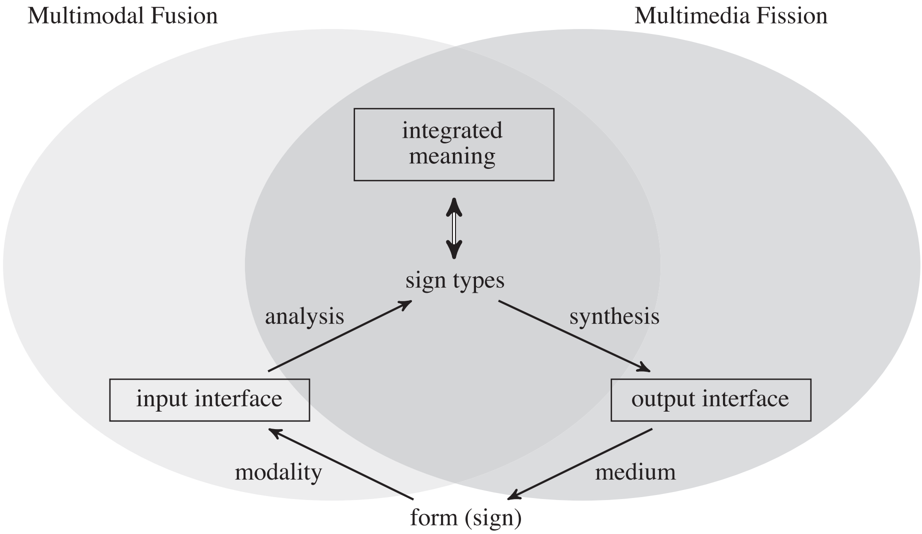


Figure 1. Processing in conversation: multimodal, multimedia, multicodal information flow.

coming multimodal signs are received by the recipient *B*. The production of multimedia, multicodal utterances is assumed to start with an unified, "amodal" *idea unit* (McNeill, 2000) that gets split and distributed onto the various sign types and media. Thus, **the modality/media interface is input/output asymmetrical.** Psycholinguistic research so far dealt mainly with the multimedia fission problem in the domain of synchronized speech-and-gesture production. The co-verbal production of a nonverbal communication means is captured within the language production model spelled out by Levelt (1989). The proposed accounts span the range from epiphenomenalism (Krauss, Chen, and Gottesmann, 2000) over common amodal information sources (McNeill, 2000; Ruiter, 2000) to a steadily inter-codal calibration (Kita and Özyürek, 2003). Multimodal, multicodal comprehension faces the converse problem. In an analysis step, the input signs have to be assigned their sign types, for example, whether they are signs of the system of gesture or proxemics. If there are various and differently code-typed signs, they have to be integrated semantically: this is **the heterogeneous information fusion problem**, a crucial benchmark for technical applications. As a consequence of the terminological framework presented in this chapter the problem of fusing heterogeneous information is actually the problem of integrating *multicodal* signs! The dimension of codality is often overlooked in the literature on multimodal corpora, multimodal dialog systems, or speech-and-gesture conversation, where the fusion problem is introduced as a challenge posed by *multimodality* (see e.g. Delgado and Araki, 2005; Gibbon, Mertins, and Moore, 2000; Allwood, 2007; Quek, McNeill, Bryll, Duncan, Ma, Kirbas, McCullough, and Ansari, 2002). For example, the novel *Die Schrecken des Eises und der Finsternis* written by the Austrian littérateur Christoph Ransmayr is unimedia and unimodal, but multicodal. Ransmayer has interlaced tabulations and figures in his prose text, that is, he combines three different types of signs. Since all three types – text, tables and figures – are realized physically as a book by ink of paper, the novel is unimedia. Furthermore, since tabulations as well as figures and written text are perceived via the seeing sense, the novel is also unimodal.

2.1. Nonverbal means

The research that comes under the heading of ‘multimodal’ can be characterized as research on communication that deals with nonverbal communication means in particular. Accordingly, the specific aspect of multimodal communication studies is their focus on communication means other than written or spoken language. Nonverbal signs are either studied in isolation (e.g. Galley, 1997) or in company with language (e.g. Kendon, 1972; Kendon, 2000). What kinds of nonverbal signs are there? In human-human, technically unmediated communication the following sign systems are used (compare the overviews provided by Luccio (1997) and Nöth (2000)):

– *Paralinguistic signals.* Aspects of phonation that are not part of the meaning differentiating vowels and consonants of a spoken language like prosody, timbre, and articulation speed are called *paralinguistic* (Birdwhistell, 1970). They are co-realized by language phonations but often stretch more than one verbal segment (vowel, consonant, syllables, …). Therefore they are also called *suprasegmentals* (Clark and Yallop, 1990). Paralinguistic signals have to be distinguished from phonetic features that are indeed part of the standard phonological language system, like tone pitches in tonal languages or sentence intonation pattern (Clark and Yallop, 1990). Besides signaling “affective” meanings, paralinguistic means are used as markers for grammatical structure (Crystal, 1971). The following acoustic features of the vocal apparatus are used for producing paralinguistic signals: *pitch*, *loudness*, *speed*, *rhythm*, and *larynx* and *oral effects* (Crystal, 1971). For example, observe the difference between a slow and a quick phonation of “Really” in English.
  The term ‘paralinguistics’ may also be understood “as including most or all of those aspects of linguistic communication that are not part of language *per se*, but are nonetheless somehow involved in the message or meaning the communicator conveys.” (Wharton, 2009, p. 5). In line with this conception of paralinguistics, Poyatos (1975) gives a “paralinguistic” account of the range of nonverbal means in face-to-face communication. Using the notion of ‘channel’ (see Subsection 2.2 below) as topmost category, Poyatos

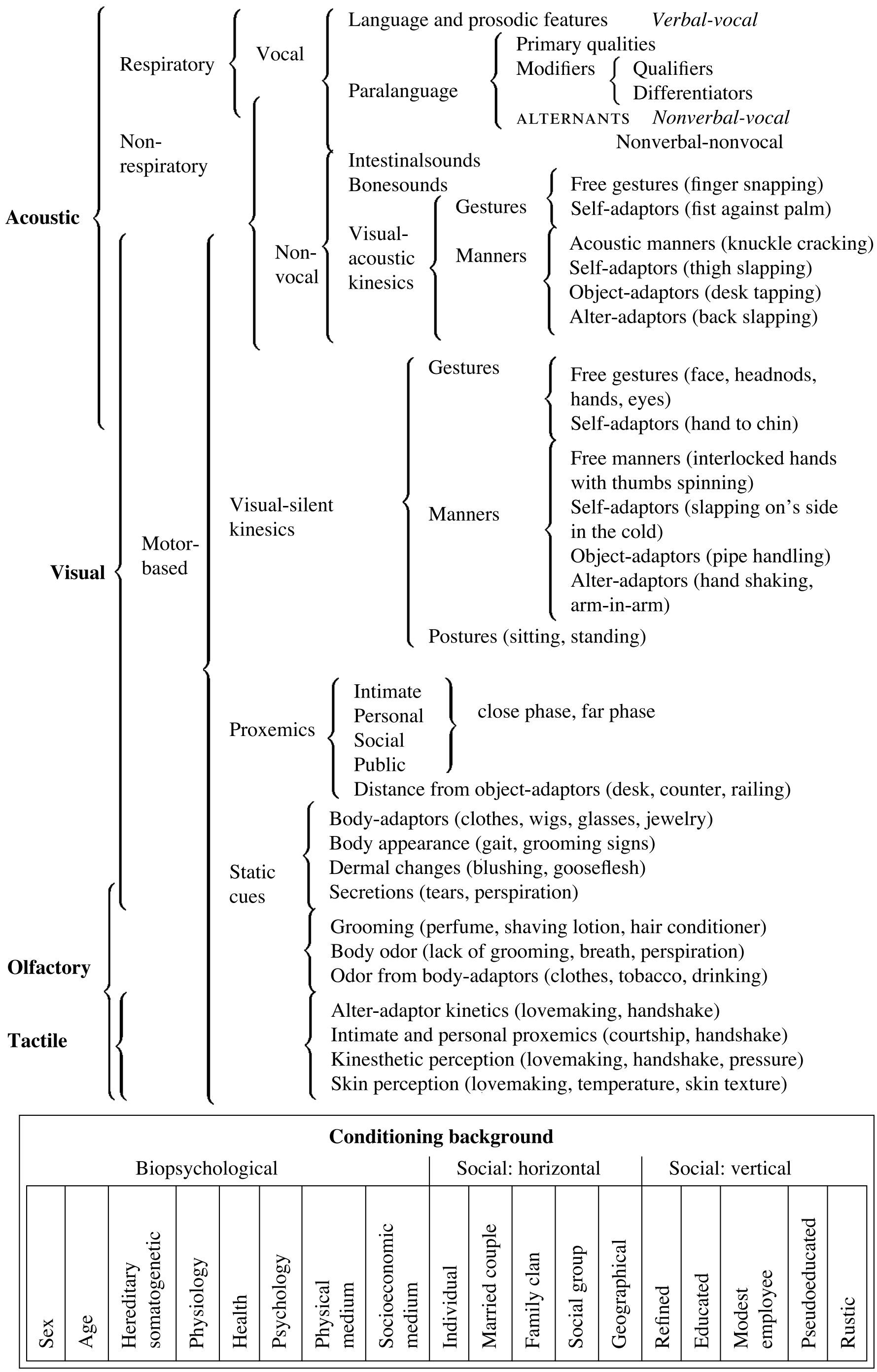


Figure 2. "Total Body Communication", diagram from Poyatos (1975, p. 292).

(1975, p. 292) presents a systematic diagram of multimodal communication, which is reprinted here as Figure 2 on the facing page.

- *Gesture.* Gestures are hand and arm movements that contribute to conversational utterances (Kendon, 2004). As reported by Rieser (2008), **gestures are associated with all levels of linguistic constituency**, from single words over phrases and sentences up to descriptions of situations. According to Kendon (1980), gestures can be differentiated along a “closeness-to-speech” continuum: on the one hand there are autonomous gestural systems obeying strict form standards including syntactic ones: sign languages. Single symbolic gestures like the “OK”-gesture in many western countries also belong to this side and are known as *emblems*. On the other hand there are spontaneous, not conventionally regulated co-verbal gesticulations that are simply called gestures. McNeill (1992a), partly reviving earlier work of Efron (1941) and Ekman and Friesen (1969), amongst others, classified gestures into *iconics* (pictorial representations), *metaphorics* (abstract representations), *deictics* (pointing gestures), and *beats* (emphasizing, not-representational gestures). Bavelas and Gerwing (2007) added the interactive gesture category of *conversational gestures*. A gesture classification scheme, relating to the work of Wundt (1911), that draws on the notion of cultural practice and action has been devised by Müller (1998), where gestures are classified according to a *representation technique* like *drawing*, *shaping* or *placing*. A gesture classification scheme that places a gesture in a socio-cultural “artisan” context has also been devised by Streeck (2009). There is some discussion on which representation techniques there are and whether they can be reduced to some basic depiction methods (see e.g. Kendon, 2004; Enfield, 2009). The representation technique notion is not undisputed: most notable Janet Bavelas and her colleagues criticize the type-view on gesture and argue for a social approach where “function constrains form” (see Gerwing and Bavelas, 2004, p. 158). Gesture as a means of rhetoric already has a long history going back at least to Quintilian (1. Century). See Murphy (2003) on some aspects on gesture classification.
- *Body Signals.* We collect bodily communication means that are not realized as hand-and-arm gestures, facial expressions, or gaze under the label of “body signals”. To here belong structuralistic accounts like the *Kinesics* of Birdwhistell (1970) in which a set of *kinemes*, say, a certain head nod, are identified. To here also belong more or less loose collections of the vague and ambiguous range of body-related signals from eye-brow raising (Grammer, 2004) to the outward appearance (Luccio, 1997), which play a role in the semiosis in a communication situation (Posner, 1997). However, notwithstanding the heuristic informativity of body signals in communication, the methodological foundation and semantic underpinning of their role in the interpretation process in communication are unclear (cf. Nöth, 2000).

- *Facial Signals.* Among the various nonverbal communication means at disposal in technically unmediated, human-human communication facial expressions play a prominent role. The facial display is nearly permanently active (Nöth, 2000). It is the main information source for the physiological as well as psychological condition of the speaker (Grammer, 2004). Facial expressions have cross-cultural validity: six basic emotions are expressed in the cultures investigated in the same way, namely surprise, angst, joy, misery, disgust and anger. They are distinguished by constant contraction patterns of facial muscles (Ekman and Friesen, 1978). The main function of facial signals is usually considered to be *emotional expressivity*: speakers signal their mood and conation as well as their attitudes towards the topic talked about (cf. Nöth, 2000). Although facial signals mainly occur in company to speech they may also be produced independently of spoken language (Argyle, 1975).
- *Gaze.* The functions of the eyes in communication are manifold. The use of eye gaze is bound up with *joint attention* (see Argyle and Cook (1976) and the articles compiled by Moore and Dunham (1995)). Gaze also plays a role in reference resolution (Prasov and Chai, 2008). The eyes contribute to dialog management (Kendon, 1967). The communicative potential of gaze has also been drawn upon in the arts, for instance in Velázquez' painting *Las Meninas* from 1656 (*Museo del Prado*, Madrid), where not only the portrayed figures but also the viewer are related to each other by means of gaze directions that cross the realms of reality and art. The eyes are more than the other nonverbal communication means seen to be bound up with the "self" of people: the eyes, in common parlance, are "the mirror of the soul".
- *Proxemics.* Proxemics is a kind of spatio-social behavior, whose semiotic significance is connected to the interpersonal distances of interlocutors. The term 'proxemics' has been coined by Hall (1968). Slightly earlier, Hall (1966) distinguished proximal *intimate* and *personal spaces* from distal *social* and *public spaces*. The spaces are centered within the subject. The average radius of each space is subject to cultural variety: what in one culture counts as social space may in another culture already be part of the personal space (Argyle, 1975).
- *Tactile Codes.* Within the intimate space of proxemics the tactile code becomes significant. Thus, *tacesics* as it also has been called (Kauffman, 1971), can be seen as a sub-discipline of proxemics. Tactile codes govern reciprocal, proximal social behaviors from hand shaking to kissing (Argyle, 1975).
- *Time behavior.* The temporal aspect of communication has been pointed at by Hall (1966). Although it is questionable whether time can be a communicative dimension on its own – time is assumed to depend on something changing (Nöth, 2000) – a semiotics of time called *chronemics* has nonethe-

less been suggested (Bruneau, 1980). Examples for chronemically significance in communication are the length of laughter, respiratory pauses, or the duration of a visit (Poyatos, 1975). In line with the observation that temporal experience depends on the changing of something, chronemics is probably best conceived of as a kind of paralinguistic or suprasegmental feature (Nöth, 2000; see also the "Paralanguage-to-Chronemics" category proposed by Poyatos, 1975).

– *Synchrony.* The register of nonverbal communication means occurring in technically unmediated HHI collected so far deals with single means in isolation. The peculiarity of *multimodal* communication, however, is that various means occur simultaneously (Gibbon, Mertins, and Moore, 2000; Delgado and Araki, 2005). The temporal relations between different communication means (including the verbal one) give rise to an own kind of significant time behavior that involves *synchrony* as a central notion (Wiltshire, 2007). For this reason, we see the temporal interrelationships between communication means as a locus of semiosis in addition to chronemics (see above). Tuite (1993) argues for "rhythmic pulses", that is, an oscillating pattern underlying the production of gesture and speech. Extending the rhythmic organization to head- and eye-movements, Loehr (2007) proposes a hierarchical model for the temporal coupling of the different communication means. He introduces the notion of a *pike*, which is a sign type-neutral generalization of peak, in honor of Kenneth Pike. Each sign system has its own basic rhythm. These rhythms eventually coincide in *meeting points*. Being based on pikes and sign type-specific oscillation patterns, the model of Loehr (2007) can unify synchrony-oriented observations (e.g., the phonetic synchrony rule of McNeill, 1992a) and offset-oriented observations (McClave, 1994). The significance of the temporal relationship between two or more signs in multimodal communication can be demonstrated by neuropsychological experiments on the combined comprehension of speech and gesture: event-related potential measurements (cf. Handy, 2005) showed that the N400 effect for incongruent speech-and-gesture input occurred only if the offset between speech and gesture was suffciently small ($\leq$ 160 ms) (Habets, Kita, Shao, Özyurek, and Hagoort, 2011). The temporal relationship between occurrences of different sign types is also a central topic in the design of multimodal HCI frameworks (Martin and Béroule, 1995). In analogy to the neologisms *chronemics*, *proxemics*, or *tacesics* this aspect of temporal significance could be called *kairemics*, derived from the Greek word καιρός (*kairos*) which refers to the right moment.

Within the semiotically rich multimodal, multimedia and multicodal communication situation, language is the most stable information source, however – and language realized as spoken speech is the primary communication medium: the privileged status of verbal means in multimodal communication is known as the **primacy of language** (Ruiter, 2004).[6] Seen from a cerebral perspective, the primacy of language is not so obvious. The proportions of sizes of cerebral cortex regions carrying neuronal representation of the body parts are distorted compared to the proportions of the actual physical body parts. Distorted representations can nicely be visualized by somatotopic maps, or cortical homunculi (Penfield and Jasper, 1954). Especially the relative cerebral representations of the hands are rather huge. Thus, the hands should be a reasonably “fine-mechanical” tool utilizable for communicative purposes. Not only evolutionary reflections (Butterworth, 2003), but also a wealth of empirical evidence indeed sustain that observation – see for instance the quite differently focused, seminal monographs of McNeill (1992a); Goldin-Meadow (2001); Kendon (2004). It seems safe to conclude that “gestures communicate” (Kendon, 1994) – at least if we add a diffident “somehow”. Similar, but case-by-case not so conspicuous, conclusions can of course be drawn for the other non-verbal communication means identified above.

2.2. Channels and medium as substance

Following a (German) standard work on the human physiology we have to acknowledge the following six sense modalities: the *visual*, the *auditory*, the *chemical*, the *tactile*, the *kinaesthetic*, and the *vestibular* modality (Schmidt and Thews, 1997). The visual and the auditory modality facilitate seeing and hearing, respectively. The chemical modality is responsible for molecular perception, that is, for smelling and tasting. The tactile and the kinaesthetic modality are subclasses of a general somatosensoric facility. Via their tactile modality, organisms haptically experience their outer environment. The main but by far not the only tactile device of humans are the hands which can be used for grasping and touching things. The kinaesthetic modality is the inner counterpart of the tactile one. It provides sensual access to stimuli from inside of the body, for instance from the muscles. Note, that the external-internal distinction of haptic modalities is not made by all authors. It nonetheless demarcates a relevant difference for technical applications, which usually are equipped with an external interface, but not with an internal one. The last modality from the above-given list, the vestibular modality, facilitates equilibrioception, the sense of balance. Since both the kinaesthetic and vestibular senses are involved in keeping track of the orientation of the body in space, the balance and the internal tactile modalities play a decisive role in proxemics and thereby in multimodal communication.

The human body in a technically unmediated communication situation constrains the media that are at the interlocutors’ disposal in that situation. The possible realizations of signs are restricted by the boundaries set by the immediate production capabilities of men. Most prominently, the media of sound waves is used, since this is the realization means of speech. For a number of nonverbal communication means the medium is the human body itself, or parts of the human body, to be more precise. Emotive signs expressed by facial expression, for instance, are realized as contraction patterns of facial muscles. Proxemic information is encoded by the relative orientation of the body within a spatial reference frame involving a reference object (say, the addressee). Since also olfactory signals can carry information, for instance about the degree of ripeness of some fruit, to give a non-communicative example, also the molecular medium has to be accounted for.

To summarize, in a technically unmediated setting the following modalities and media as substances are at disposal (see Jewitt (2009) for a different construal of 'modality', one that corresponds roughly to our codality):

modality visual, auditory, chemical, tactile, kinaesthetic, vestibular
medium sound waves, molecules, light waves (the latter one is the medium that transmits information of visually perceivable objects like the human body which in turn is related to an external spatial frame and possible reference objects, as is the case in proxemics)

There is a tight connection between transmission substance and receiving sense. Each sense is responsive to a certain physical substance. For instance, it is not possible to access light waves via the auditory modality. Perception constraints like this give rise to "pathways" of information transmission or *channels*. Channels can be conceived of as a connection of a medium as substance with its corresponding modality. The following channels can be distinguished as the *loci* of semiosis in HHI:

- the optical channel (Landwehr, 1997),
- the acoustic channel (Strube and Lazarus, 1997),
- the chemical channel (Kröller, 1997),
- the thermic channel (Brück, 1997),
- the tactile channel (Heuer, 1997),
- and the electric and magnetic channels (Moller, 1997).

The general scheme used as contextual blueprint for terminological distinctions at the beginning of Section 2 above makes use of the transportation simile of communication and focuses on the transmission vehicle *M* rather than on the transmission path. The notion of vehicle, however, allows us to capture the difference between technically mediated and technically unmediated communication. This difference cannot be expressed by the notion of channel since it abstracts over natural and non-natural media: whatever the vehicle is, it has to ride on a channel. Actually, any artificial device used as a medium in communication emits signals in the medium as substance sense. For instance, the medium of ink of paper emits light waves on the optical channel. Thus, the notion of channel and the notions of medium as substance vs. medium as device provide two different viewpoints on multimodal communication. The primary accomplishment of medium as device is to have a notion of transmission vehicles that are detached from the sign producers or the *in situ* of the communication situation. This accomplishment is taken up in the subsequent Subsection 2.3.

First, however, note that the physiological notion of modality proposed here diverges from the notion of modality held among logicians which goes back to the founder of semiotics, Charles Sanders Peirce (1839–1914). For Peirce (CP 2.382), modality consists of the four logical modes "possible", "impossible", "necessary", and "contingent". The modality of a sign, then, is determined by its "truth-value transparency". From this perspective, written language, for instance, has *low* modality, while a photograph has *high* modality. The Peircean conception of modality, unlike many of his other semiotic insights, has not gained wider acceptance.[7] We leave it open whether this is due to the status of signs "in the age of mechanical reproduction" (Walter Benjamin) or to the philosophical involvement of the Peircean account. We advocate modality as the sensual aspect of information in this article.[8]

2.3. Representations and medium as device

Medium as device is the result of replacing the natural medium as substance by some non-natural artifact. As seen in the previous Subsection 2.2 the primary achievement of artifacts used as media is to transfer ephemeral, evanescent communication signals into a more permanent form. This achievement can be realized in three ways:

- *physically*, by means of recording (on the comprehension side) or fabricating (on the production side) the material form of a sign;
- *notationally*, by means of a symbolic representation of a sign (notation is neutral with respect to production or comprehension);
- *annotationally*, a combination of the first two methods, where previously recorded communication instances are transcribed and possibly coded for further information.

Recording techniques are ubiquitous in everyday life. We use cameras to photograph sights and people, we record our outspoken thoughts by means of a dictaphone, and we capture temporal dynamic scenes with video cameras. These capturing devices can be used as artificial proxies for respective human sense modalities – and for that reason have been called "protheses" by Eco (1985/2000). Such input devices are introduced in Subsection 4. Some examples for technical devices that output forms as imprints on a material medium are given in subsection 6.2. A common example are computer screens that display textual and pictorial information. Accomplishing physical representations of signs is a precondition especially for anthropomorphic avatars that mimic human communication behavior – see subsection 6.1 for a virtual robot that is able to produce speech and gesture.

Notational systems provide a syntax and a semantics for the symbolic representation of signs: the syntax consists in symbols that stand for elements from the domain of signs to be noted and some rules for combining them. The semantic domain is the kind of sign system to be notated plus rules of mapping elements from the domain onto the symbolic inventory of the notation system's syntax (cf. Goodman, 1976). For instance, both spoken and written language are often represented in terms of alphabetical representation (the latter quite trivially so). A symbolic representation system for spoken language that suits better to the specific features of articulation is provided by the IPA (*International Phonetic Alphabet*[9]). In this vein, there are a couple of notation system specifically designed for certain nonverbal communication means. Sign languages can represented by the Stokoe notation (Stokoe, Casterline, and Croneberg, 1965) or HamNoSys, the *Hamburger Notationssystem für Gebärdensprache* (Prillwitz, Leven, Zienert, Hanke, and Henning, 1989). A notation system for body movements and an inventory for facial expressions has been devised by Birdwhistell (1970). Dance figures can be notated using *Labanotation* (Hutchinson, 1991), a notation system for body movements. There are a couple of coding schemes for gesture annotation: FORM (Martell, Osborn, Friedman, and Howard, 2002), an annotation scheme for conversational gestures (Kipp, Neff, and Albrecht, 2007), and a "morphological" coding scheme used in the SaGA (speech and gesture alignment) project (Lücking, Bergmann, Hahn, Kopp, and Rieser, 2010). Gaze annotation has been a part of the corpus work of Jovanovic, Akker, and Nijholt (2006).

Annotation is not only appropriate for data analysis and reconstruction of (multimodal) behavior (cf. Lehmberg and Wörner, 2007) but can also be employed for the generation of movements of animated avatars (cf. Martin, 2005; see for example Neagle, Ng, and Ruddle, 2004). Annotation data can be modeled by general annotation models such as the annotation graph model (Bird and Liberman, 2001) or the abstract corpus model (Brugman and Wittenburg, 2001) – see Wittenburg (2007) for an overview. More complex interrelationships can be captured by means of the Graph Exchange Language (GXL) (Winter, Kullbach, and Riediger, 2002). See also Stührenberg (2012), in this volume, on annotation standards.

There is a subtle, but very important difference between physical representation and notation. Recording or producing a sign are *token-related* procedures; the objects of the procedures are necessarily *sign occurrences*. A *type-related* way of representing signs is by describing them according to a notation system. A notation system allows to describe a sign that has not previously been recorded (the difference to annotation is simply that annotation is not used in this way) and will not be physically realized. A notational representation is type-related in such a way that different tokens that are similar in a respect set by the notation system will be notated in the same way. That is, a notation system is blind to those features of tokens that are not provided with notational means (a token is always underspecified under description (cf. Waisman, 1951)). A technical communication system that aims at generality has to be able to abstract away from formal differences between the letter tokens *a*, A, and a, and to recognize that they are all instances of the letter type "a". For a discussion of the prime importance of types over tokens in linguistics see Bromberger (1988).

2.4. Sign systems, icon, index, and symbol

Signs are distinguished into *icons*, *indices*, and *symbols* according to the relation between the sign's form and its meaning (see Allwood (2007) for motivating this distinction in the context of multimodal data). The three kinds of relations between sign vehicles and their objects have been inherited from Peirce (1867), who devised the following tripartition of signs:

> First. Those whose relation to their objects is a mere community in some quality, and these representations may be termed *Likenesses*.
> Second. Those whose relation to their objects consists in a correspondence in fact, and these may be termed *Indices* or *Signs*.
> Third. Those the ground of whose relation to their objects is an imputed character, which are the same as general signs, and these may be termed *Symbols*.

Later, "Likenesses" have been called "Icons", the name they are known by today.

The three kinds of signs play an important role in multimodal communication means. The ready means for displaying (iconic) or indicating emotions or attitudes are nonverbal means like facial expression of voice quality, which make up 80–90 % of the information exchanged in communication (Allwood, 2002). Iconicity and indexicality also are of special impact in gesture studies. Using the Peircean nomenclature, the gestural sign systems has been populated by the classes of indexical and iconic gestures, amongst others (cf. McNeill, 1992a). The distinction of different kinds of sign, especially the identification of non-symbolic sign, gives rise to a further, semantic benchmark of multimodal communication: **the meaning of nonverbal communication means need not be like the semantics of verbal predicates.**

However, the Peircean tripartition is often adopted in unaudited form, leading to an oversimplification of matters. Since the Peircean distinction is rooted in his *phaneroscopy*, that is, his theory of the elements of representation and intuition (*Vorstellung* and *Anschauung*[10])[11], you cannot – as is nonetheless usually done – get Peirce's object-related kinds of signs without buying their phaneroscopic underpinning, including the two other relationships of triadic signs, namely sign vehicle–interpretant and interpretant–object. Therefore, one could follow the conception of Keller (1995), which is far less metaphysically committing. According to Keller (1995), the sign vehicles of icons, indices and symbols serve different roles in their interpretation by a sign receiver. An icon triggers some association that helps to "read" the referent. An index is connected to its referent in a causal way, and therefore automatically points to its object. The form-meaning relation of symbols has to be learned, since there is nothing in the form of symbols that bears an inherent or associative relation to their referents.

In particular icons and the notion of similarity bound up with them have produced some objections. Burks (1949), Bierman (1962), Eco (1976), and Goodman (1976) argue that similarity is a far too loose notion to base the sign relation on. For in a certain respect everything becomes similar to everything else. The objections put forward, however, did not remain unchallenged. For instance, in his semiotic work Sonesson (1997); Sonesson (1998) tries to maintain a resemblance account to icons. In a related vein, Kauther and Müller (2000) point out that the associative potential of icons as proposed by Keller (1995) cannot be explained without reference to similarity. A workable philosophical account for informativity qua similarity has, for instance, been proposed in the form of a theory of depiction based on visual perception by Peacocke (1987).

2.5. Multiplicity in multimodal communication

Multimodal communication describes the phenomenon that information exchanged in a communication situation is perceived via more than one sensory channel. Since this conception of multimodality does not say anything about the number or kind of signs perceived, two kinds of multimodal communication can be distinguished:

– *Multiple access:* one sign is accessed in several sensory ways (for instance in simultaneously touching and reading Braille – if the text is also read aloud, there would be even a triple access to one sign); and
– *Multiple signs:* two or more signs are perceived simultaneously by different modalities. A common example would be co-occurring speech and gesture.

Further distinctions for *multiple signs*-communication can be made. Firstly, there can be combinations of multiple signs and multiple access. Suppose the perception of perfume probes in addition to the multiple access of a Braille text on fragrances. Secondly, if there are more than two signs involved in a communication exchange, it suffices that at least two of them are of different modality in order to speak of multimodality. Thirdly, multimodality covers all kinds of sign constitution, that is, aggregates that form cross-modal super-signs as well as couples of co-occurring signs. Multimodality is typically reserved for the latter case, while the former has been termed 'ensemble' by Kendon (2004) and Lücking, Mehler, and Menke (2008). Fourthly, multimodality is typically used for speech combining with at least one non-verbal means. The combination of two or more non-verbal means (excluding speech) can also be multimodal. We know of no systematic distinction, involving a terminological one, between the possible range of realizations of multimodality (and multicodality). However, branching points for such a distinction are outlined by the four enumeration articles above.

The various communication means usually do not occur consecutively, rather they occur at once. The interpretation of co-occurring information is guided by the following heuristics: **if multiple signs occur simultaneously, take them as one!** (Enfield, 2009, p. 6). On the cerebral side there is sustaining evidence that fused perception is bound up with parallel processing in the neural network (Engel and Singer, 2001). Note, that the simultaneity-rule is just a rule of thumb: channel-crossing bits belonging together in a single multimodal utterance might not overlap in time – see the respective remarks in the entry on *Synchrony* in Subsection 2.1 above.

A further dimension of multiplicity in multimodal communication concerns semantics. The basic question here is whether co-occurring signs present different information (*complementarity*), or whether one information is replicated between the signs (*redundancy*). In the case of speech and accompanying gestures it has been found that redundant and complementary gestures occur at equal proportion (Cassell and Prevost, 1996). The informational relationship between speech and gestures is influenced by whether there is a ready verbal description of the referent or not (Bavelas, Kenwood, Johnson, and Philips, 2002). The ability to describe things properly depends on individual spatial and verbal skills (Hostetter and Alibali, 2011). This finding traces the redundancy-complementary ratio in part back to individual facilities of speakers. Whether a gesture is redundant or complementary to its affliated speech is not an all-or-none matter; rather, a gesture may replicate certain semantic features of its affiliated speech but add certain others (Bergmann and Kopp, 2006). Thus, redundancy or complementary is a *gradual* relationship. A feature-based approach is the best method to explore complementarity between gesture and speech anyway, as is argued in the methodological comparison by Gerwing and Allison (2009). Redundancy or complementarity has also been shown to influence humans' reactions to multicodal behavior of avatars in HCI (Buisine and Martin, 2007) and therefore is a multiplicity level relevant for technical communication systems. However, since the "meaningfulness" of different types of signs usually is grounded in different ways – cf. Subsection 2.4 above – it is questionable whether one can talk of redundancy and complementarity of information at all (see e.g. Lücking (2011, pp. 64f)).

We would like to emphasize that multimodality in a sensory understanding is strongly related to barrier-free communication (see Kubina and Lücking (2012) in this volume). For what a sign makes barrier-free is that it can be accessed in several ways. If one modality is blocked, be it for reasons internal or external to the sign perceiver, the sign can be accessed via another modality – recall the afore-mentioned Braille example as an illustration. This example also illustrates that barrier-free communication is not just given in advance, it frequently has at first to be made possible *inter alia* by multimedia realizations of a single sign – see Lazar, Allen, Kleinman, and Malarkey (2007) on screen readers for blind users and some usage diffculties.

2.6. Benchmarking: A brief summary

We achieved the following list of benchmarks that multimodal technical communication systems have to fulfill in order be linguistically adequate:

- The primacy of speech (Subsection 2.1);
- The modality–channel/media interface is I/O asymmetrical (Section 2);
- Multimodal information is fused (Section 2);
- Non-verbal interaction can be related to each level of linguistic constituency (Subsection 2.1);
- Multimodal information fusion is guided by synchrony, as an overridable default (Subsection 2.5);
- The semantics of non-verbal communication means need not be like the verbal one (Subsection 2.4).

Of course, this rough outline might be extended by more specific points, especially in the context of multimedia fission, for instance, in the area of models of speech-and-gesture production. However, such more specific points are often themselves still object of research (see e.g. Alibali, Kita, and Yound, 2000; Kita and Özyürek, 2003; Kita, Özyürek, Allen, Brown, Furman, and Ishizuka, 2007; Kelly, Manning, and Rodak, 2008) and not settled within the scientific community. The features identified in the previous list, however, to our minds have to be fulfilled within a system that claims to be multimodal and communicative. Accordingly, the list is used in Section 7.1 as a reference point for assessing technical systems and applications introduced in the following sections.

3. Outline of technical applications for multimodal communication

Compared to the processing of non-verbal communication signals, the technology for detecting and recognizing speech is already quite advanced. Most approaches are based on Hidden Markov models (Baum and Petrie, 1966; Baum, Petrie, Soules, and Weiss, 1970; Rabiner, 1989; see also Paaß (2012) in this volume). Challenging areas of current research are speaker-independent speech recognition, the separation of speech from speakers within a group or from general noise, or the identification of speakers or the detection and interpretation of non-verbal aspects of speech, which is, for example, used for stress-detection (Waibel and Lee, 1990). Speech recognition technology has already entered the market and even mainstream operating systems, such as Windows VISTA or Windows 7, already include at least speaker-dependent voice recognition technology. The results of speech recognition systems are often dramatically improved when using high-quality hardware, such as specific sound cards and head-mounted microphones. In addition, more recent smart phones are equipped with an intelligent voice control (e.g., Siri[12]).

If speech recognition is used in a multimodal set-up, the speech recognizer should be verbose in its output and not restrict itself to providing the most probable answer only. During multimodal fusion, competitive hypotheses can often be decided upon later based on additional crossmodal information. However, it is crucial to have access to the hypotheses as quickly as possible. Early approaches to multimodal fusion used a time threshold to wait for the end results of all multimodal recognizers (Oviatt, DeAngeli, and Kuhn, 1997). Newer approaches use an incremental processing strategy, trying to interpret intermediate results as early as possible (Johnston and Bangalore, 2000; Latoschik, 2002; Sun, Chen, and Chung, 2006).

The human body and the human senses already equip people with various channels that can be used for communicative means. Machines, on the other hand, have at first to be equipped with devices that facilitate to produce or to receive signals on different channels. The advent of non-verbal input and output in human-machine interaction was the development of graphical interfaces. The possibility to represent and manipulate non-symbolic content, such as images and diagrams, set the stage for new input devices, such as computer mice and digitizers, which allowed for manual gesture input (Shneiderman and Plaisant, 2009). Interaction with computers since then has been shaped by a quite pragmatic view. Non-verbal input by the user is predominantly treated as directed manipulations, often mimicking a physical cognate, such as placing a cursor, moving icons, drawing lines, etc.

To focus the following overview, we will concentrate on two modalities: manual gestures, which are performed by the hands and arms, and eye gaze. We will omit the work on interpreting full body postures, facial expressions, force-based interactions, or, for example, physiological signals. We will also only consider technical applications where at least two modalities are integrated and thus multimodal fusion is relevant. Interaction with the iPhone, for example, will not be considered, as it is able to accept input with different modalities, such as spoken and textual speech and multi-touch gestures, but it does not integrate the input from these modalities.

Regarding the production of multimodal output, we will focus mainly on the work on *Embodied Conversational Agents* (ECAs) (Cassell, Sullivan, Prevost, and Churchill, 2000), be it virtual or robotic, because this work is closely related to the scientific focus of linguistics: ECAs are targeted at human-level communication skills using human-like modalities (Traum and Rickel, 2002, see also Kopp and Wachsmuth (2012) in this volume).

4. Input devices

4.1. Devices for detecting manual gestures

In computer science, manual gestural input refers to any continuous input which is operated by the hands and has a spatial extension. This excludes input on a keyboard, but includes writing something down on a digitizer. Typical gestural input devices are the computer mouse, the touchpad, the digital pen or multi-touch surfaces (Kurtenbach and Hulteen, 1990). The most common interactions with these devices are direct manipulations, such as drag-n-drop or multi-touch zoom. It is arguable whether these interactions are counted among communicative gestures. Interactions based on symbolic gestures are less common, examples are mouse gestures which are supported by some internet browsers or the pen-based symbolic gestures recognized by Microsoft's tablet operating systems. These types of gestures, however, are technical gestures and not communicative gestures in the sense we are interested in.

A classical scenario that already comes under the heading of 'multimodal human-computer interaction' is a combination of spoken utterance and pen input. This kind of communication, however, usually is strictly command-bound and merely loosely related to natural conversation. An example for such a multimodal human-computer interface (HCI) is the QuickSet tool of Johnston, Cohen, McGee, Oviatt, Pittman, and Smith (1997) where a pen input on a touchscreen is interpreted as a location command ("put-it-here" or "move-that-there").

For recognizing gestures, one needs to perceive the movements of the hands (including fingers) and arms. The technical term for this process is 'tracking' (Blackman and Popoli, 1999). There are several approaches for tracking hand and arm movements (a sample of systems is depicted in Figure 3), which can be roughly assigned to one of two categories:

1. 'Marker-based Tracking Systems' rely on the detection of specific markers by tracking devices. If the system operates 'outside-in', the markers are attached to the relevant body parts while the tracking devices are observing a certain interaction space. Many optical tracking systems operate that way, such as the systems from ART (ART, 2011) and Vicon (Vicon Motion Systems, 1984). 'Inside-out' systems work the other way round in that the markers are used as landmarks in the environment and the tracking devices are attached to the relevant body parts. A well-known example for inside-

(a) Hand Tracking: Immersion's CyberGlove

(b) Inside-Out Magnetic Tracking: Ascension's Flock of Birds

(c) Outside-In Optical Tracking: DTrack System by ARTracking GmbH

Figure 3. A collection of tracking systems. The CyberGlove (Virtual Technologies (Virtex), 1991) detects hand form configurations based on resistive bend-sensing technology. The Flock of Birds (Scully and Blood, 1986) detects the position of the wired sensor in a magnetic field. The DTrack (ART, 2011) system is an optical tracking system with light-weight markers, here for detecting position and orientation.

out tracking is the Wii-Remote developed by Nintendo. Several technologies can be used for tracking, such as visual tracking in different domains (for example infra-red) and electromagnetic tracking.

2. 'Marker-less Tracking Systems' do only require tracking devices. They also come in two different flavors. The 'inside-out' systems for example are small devices that are attached to different body parts. They can detect accelerations or changes in the magnetic field. These systems, however, are only capable of detecting relative movements, not absolute positions. 'Outside-in' systems observe the movement of the body from a distance, most systems operate in the visual domain. Examples for outside-in systems are the Microsoft Kinect (Microsoft Kinect SDK, 2010) or the Swiss-ranger cameras by MESA Imaging AG (MESA Imaging AG, 2006).

The choice on which system to use for recognizing manual gestures depends on the requirements and personal preferences. Marker-based tracking systems today offer the highest accuracy and precision for these motions (on these notions see Paaß (2012) in this volume). They are also very robust. If large areas are to be covered, inside-out tracking systems are less expensive, as typically the markers are less expensive than the tracking devices and it is thus cheaper to cover the area with markers than with tracking devices. If the interaction space is small (several meters), outside-in tracking might be more convenient to use, as the markers are typically more light-weight and easier to attach than the tracking devices. Marker-based tracking systems typically provide absolute positions and depending on the configuration of the supported tracking targets, may also offer information about the orientation of a marker (ART, 2011).

Marker-less tracking systems that operate inside-out are attractive for mobile use, where the area of operation is unrestricted. However, as has been pointed out, they provide relative data which, depending on the task, is often diffcult to interpret reliably. The least obtrusive method is the outside-in marker-less tracking. Here the user can enter the interaction space without any markers being attached. There might be, however, a calibration procedure. Such systems provide absolute positions which are easier to handle than relative ones. First products, such as the Microsoft Kinect, have hit the consumer market, so we expect much progress regarding the interpretation of multimodal gestures using unobtrusive outside-in marker-less tracking.

4.2. Devices for detecting eye gaze

Interaction via eye gaze, although already a classic of HCI (see Levine, 1981; Hutchinson, White, Martin, Reichert, and Frey, 1989), is less visible than interaction via gestures. Eye tracking is the observation of eye movements by technical means, today usually an optical camera system. Eye tracking is commonly used in order to determine the *point of regard*, that is, the part of a medium looked at. An eye tracking research niche is *gaze-typing* (Ten Kate, Frietman, Willems, Ter Haar Romeny, and Tenkink, 1979), that is text input by means of gaze. Gaze-typing has been applied early as a kind of interaction for impaired (physically as well as occupied) persons. Apart from this special purpose application, the most common uses for eye tracking are as measuring device in scientific studies or usability tests.

Approaches to interaction by gaze have been mainly remained in research laboratories (Bolt, 1981; Levine, 1981; Levine, 1984; Starker and Bolt, 1990; Jacob, 1993; Jacob and Karn, 2003). One reason for that is to be found in the great cost of camera systems capable of image processing of the required level of detail and precision. There are, however, some projects working on solutions for webcams and home computers, for instance the European network project *COGAIN*[13] (COmmunication by GAze INteraction, 2004–2009, EU FP6 San Agustin, Skovsgaard, Hansen, and Hansen, 2009).

Currently, eye tracking is primarily done using optical outside-in marker-less tracking. There are, however, some approaches where small pigment markers are applied to the eye. While accuracy and overall system latency are improved this way, this probably imposes problems of acceptability. Regarding the positioning of the tracking cameras, there are currently two options: they are either head-mounted (HM) or remote systems, as shown in Figure 4 (Duchowski, 2007). Similar to the tracking of hands and arms, the remote tracking systems are the least obtrusive ones. However, the interaction space is very re-

(a) Remote Eye Tracker: LC Technology's Eyegaze System

(b) Head-Mounted Eye Tracker: SMI's EyeLink II

(c) Light-weight HM Eye Tracker: Arrington Research's ViewPoint System

Figure 4. A collection of different eye tracking systems based on marker-less optical tracking.

stricted. The remote systems are well suited when the participants are seated, which might be at a still table, but also in cars or planes. They are especially interesting for studying human-human interactions, for example, dialogs, as both participants have an unrestricted view of the interlocutors face. When interacting with a machine, however, functionality comes first and the advantage of a more-or-less unrestricted interaction space increases the attractiveness of head-mounted systems.

5. Integration of multimodal information

5.1. Stages of uni-modal processing

Before we are discussing the integration of the different chunks of information contributed by several modalities in the next section about multimodal fusion, we will shortly describe the stages of uni-modal processing. Our main focus will thereby be on the processing of manual gestures and eye gaze.

First, 'raw data' needs to be collected from the input devices, such as the ones described before. The type of the data collected depends on the method used for tracking. In the case of an optical tracking system, it could be a still video image of one of the cameras. This data will then be preprocessed to reduce noise or to emphasize certain aspects. When detecting eye gaze, for example, filters are used to help finding the pupil and the glint. This 'feature extraction' process will reduce the raw data to a set of relevant features (Guyon and Elisseeff, 2003). These features could be, for example, detected edges or points. In a subsequent ‘classification’ step, these feature sets will then be classified into different categories. In the case of the eye-gaze detection, the detected visual features will be classified as being part of the pupil or part of the glint, otherwise they will be rejected.

This pattern of a three-stage process (raw data collection, feature extraction and classification) can be applied to several levels of uni-modal processing operating on different levels of data abstraction. For example, when using an optical tracking system for gestures, the first three-stage process will take a single video image as raw input and classify bright pixels into potential marker positions as 2D coordinates. The 2D coordinate sets provided by several cameras are then the raw data input for the next three-stage process, which will classify them into 3D positions and eventually 3D orientations. Then again, these positions will be taken as raw input and in the end they will be classified into a 3D posture (or sub-posture) of the user interacting with the system. A time series of such postures, again, can be input for another process chain which in the end leads to a classification into gestures such as hand-waving or pointing, just to name a few (McNeill, 1992b; Kendon, 2004).

The description of the processes involved from acquiring a single image to the classification into a label for the gesture observed by an optical tracking system shows that the collected information chunks are transformed from one representation into another. Information which is a set of features on one level forms raw input for the next level. The description given above also suggests a linear chain of the processes, which is not necessarily the case. The processes can actually fork so that generated classifications form the input of two or more following process chains. Some architectures also use backwards feedback from later processing steps to inform earlier processing.

5.2. Levels of multimodal fusion

Multimodal fusion is the process of integrating data from different modalities which aims at creating a comprehensive description suitable for further interpretations, for example, in the dialog manager. Multimodal fusion has been applied on several levels of abstraction, starting with very low-level representations generated early in the uni-modal processing chain up to the high-level representations generated towards the end of the chain.

Sometimes, technically there is also the need for multi-sensor fusion. For example, if arm and hand gestures are to be assessed, tracking might be done by combining a marker-based tracking for the back of the hand and the elbow with a data glove to capture fine-grained finger movements (Fröhlich and Wachsmuth, 1998). In this case the sensors provide complementary information.

One of the progenitors of multimodal HCI was Bolt (1980). His “Put-that-there”-system integrates speech and a simple pointing gesture. However, the integration rule was highly stereotyped. Speech input had to be very slow and interrupted by frequent pauses in order to be manageable by the speech recognizer. Above all, even multimodal input had to be given in a sequential way. Nonetheless, Bolt's system demonstrated how appealing multimodal HCIs could be and inspired a lot of researchers to proceed on this way. Just one year after his seminal work, Bolt himself presented more innovative work on multimodal human-computer interaction, but this time in the area of gaze: the "World of Windows" (Bolt, 1981). In his scenario, single video windows on a wall of video windows could be replayed by looking at them.

A combination of speech, gaze and gesture input has been interpreted by Koons, Sparrell, and Thorisson (1993) as input for a 2D interface. They used frame-based representations to integrate multimodal information. Each modality was interpreted by a specific parser, which created a modality-specific frame that later was integrated. In their first prototype, they concentrated on deictic gestures (manual and gaze pointing) in an application scenario where objects had to be placed and moved on a map. Iconic gestures were interpreted in a second prototype to manipulate 3D objects (Sparrell and Koons, 1994).

The described systems are able to recognize and interpret a subset of the multimodal non-verbal communicative signals humans produce as commands. They are not communicating with humans on a comparable level. Considering multimodal non-verbal communication, *Embodied Conversational Agents* (ECAs) (see Cassell, Sullivan, Prevost, and Churchill, 2000) are the most advanced kind of human-machine interfaces.

Besides deictic gestures, turn-taking gestures have quite early been interpreted by ECAs (Thórisson, 1997; Cassell, Bickmore, Billinghurst, Campbell, Chang, Vilhjálmsson, and Yan, 1998; Traum and Rickel, 2002; Lessmann, Kranstedt, and Wachsmuth, 2004).

For the interaction with a virtual agent, Oh, Fox, Kleek, Adler, Gajos, Morency, and Darrell (2002) have implemented a look-to-talk interface to trigger the conversation. Besides eye gaze, the orientation of the head can provide useful information, for example, to facilitate mutual understanding. This has been shown by Nakano, Reinstein, Stocky, and Cassell (2003) in a dialog with the ECA MACK. Pfeiffer-Lessmann and Wachsmuth (2008) describe how alignment between a human and the ECA Max can be achieved by monitoring the user's gaze and establishing joined attention. Compared to gestures with arms and hands, however, the interpretation of the interlocutor's gaze is still underrepresented. This will change in the future, as eye tracking systems have become smaller, less obtrusive, and cheaper.

5.2.1. Early multimodal fusion

Multimodal fusion in early stages of the processing is also said to operate on the 'signal level', if the representation of the input is very close to that of the input device. On this level, fusion is primarily based on timing, that is multimodal events which have a certain correlation in time are fused. An exact co-occurrence is often not required, rather patterns that have been extracted during a learning phase or through explicit modeling are used. A second basis for fusion can be spatial co-location, that is multimodal events which have been detected within the same area are fused. An example of this might be the detection of a speaker in group discussions based on audio triangulation and the detection of lip movements (Liew and Wang, 2008). Often, spatial fusion is done implicitly, as markers and sensors are explicitly attached to the spatial area of interest, thus anticipating spatial fusion by design.

Targeting at applications in video telephony, Vermaak, Gangnet, Blake, and Perez (2001) use a particle filtering approach to fuse multimodal data on audio and video to identify and track the speaker. Although not explicitly stated by the authors, their approach enables the early detection of turn-taking signals and the assignment of conversational roles. Besides probabilistic models, also neural networks are used in early multimodal fusion. An example is the work of Meier, Stiefelhagen, Yang, and Waibel (2000), who improved speech recognition by early multimodal fusion of the audio signal with visual information about the lip movements. Their fusion algorithm works on the basis of a Multi-State Time-Delay Neural Network.

5.2.2. Multimodal fusion at the level of features

Fröhlich and Wachsmuth (1998) track hand movements and hand shape by means of data gloves within a magnetic tracking system and classify the captured data into a feature set described by the symbolic notation system for sign languages *HamNoSys* (Hamburg Notation System Prillwitz, Leven, Zienert, Hanke, and Henning, 1989). This feature set provides the basis for a top-level processing of gestures in terms of a gesture lexicon, which in turn assign each gesture (in HamNoSys representation) a meaning.

In addition to the discrete symbolic interpretation, Latoschik, Fröhlich, Jung, and Wachsmuth (1998) provide a continuous interpretation for direct manipulation. This line of research finally led to the development of ProSA (*Patterns of Sequences of Attributes* Latoschik, 2001a), a framework for gesture detection. ProSA allows to feed input sensory data into nodes of a processing network. The nodes perform different detection tasks. Edges connecting the nodes not only determine the flow of data, they also determine the chronological order of the processing steps. In this way, the ProSA framework facilitates the building of complex recognizers for hand shapes and movement trajectories.

Simultaneously, in ProSA the three stages are dissolved, as every result of every single processing step is immediately made available in the data-flow network as input for any other process. This immediacy is required for direct manipulations, as is highlighted by the example of the rotation of objects: once detector nodes in ProSA have detected a gesture constituting a request for rotation, a direct coupling between early detectors for the angular rotation of the gesture and the representation of the object in the scenegraph of their virtual reality environment is established. Every further rotational movement of the hands is then closely tied to the rotation of the object in the application, unless the rotation gesture is relaxed. Detecting and acting can be tightly coupled in ProSA.

ProSA is strongly connected to a tATN (temporal augmented transition network, Latoschik, 2002), a framework for multimodal integration. Selected outputs from ProSA are pulled up to a symbolic representation which is processed by the tATN. A temporal ATN, like the name suggests, extends common ATNs by explicit time-related aspects, that are needed in order to capture the synchronization between signals in various modalities. A state of a tATN is a representation of a multimodal grammar modeling a multimodal input. New input leads to new states. Usually, several conflicting hypotheses (states) are represented in a parallel way, before finally the best-fitting one is made out.

5.2.3. Semantic multimodal fusion

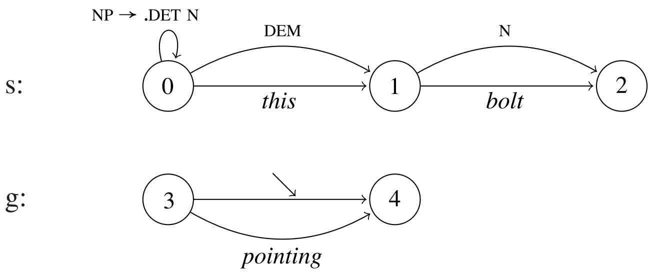
In the course of the project QuickSet, Johnston, Cohen, McGee, Oviatt, Pittman, and Smith (1997) used multimodal fusion by unification of typed feature structures to interpret speech together with pen gestures. Both speech and gesture are assigned a representation in terms of *Attribute-Value Matrices* (AVMs) by speech and gesture recognizers. With this approach, their multimodal fusion works on the feature level and on the semantic level simultaneously, as illustrated by the following example.

Users interacting with the QuickSet system can point (at *X*) and by doing so they introduce a certain point in space represented as a *latitude-longitude* coordinate pair. This locational function of pointing is captured in the representation given in (1), which shows that the semantics (*content*) of an object of the category (*cat*) named *spatial_gesture* is a point in space (Johnston, 1998):

(1)
$$\begin{bmatrix} \text{cat:} & \textit{spatial\_gesture} \\ \text{content:} & \begin{bmatrix} \text{fsType:} & \textit{point} \\ \text{coord:} & \textit{latlong}(x,y) \end{bmatrix} \end{bmatrix}$$

The AVM-grammar formalism rests on a multimodal chart parser. A multimodal chart extends a conventional chart in that the former covers channel-crossing edges defined in terms of sets of identifiers of gestural (g) and speech (s) terminals:

(2)



The basic rule allowing to “bridge” between the modalities is the *basic integration scheme* depicted in (3):

(3)

$$\begin{bmatrix} \text{lhs}: & \begin{bmatrix} \text{cat}: & \textit{command} \\ \text{modality}: & \boxed{2} \\ \text{content}: & \boxed{1} \\ \text{time}: & \boxed{3} \end{bmatrix} \\ \text{rhs}: & \begin{bmatrix} \text{dtr1}: & \begin{bmatrix} \text{cat}: & \textit{located\_command} \\ \text{modality}: & \boxed{6} \\ \text{content}: & \boxed{1}[\text{location } \boxed{5}] \\ \text{time}: & \boxed{7} \end{bmatrix} \\ \text{dtr2}: & \begin{bmatrix} \text{cat}: & \textit{spatial\_gesture} \\ \text{modality}: & \boxed{5} \\ \text{content}: & \boxed{9} \\ \text{time}: & \boxed{10} \end{bmatrix} \end{bmatrix} \\ \text{constraints}: & \left\{ \begin{array}{l} \text{overlap}(\boxed{7},\boxed{10}) \vee \text{follow}(\boxed{7},\boxed{10},4) \\ \text{total-time}(\boxed{7},\boxed{10},\boxed{3}) \\ \text{assign-modality}(\boxed{6},\boxed{9},\boxed{2}) \end{array} \right\} \end{bmatrix}$$

The AVM for the integration scheme is stated very closely to a rule of a context free grammar of the form *lhs* → *rhs*; the right-hand side (*rhs*) is made up of two constituents, namely *dtr1* and *dtr2* (see (3)).

Thus, mapping the rule to a tree, they are the daughters of their mother constituent on the rule’s left-hand side (*lhs*). The *rhs*-part of the AVM-structure is made up of a verbal *located command* and a *spatial gesture*. The gesture determines the location value of the word’s content. The mother structure (*lhs*) then is a complete multimodal command. The cross-codal integration is constrained by a set of restrictions given as the value of the feature *constraints*. Most notably, co-occurrence constraints are expressed as temporal requirements, see the use of tags [7] and [10] in (3).

Another example for multimodal fusion on the level of semantics is the reference resolution process developed by Pfeiffer and Latoschik (2004). They represent the contributions of individual modalities, such as speech, eye gaze and manual pointing gestures, as constraints. During multimodal fusion, they construct a fuzzy-based *Constraint-Satisfaction Problem* (CSP), which is a full representation of semantic constraints detected in the multimodal expression. In a second step, this is augmented with pragmatic constraints, such as preferences for objects that are easily perceivable or within grasping reach of the addressee. Finally, this CSP is solved over the set of objects in the current context.

While the systems described in this section so far concentrated on deictic gestures, there are also approaches that integrate other kind of gestures, for example, iconic gestures. We have already mentioned Sparrell and Koons (1994) for a command-like multimodal interface capable of recognizing speech in conjunction with iconic gestures.

5.2.4. Conceptual multimodal fusion

Based on his own empirical findings and previous work, for example, by Lang (1989) on object schemata for the semantic modeling of spatial language, Sowa (2005) developed a representation format for shapes of connected rigid objects called *Imagistic Description Trees* (IDTs). His model for multimodal fusion of speech and iconic shape-related gestures has four processing levels: phonological/kinetic, syntactic, semantic and conceptual/imagistic. The integration of the different modalities is done on the syntactic and on the conceptual level. On his syntactic stage, the integration is done solely based on temporal co-occurrence of speech and gesture. The integration is thereby primarily a chunking of information packages expected to refer to the same idea. There is no common representation format used that would embrace both speech and gesture. The primary multimodal fusion happens at the conceptual stage, after speech and gesture content has been used to create two descriptions of the shape-related contributions of each modality as independent IDTs. The conceptualizer in this stage has access to a spatial working memory, where a current IDT is kept and incrementally updated. The IDTs derived from both modalities are then unified with the current IDT to a common representation independent of the originating modality. The applicability of Sowa's approach is demonstrated in a system for reference resolution (Sowa, 2006). In addition, an example for gestural imi-

```
<definition>
  <utterance>
    <specification>
      Take <time id="t1"/> this bar <time id="t2"/>!
    </specification>
    <behaviorspec id="gesture_1">
      <gesture>
        <affiliate onset="t1" end="t2"/>
        <function name="refer_to_loc">
          <param name="refloc" value="Loc-Bar_1"/>
        </function>
      </gesture>
    </behaviorspec>
  </utterance>
</definition>
```

Listing 1. Specification of a deictic expression with speech and manual pointing gesture in MURML.

tation demonstrates that the IDTs can also be used for the production of multimodal output. However, Sowa's imitation game only uses gestural input and output, which leaves the question open, on how the distribution of content between modalities should be handled in his model. This problem is taken up in Section 6 below.

5.2.5. *Pragmatic multimodal fusion*

The level of pragmatics is one of the highest levels of multimodal fusion. One example of pragmatic fusion is the resolution of multimodal deictic expressions to objects in the visual context used by Latoschik (2001b). Different modalities such as speech and manual gestures contribute to this process. The results of the independent uni-modal processing of the individual contributions to the deictic expression is represented as sets of rated candidates, one set for each modality. The late pragmatic multimodal fusion then classifies the candidates of all modalities and provides a single set of candidates as a result. The highest ranked candidate will then be chosen as the referent (see also Pfeiffer and Latoschik, 2004).

Note that fusion on one level does not prohibit a subsequent fusion on a second level. Rather the opposite is the case. Take for example the fusion of manual gesture and speech. First, it makes sense to fuse on the signal level to group speech and gestures that happen at the same time and place. Second, after a unimodal processing, the fusion at the pragmatic level provides the candidate of a detected deictic expression.

6. Multimedia fission

6.1. Example: Speech-gesture production

A precondition for multimodal fusion as well as multimedia fission is a representation format that allows for an integrated representation of multimodal content – be it as the goal processing (fusion) or as the starting point (fission). This has led to the development of markup languages for multimodal interaction, such as the *Extensible Multi-Modal Annotation markup language EMMA*; Baggia, Burnett, Carter, Dahl, McCobb, and D. 2009) defined by the W3C, which primarily targets interactions with web interfaces. For human multimodal interactions, Kranstedt, Kopp, and Wachsmuth (2002) developed a markup language format called *Multimodal Utterance Representation Markup Language* (MURML). MURML is able to represent utterances consisting of manual gestures, speech, eye-gaze and facial expressions. MURML is an XML dialect that allows to specify the contributions of different modalities to a single utterance – see Listing 1.1 for an example. The contributions are synchronized by anchors in the text to be verbalized which can then be taken up by the non-verbal elements of a multimodal contribution.

MURML has been employed to steer the multimodal communication behavior of the virtual agent Max (Kopp, Jung, Lessmann, and Wachsmuth, 2003). Max' gesture planning and realization has extended to iconic gestures, starting from Kopp, Sowa, and Wachsmuth (2004). The major challenge in designing a serious system that is able to do multimedia fission and to output speech and iconic gesture is to handle the iconicity-part in a proper way. As discussed above, iconic signs are special due to their form bearing the key to interpreting them. For this reason, early attempts to model multimodal output that rests on lexical storage, that is, on a conventional form-meaning mapping of gesture information (for example, REA *Real Estate Agent* by Cassell, Bickmore, Billinghurst, Campbell, Chang, Vilhjálmsson, and Yan, 1999). The lexical approach to iconic gestures is given up in the NUMACK system developed by Kopp, Tepper, and Cassell (2004). Here, IDTs – Sowa (2005), cf. Subsubsection 5.2.4 – are used as an intermediate representation format bridging between gesture morphology and "spatial meanings".

Continuing the line of research on speech-gesture output, Bergmann and Kopp (2009) developed a multimodal production system in terms of Bayesian networks (Pearl, 1985) that is trained on the richly annotated empirical data from the SaGA corpus (Lücking, Bergmann, Hahn, Kopp, and Rieser, 2010). Their model, *GNeTIc* (*Gesture Net for Iconic Gestures*), employs a Bayesian Decision Network (Zhang, 1993), that is a probabilistic Bayesian network enhanced by rule-based decision nodes. The general architecture of GNeTIc comprises four dimensions found to be relevant for the production of an iconic gesture, namely information state of the discourse context, features of the referent to be depicted, the kind of the previously produced gesture, and morphological gesture features. Given the message to be produced, GNeTIc determines – via rule-based as well as probabilistic decisions – whether a gesture is output at all and if so, how it is performed. The result of the Bayesian decision network is then passed to an output device for embodied conversational agents.

6.2. Devices for multimodal output

There are at least two types of output devices for displaying gestures. The first category is comprised of those devices using a physical body and, for example, includes robots (Breazeal and Scassellati, 2000; Beuter, Spexard, Lütkebohle, Peltason, and Kummert, 2008) and androids (Ishiguro, 2007). The second category uses visual projection technology to display gestures using articulated bodies, such as animated characters. A very advanced kind of display are embodied conversational agents (ECAs, see Cassell, Sullivan, Prevost, and Churchill, 2000). These ECAs not only have a visual appearance, they also often include a complete multimodal dialog system, capable of interpreting and producing two or more of (lip-sync) speech, gestures and other non-verbal cues, such as facial expressions.

Jack Presenter was one of the first ECAs with human-like 3D appearance using speech and manual gestures (Noma and Badler, 1997), but his production of non-verbal signals could only coarsely be synchronized to speech. The Steve-system (Johnson and Rickel, 1997) was developed at about the same time and offered similar capabilities, for example, head movements as non-verbal signals. Gandalf (Thórisson, 1997) was especially targeted at multimodal real-time interactions. Maybe it was to this end that the geometry of his visual appearance was quite simple, only a head and a hand in 2D, without a connecting body. REA (Cassell, Bickmore, Billinghurst, Campbell, Chang, Vilhjálmsson, and Yan, 1999) improved upon this work and provided a temporal synchronization between speech and gestures. Later ECAs, such as Max (Multimodal Assembly eXpert, Kopp, Jung, Lessmann, and Wachsmuth, 2003), provided synchronized speech and gestures, together with modulated intonations and mimics.

To exhibit a plausible behavior, eye gaze patterns have been modeled for ECAs to support turn-taking (Thórisson, 1996), as a function of discourse structure (Torres, Cassell, and Prevost, 1997) or depending on conversational state, cognitive state and visual context (Lee, Marsella, Traum, Gratch, and Lance, 2007).

7. Conclusion

Virtual embodied agents as the most ambitious representative for multimodal technical communication systems should cover the multifarious panoply of communication means known from human face-to-face interaction (see Kopp and Wachsmuth (2012) in this volume). ECAs and further applications that are able to process multimodal input or capable of outputting non-verbal behavior, have been covered by the previous sections. To complete this chapter, we want to appraise the systems in light of the benchmarks found underway and collected in Subsection 2.6. This is done in the next Subsection 7.1. In the subsequent Subsection 7.2 we try to provide some hints onto where multimodal technical communication might go.

7.1. Brief assessment of multimodal technical communication systems

The reference points for assessing multimodal technical communication systems are the list of benchmarks compiled in Subsection 2.6.

Most multimodal technical communication systems capture the I/O asymmetry merely in a trivial way. Since they usually are *either* processing *or* production tools, the respective benchmark does not apply in a sensible way. The possibility for I/O asymmetry only arises in a system that is capable of both, multimodal fusion and multimedia fission.

The fusion of multimodal information is in the case of humans to be conceived as semantic integration: the multimodal input is processed in such a way that it gets an encompassing or unified interpretation. As we have seen in Section 5, semantic integration is not the only level technical systems operate upon. On a pre-semantic level, the challenge for applications is to provide a unified representation for multimodal signals (that subsequently can be used for integrated processing higher order). Thus, the general form of the fusion problem is that of a representation problem. Fusion of multimodal input is accomplished, if the data streams of the signs involved can be captured within a single representation framework. Such a unifying representation framework can apply at various levels in multimodal processing, from basic data fusion up to conceptual integration. Since communication is bound up with at least the semantic level, systems that employ signal-based integration can be seen as a preliminary stage on the way to higher-level, ones like, say, ECAs. Note, that the representation problem also prevails in theoretical-linguistic accounts to nonverbal signs, such as gestures. A precondition for giving a formal semantic reconstruction of speech-gesture-samples is the availability of a representation format that can be applied to gestures. Seen from a representationalistic point of view, the problem of multimodal fusion is equal to the problem of integrated representation. Only if integration is done on at least a semantic layer, we want to rate a technical application as being communicative.

The first benchmark states that within an integrated representation not all parts are of equal rank: language is the chief sign system. In multimodal fusion, nonverbal information is integrated *into* the verbal one. The *raison d'être* for the primacy-of-speech benchmark is the dominant role speech plays in human communication. The contributions of other signal systems are interpreted in the context of and in relation to the co-occurring portions of speech (under the regiment of the subsequent benchmark, viz. temporal contiguity). Of course, only multimodal systems that include a language modality can comply to the respective benchmark. If they do, the relation between speech and other non-verbal communication means is often quite epiphenomenal. For instance, the QuickSet system of Johnston (1998) (see Subsection 5.2.3) exploits the command-like interpretation of user input in such a way, that purely nonverbal inputs are possible in principle: two "pointing gestures" (pen strokes) are processed as acting out a "put-that-there" order, even if no respective speech act is produced. Thus, the interpretive context for the gesture is not speech, but the command-driven architecture of the QuickSet system. Even the most advanced technical communication system, viz. ECAs, treat gesture and speech as epiphenomena. The reason is, that currently there is no semantic integration between input on the gesture and input on the speech modality; the integration scheme provided by, for example, MURML (Kranstedt, Kopp, and Wachsmuth, 2002) (see Subsection 6.1 above), only allows for a shallow combination of speech and gesture, which nevertheless suffices to capture the synchrony benchmark, since production time is the decisive information that triggers fusion. Note also that the ECAs briefly introduced above are *production* systems. That means that the ECAs can make use of the processing direction in such a way that they start from a single, unified meaning representation that becomes modally divorced in the production process, thereby circumventing the integration problem.

In Subsection 2.4 it has been discussed that different signs, even different aspects of one sign, can be decoded in three different ways, namely symptomatically or indexically (index), associatively (icon), and conventionally (symbol). However, the former two (icons and indices) have not been taken too serious by technical applications. Take for instance a touchpad input interface. The user can operate the operating system by means of finger strokes and "gestures", that is, by signals that are iconic and indexical in nature. However, the continuous input (at least continuous in the boundaries set by the physical display) is mapped onto a discrete set of quasi-lexical input repertoire, that is on a set of "gestural" symbols the touchpad is able to process. A related simplification has been carried out by early speech-and-gesture capable avatars that rest on a "gesticon" (Cassell, Bickmore, Billinghurst, Campbell, Chang, Vilhjálmsson, and Yan, 1999).

At the bottom line, the rating of state-of-the-art multimodal technical communication systems against a broad semiotic background results in a mixed picture: on the one hand there is a wealth of devices and applications dealing with non-verbal input or output. In particular, considerable technical effort is undertaken to exploit and to produce signals that encounter the devices on various channels and are transmitted by various media. On the other hand, the fusion of multiple signs as well as the communicative facilities are far away from simulating what humans achieve in face-to-face conversation. Here opens up a semiotic field of work for artificial intelligence applications. To conclude, let us have a brief outlook on some still underused semiotic means.

7.2. Outlook

According to the technical and the communication share in multimodal technical communication, there are two outlook areas: the one focuses on engineering of *prima facie* marginal channels and media that may nonetheless be used for HCI applications; the other points to possible improvements for the communication facilities of ECAs.

The survey on multimodal technical devices provided in Section 3 to 6, reveals that the lions share on nonverbal engineering deals with producing output on those media and processing input from those channels that are at first sight predominant in (human) communication, that is, hand-and-arm gestures and gaze.

Currently still underdeveloped channels like the chemical (for example: smelling) or the kinaesthetic (for example: balance) one provide further extensions of the bandwidth of HCI. Broadening the range of modalities, channels and media requires interface systems that are able to process information ultimately emitted via all media as substances (cf. Subsection 2.2). Such encompassing systems might be provided by so-called *ambient rooms* (Ishii, Wisneski, Brave, Dahley, Gorbet, Ullmer, and Yari, 1998), that are intelligent surroundings augmenting the user's usual life and work environments (see also Lauth, Berendt, Schmidt, and Pfleging (2012) in this volume).

Systems that aim more towards the communication side of multimodal technical communication will try to become more realistic, that is, more humanlike. The perfect ECA will be an avatar that can easily and in a semiotically rich manner interact with a human interlocutor. There are (at least) three steps that have to be accomplished:

– from input to interaction;
– from manipulations to signs;
– modality-crossing flexibility.

Current input devices lack communicative interaction; the signals they receive are processed as quite hard-wired commands – see for instance the QuickSet system introduced in Section 5.2.3. Signs in a communication situation however, can be used creatively and may be subject to clarification and negotiation. This plasticity of signs is not yet captured by the mere manipulation-based HCI. Somewhat related, natural communication exhibits a remarkable feature, namely that a sign from one modality is transposable or translatable to a sign from another modality. The ability, for instance, to "talk" by hand and feet in a foreign language environment is legion. Information from the spoken language modality is transferred to the gesture or pantomime modality. Research focusing on this aspect of technical communication is already under way, it is the core of barrier-free communication (see Kubina and Lücking (2012) in this volume) where information has to be made freely accessible, that is accessible via various channels.

Though there are yet a couple of multimodal technical applications there is still a long way to go until the full semiotic panoply of communication is exploited. After all, mankind is probably not just the "symbolic species" (Deacon, 1997), it appears to be the "semiotic species" (Sonesson, 2009).

Notes

1. This view is certainly not very original (see as an early example, though of different provenance, the rhetorical advices for effective use of gestures in giving a talk by Quintilian (1. Century)) and probably has popped up in various publications. We owe the semiotic-display-viewpoint to (Goodwin, 2003), for whom the body "[performs] a hierarchy of displays" (p. 222), that is, he conceives "the body as a socially organized field for temporally unfolding displays of meaning tied to relevant action" (p. 238).
2. We use the broad phrase "under normal conditions" to emphasize that we refer to a standard or regular behavior that is not influenced by overriding factors outside or inside the participants. Outside factors could be narrow locations the participant meet in, like elevators. Inside factors are for instance intentional violations of proxemic rules by one interlocutor in order to indicate something to the other interlocutor.
3. The phenomenon of multimodality is not affected by the number of interlocutors. Hence, introducing it and its cognate terms against the background of a dialog setting does no harm to the exposition. As Branigan (2006) argues, dialog does not even differ from multilog *in principle*.
4. Information and meaning are regarded as equivalent in situation semantics, since both are understood as systematic linkages between (types of) situation or events (Turvey and Carello, 1985). Usually, however, information is treated as the broader term covering, inter alia, non-natural meaning (Grice, 1957), whereas meaning in the narrower sense is linguistic meaning, that is, applied only to linguistic types. The tension between a wider (information) and a narrower (linguistic meaning) sense of meaning is amplified in the context of multimodal communication, i.e., communication that does not rest exclusively on linguistic signs. This problem is briefly taken up in Subsection 2.4.
5. Word sense distinction according to the Merriam Webster online dictionary, http://www.merriam-webster.com/dictionary/medium, accessed at August 9, 2011. There is a third sense which is of no concern to the present topic, namely medium as "an individual held to be a channel of communication between the earthly world and a world of spirits".
6. Oviatt (1999, pp. 77 f.) consigns the primacy of speech in multimodal communication to the realms of myths. However, her evidence is concerned with input behavior and

the case-wise importance of non-verbal input means, and does not touch the levels of meaning and communication.

7. Along the Peircean truth-transparency conception of modality, one could introduce a mode as a sign system, since graphical depictions are said to be of higher modality than arbitrary written text. What here is called multicodal would there be called multimodal, then.
8. Nonetheless, a link can be made up between the physiological and the Peircean view on modality along the line of perception reports. Perception reports which arguments are syntactically realized as "naked infinitives" are veridical, that is, the following "principle of veridicality" holds for naked infinitive reports φ: "If *a* sees φ, then φ" (Barwise, 1981, p. 376). Visual perception reports seems to be more trustworthy than non-visual perception, at lest in general. It is not so obvious whether the principle of veridicality applies also to the olfactory modality, for instance in *John smells the lunar space probe landing on the moon*. The veridicality of perception reports seems also to depend on the "epistemic range" of the kind of perception under discussion in relation to the scene (purportedly) perceived, making the veridicality principle a modality-dependent one, sustaining the Peircean view.
9. http://www.langsci.ucl.ac.uk/ipa/, accessed May 5, 2011.
10. The terms refer to Immanuel Kant (1998): *Critique of Pure Reason*. Translated and edited by Paul Guyer & Allen W. Wood. Cambridge University Press, published as part of the series "The Cambridge Edition of the Works of Immanuel Kant".
11. "Phaneroscopy is the description of the *phaneron*; and by the *phaneron* I mean the collective total of all that is in any way or in any sense present to the mind, quite regardless of whether it corresponds to any real thing or not." (CP 1.284)
12. http://www.apple.com/iphone/features/siri.html, accessed November 17, 2011
13. www.cogain.org, accessed November 15, 2011.

References

Alibali, Martha W., Sotaro Kita, and Amanda J. Yound

2000 "Gesture and the Process of Speech Production: We Think, Therefore We Gesture." In: *Language and Cognitive Processes* 15.6, pp. 593–613.

Allwood, Jens

2002 "Bodily Communication Dimensions of Expression and Content." In: *Multimodality in language and speech systems*. Ed. by Björn Granström, David House, and Inger Karlsson. Dordrecht, The Netherlands: Kluwer Academic Publishers, pp. 7–26.

2007 "Multimodal Corpora." In: *Corpus Linguistics. An International Handbook of the Science of Language and Society.* Ed. by Anke Lüdeling and Merja Kytö. Vol. 1. Handbooks of Linguistics and Communication Science. Berlin and New York: De Gruyter. Chap. 12, pp. 207–223.

Anderson, Stephan R. and Edward L. Keenan

1985 "Deixis." In: *Language Typology and Syntactic Description*. Ed. by Timothy Shopen. Vol. III. Cambridge: Cambridge University Press. Chap. 5, pp. 259–308.

Argyle, Michael
1975 *Bodily Communication*. New York, NY: Methuen & Co.
Argyle, Michael and Mark Cook
1976 *Gaze and Mutual Gaze*. Cambridge, UK: Cambridge University Press.
ART
2011 *ART – advanced realtime tracking GmbH, Homepage*. retrieved September 2011.
Austin, John L.
1962 *How To Do Things With Words*. 2nd ed. Cambridge, MA: Harvard University Press.
Baggia, Paolo, Daniel C. Burnett, Jerry Carter, Deborah A. Dahl, Gerry McCobb, and Dave Raggett D.
2009 *W3C: EMMA: Extensible MultiModal Annotation markup language*. http://www.w3.org/TR/emma/.
Barwise, Jon
1981 "Scenes and Other Situations." In: *The Journal of Philosophy* 78.7, pp. 369–397.
Baum, Leonard E. and Ted Petrie
1966 "Statistical inference for probabilistic functions of finite state Markov chains." In: *The Annals of Mathematical Statistics* 37.6, pp. 1554–1563.
Baum, Leonard E., Ted Petrie, George Soules and Norman Weiss
1970 "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains." In: *The annals of mathematical statistics* 41.1, pp. 164–171.
Bavelas, Janet and Jennifer Gerwing
2007 "Conversational Hand Gestures and Facial Displays in Face-to-Face Dialogue." In: *Social Communication*. Ed. by Klaus Fiedler. New York, NY: Psychology Press. Chap. 10, pp. 283–308.
Bavelas, Janet, Christine Kenwood, Trudy Johnson and Bruce Philips
2002 "An Experimental Study of When and How Speakers Use Gestures to Communicate." In: *Gesture* 2.1, pp. 1–17.
Bellotti, Victoria, Maribeth Back, W. Keith Edwards, Rebecca E. Grinter, Austin Henderson and Cristina Lopes
2002 "Making sense of sensing systems: five questions for designers and researchers." In: *Proceedings of the SIGCHI conference on Human factors in computing systems: Changing our world, changing ourselves*. CHI'02. Minneapolis, Minnesota, USA: ACM, pp. 415–422.
Bergmann, Kirsten and Stefan Kopp
2006 "Verbal or visual? How information is distributed across speech and gesture in spatial dialog." In: *Proceedings of the 10th Workshop on the Semantics and Pragmatics of Dialogue*. Ed. by David Schlangen and Raquel Fernández. Brandial'06. Potsdam: Universitätsverlag Potsdam, pp. 90–97.
2009 "GNetIc – Using Bayesian Decision Networks for Iconic Gesture Generation." In: *Proceedings of the 9th Conference on Intelligent Virtual Agents*. LNAI 5773. Berlin; Heidelberg: Springer, pp. 76–89.

Beuter, Niklas, Thorsten P. Spexard, Ingo Lütkebohle, Julia Peltason and Franz Kummert
2008 "Where is this? – Gesture based multi-modal interaction with an anthropomorphic robot." In: *International Conference on Humanoid Robots, Daejeon, Korea. IEEE-RAS.*

Bierman, Arthur K.
1962 "That There Are No Iconic Signs." In: *Philosophy and Phenomenological Research* 23.2, pp. 243–249.

Bird, Steven and Mark Liberman
2001 "A formal framework for linguistic annotation." In: *Speech Communication* 33.1-2. Speech Annotation and Corpus Tools, pp. 23–60.

Birdwhistell, Ray L.
1970 *Kinesics and Context: Essays on Body Motion Communication*. Philadelphia: University of Pennsylvania Press.

Blackman, Samuel and Robert Popoli
1999 *Design and analysis of modern tracking systems*. Vol. 685. Artech House Noorwood, MA.

Bolt, Richard A.
1980 "Put-That-There: Voice and gesture at the graphics interface." In: *ACM SIGGRAPH Computer Graphics*. New York: ACM Press, pp. 262–270.
1981 "Gaze-orchestrated dynamic windows." In: *Proceedings of the 8th Annual Conference on Computer Graphics and Interactive Techniques*. ACM Press New York, NY, USA, pp. 109–119.

Branigan, Holly
2006 "Perspectives on Multiparty Dialogue." In: *Research on Language & Computation* 4 (2), pp. 153–177.

Breazeal, Cynthia and Brian Scassellati
2000 "Infant-like Social Interactions between a Robot and a Human Caregiver." In: *Adaptive Behavior* 8.1, pp. 49–74.

Bromberger, Sylvain
1988 *Types and Tokens in Linguistics*. Tech. rep. CSLI–88–125. Stanford: CSLI.

Brück, Kurt
1997 "Der thermische Kanal." In: *Semiotik: ein Handbuch zu den zeichentheoretischen Grundlagen von Natur und Kultur.* Ed. by Roland Posner, Klaus Robering, and Thomas A. Seboek. Vol. 1. Handbücher zur Sprach- und Kommunikationswissenschaft. Berlin: de Gruyter. Chap. II. Gegenstand I: Aspekte der Semiose – Kanäle, Medien und Kodes / General Topics I: Aspects of Semiosis – Channels, Media, and Codes, pp. 325–330.

Brugman, Hennie and Peter Wittenburg
2001 "The application of annotation models for the construction of databases and tools – Overview and analysis of MPI work since 1994." In: *Proceedings of the IRCS Workshop on Linguistic Databases*. Philadelphia: LDC, pp. 65–75.

Bruneau, Thomas J.
1980 "Chronemics and the Verbal-Nonverbal Interface." In: *The Relationship of Verbal and Nonverbal Communication*. Ed. by Mary Ritchie Key. The Hague, The Netherlands: Mouton, pp. 101–118.

Buisine, Stéphanie and Jean-Claude Martin
2007 "The effects of speech-gesture cooperation in animated agents' behavior in multi-media presentations." In: *Interacting with Computers* 19.4, pp. 484–493.

Burks, Arthur W.
1949 "Icon, Index, and Symbol." In: *Philosophy and Phenomenological Research* 9.4, pp. 673–689.
Butterworth, George
2003 "Pointing Is the Royal Road to Language for Babies." In: *Pointing: Where Language, Culture, and Cognition Meet*. Ed. by Sotaro Kita. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc. Chap. 2, pp. 9–33.
Cassell, Justin, T. Bickmore, M. Billinghurst, L. Campbell, K. Chang, H. Vilhjálmsson and H. Yan
1998 "An architecture for embodied conversational characters." In: *Proceedings of the First Workshop on Embodied Conversational Characters, Tahoe City, California October 1998*, pp. 109–120.
Cassell, Justin, T. Bickmore, M. Billinghurst, L. Campbell, K. Chang, H. Vilhjálmsson and H. Yan
1999 "Embodiment in conversational interfaces: Rea." In: *Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit*. CHI '99. Pittsburgh, Pennsylvania, United States: ACM, pp. 520–527.
Cassell, Justin and Scott Prevost
1996 "Distribution of Semantic Features Across Speech and Gesture by Humans and Computers." In: *Proceedings of the Workshop on Integration of Gesture in Language and Speech*.
Cassell, Justin, Joseph Sullivan, Scott Prevost and Elizabeth Churchill (eds.)
2000 *Embodied Conversational Agents*. Cambridge, MA: MIT Press.
Chandler, Daniel
2002 *Semiotics: The Basics*. New York, NY: Routledge.
Clark, Herbert H. (ed.)
1992 *Arenas of Language Use*. Chicago and Stanford: University of Chicago Press & CSLI.
Clark, John and Colin Yallop
1990 *An Introduction to Phonetics and Phonology*. 2nd ed. Blackwell Textbooks in Linguistics. Oxford, UK and Cambridge, MA: Blackwell.
Cohen, Philip R., Jerry R. Morgan and Martha E. Pollack (eds.)
1990 *Intentions in Communication*. Cambridge, MA: MIT Press.
Condon, William S. and William D. Ogsten
1967 "A Segmentation of Behavior." In: *Journal of Psychiatric Research* 5, pp. 221–235.
Crystal, David
1971 "Paralinguistics." In: *Linguistics at Large*. New York: Viking Press, pp. 162–174.
Deacon, Terrence
1997 *The Symbolic Species*. London: Penguin Press.
Delgado, Ramón López-Cózar and Masahiro Araki
2005 *Spoken, Multilingual, and Multimodal Dialogue Systems: Development and Assessment*. Chichester, GB: John Wiley & Sons Ltd.
Duchowski, Andrew T.
2007 *Eye Tracking Methodology: Theory and Practice*. Second Edition. Springer-Verlag London Limited.

Eco, Umberto
1976 *A Theory of Semiotics*. Bloomington: Indiana University Press.
1985/2000 *Über Spiegel und andere Phänomene*. München: DTV.
Efron, David
1941 *Gesture and Environment*. Republiziert 1972 als *Gesture, Race, and Culture*. The Hague: Mouton. Morningside Heights, NY: King's Crown Press.
Ekman, Paul and Wallace V. Friesen
1969 "The Repertoire of Nonverbal Behavior: Categories, Origins, Usage, and Coding." In: *Semiotica* 1.1, pp. 49–98.
1978 *Facial Action Coding System*. Stanford University, Palo Alto: Consulting Psychologists Press.
Enfield, Nick J.
2009 *The Anatomy of Meaning: Speech, Gesture, and Composite Utterances*. Language, Culture and Cognition 13. Cambridge, UK: Cambridge University Press.
Engel, Andreas K. and Wolf Singer
2001 "Temporal binding and the neural correlates of sensory awareness." In: *Trends in Cognitive Sciences* 5.1, pp. 16–25.
Fröhlich, Martin and Ipke Wachsmuth
1998 "Gesture recognition of the upper limbs – from signal to symbol." In: *Gesture and Sign Language in Human-Computer Interaction*. Ed. by Ipke Wachsmuth and Martin Fröhlich. LNAI 1371. Springer-Verlag, pp. 173–184.
Galley, Niels
1997 "Die Organisation von Augenbewegungen: Fallstudie einer mehrkanaligen Semiose." In: *Semiotik: ein Handbuch zu den zeichentheoretischen Grundlagen von Natur und Kultur*. Ed. by Roland Posner, Klaus Robering, and Thomas A. Seboek. Vol. 1. Handbücher zur Sprach- und Kommunikationswissenschaft. Berlin: de Gruyter. Chap. II. Gegenstand I: Aspekte der Semiose – Kanäle, Medien und Kodes / General Topics I: Aspects of Semiosis – Channels, Media, and Codes, pp. 330–345.
Garfinkel, Harold
1967 *Studies in Ethnomethodology*. Malden, MA: Prentice Hall.
Gerwing, Jennifer and Meredith Allison
2009 "The relationship between verbal and gestural contributions in conversation: A comparison of three methods." In: *Gesture* 9.3, pp. 312–336.
Gerwing, Jennifer and Janet B. Bavelas
2004 "Linguistic influences on gesture's form." In: *Gesture* 4.2, pp. 157–195.
Gibbon, Dafydd, Inge Mertins, and Roger K. Moore, eds.
2000 *Handbook of Multimodal and Spoken Dialogue Systems: Resources, Terminology and Product Evaluation*. The Springer International Series in Engineering and Computer Science. New York, NY: Springer.
Goldin-Meadow, Susan
2001 *Hearing Gesture: How Our Hands Help Us Think*. Cambridge, MA: Belknap Press.
Goodman, Nelson
1976 *Languages of Art. An Approach to a Theory of Symbols*. 2nd ed. Idianapolis: Hackett Publishing Company, Inc.

Goodwin, Charles
1981 *Conversational Organization: Interaction between Speakers and Hearers*. New York: Academic Press.
2003 "Pointing as Situated Practice." In: *Pointing: Where Language, Culture, and Cognition Meet*. Ed. by Sotaro Kita. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc. Chap. 2, pp. 217–241.
Grammer, Karl
2004 "Körpersignale in menschlicher Interaktion." In: *Semiotik: ein Handbuch zu den zeichentheoretischen Grundlagen von Natur und Kultur*. Ed. by Roland Posner, Klaus Robering, and Thomas A. Seboek. Vol. 4. Handbücher zur Sprach- und Kommunikationswissenschaft. Berlin: de Gruyter. Chap. XV. Ausgewählte Gegenstände der Semiotik / Selected Topics of Semiotics, pp. 3448–3487.
Grice, Herbert Paul
1957 "Meaning." In: *Philosophical Review* 66.3, pp. 377–388.
1975 "Logic and Conversation." In: *Syntax and Semantics*. Ed. by Peter Cole and Jerry L. Morgan. Vol. 3, Speech Acts. New York: Academic Press, pp. 41–58.
Guyon, Isabelle and André Elisseeff
2003 "An introduction to variable and feature selection." In: *The Journal of Machine Learning Research* 3, pp. 1157–1182.
Habets, Boukje, Sotaro Kita, Zeshu Shao, Asli Özyurek and Peter Hagoort
2011 "The Role of Synchrony and Ambiguity in Speech-Gesture Integration during Comprehension." In: *Journal of Cognitive Neuroscience* 23.8, pp. 1845–1854.
Hager, Joseph C. and Paul Ekman
1983 "The Inner and Outer Meanings of Facial Expressions." In: *Social Psychophysiology: A Sourcebook*. Ed. by J. T. Cacioppo and R. E. Petty. New York, N. Y.: The Guilford Press. Chap. 10, pp. 287–306.
Hall, Edward T.
1966 *The hidden dimension*. New York: Doubleday. –
1968 "Proxemics." In: *Current Anthropology* 9.2-3, pp. 83–108.
Handy, Todd C. (ed.)
2005 *Event-Related Potentials: A Methods Handbook*. Cambdridge, MA: MIT Press.
Hess-Lüttich, Ernst W. B. and Dagmar Schmauks
2004 "Multimediale Kommunikation." In: *Semiotik: ein Handbuch zu den zeichen-theoretischen Grundlagen von Natur und Kultur*. Ed. by Roland Posner, Klaus Robering, and Thomas A. Seboek. Vol. 4. Handbücher zur Sprach- und Kommunikationswissenschaft. Berlin: de Gruyter. Chap. XV. Ausgewählte Gegenstände der Semiotik / Selected Topics of Semiotics, pp. 3487–3503.
Heuer, Herbert
1997 "Der taktile Kanal." In: *Semiotik: ein Handbuch zu den zeichentheoretischen Grundlagen von Natur und Kultur*. Ed. by Roland Posner, Klaus Robering, and Thomas A. Seboek. Vol. 1. Handbücher zur Sprach- und Kommunikationswissenschaft. Berlin: de Gruyter. Chap. II. Gegenstand I: Aspekte der Semiose – Kanäle, Medien und Kodes / General Topics I: Aspects of Semiosis – Channels, Media, and Codes, pp. 300–306.

Hostetter, Autumn B. and Martha W. Alibali
2011 "Cognitive skills and gesture–speech redundancy: Formulation diffculty or communicative strategy?" In: *Gesture* 11.1, pp. 40–60.
Hutchinson, Ann
1991 *Labanotation: The System of Analyzing and Recording Movement*. 3rd ed. Theatre Arts Books. New York: Routledge.
Hutchinson, T. E., K. P. White, W. N. Martin, K. C. Reichert and L. A. Frey
1989 "Human-computer interaction using eye-gaze input." In: *IEEE Transactions on systems, man and cybernetics* 19.6, pp. 1527–1534.
Ishiguro, Hiroshi
2007 "Android Science Toward a New Cross-Interdisciplinary Framework." In: *Robotics research: results of the 12th International Symposium ISRR*. Vol. 28. Springer Verlag, p. 118.
Ishii, Hiroshi, Craig Wisneski, Scott Brave, Andrew Dahley, Matt Gorbet, Brygg Ullmer and Paul Yari
1998 "ambientROOM: Integrating Ambient Media with Architectural Space." In: *Proceedings of the Conference on Human Factors in Computing systems*. CHI 98. ACM. Los Angeles, CA, pp. 18–23.
Israel, David and John Perry
1990 "What is Information?" In: *Information, Language and Cognition*. Ed. by Philip Hanson. Vancouver: University of British Columbia Press, pp. 1–19.
Jacob, Robert J. K.
1993 "Eye-movement-based human-computer interaction techniques: Toward non-command interfaces." In: ed. by H. R. Hartson and D. Hix. Vol. 4. Advances in Human-Computer Interaction. Norwood, New Jersey: Ablex Publishing Corporation. Chap. 6, pp. 151–190.
Jacob, Robert J. K. and Keith S. Karn
2003 "Eye Tracking in Human-Computer Interaction and Usability Research: Ready to Deliver the Promises (Section Commentary)." In: *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*. Ed. by J. Hyona, R. Radach, and H. Deubel. Amsterdam: Elsevier Science, pp. 573–605.
Jewitt, Carey (ed.)
2009 *The Routledge Handbook of Multimodal Analysis*. London and New York: Routledge.
Johnson, W. Lewis and Jeff Rickel
1997 "Steve: an animated pedagogical agent for procedural training in virtual environments." In: *SIGART Bull.* 8 (1–4), pp. 16–21.
Johnston, Michael
1998 "Unification-based Multimodal Parsing." In: *Proceedings of the 36th Annual Meeting on Association for Computational Linguistics – Volume I*. Annual Meeting of the ACL. Montreal, Quebec, Canada: Association for Computational Linguistics, pp. 624–630.
Johnston, Michael and Srinivas Bangalore
2000 "Finite-state Multimodal Parsing and Understanding." In: *Proceedings of the 18th Conference on Computational Linguistics – Volume I*. International Conference on Computational Linguistics. Saarbrücken, Germany: Association for Computational Linguistics, pp. 369–375.

Johnston, Michael, Philip R. Cohen, David McGee, Sharon L. Oviatt, James A. Pittman and Ira Smith

1997 "Unification-based Multimodal Integration." In: *Proceedings of the Eighth Conference on European Chapter of the Association for Computational Linguistics*. European Chapter Meeting of the ACL. Madrid, Spain: Association for Computational Linguistics, pp. 281–288.

Jovanovic, Natasa, Rieks op den Akker and Anton Nijholt

2006 "A corpus for studying addressing behaviour in multi-party dialogues." In: *Language Resources and Evaluation* 40 (1), pp. 5–23.

Kauffman, Lynn E.

1971 "Tacesics, the Study of Touch: A Model for Proxemic Analysis." In: *Semiotica* 4.2, pp. 149–161.

Kauther, Ralf and Michael Müller

2000 "Rudi Keller: Zeichentheorie. Zu einer Theorie semiotischen Wissens." In: *Journal for General Philosophy of Science* 31. Book Review, pp. 347–356.

Keller, Rudi

1995 *Zeichentheorie. Zu einer Theorie semiotischen Wissens*. Tübingen: Francke.

Kelly, Spencer D., Sarah M. Manning and Sabrina Rodak

2008 "Gesture Gives a Hand to Language and Learning: Perspectives from Cognitive Neuroscience, Developmental Psychology and Education." In: *Language and Linguistics Compass* 2.

Kendon, Adam

1967 "Some functions of gaze-direction in social interaction." In: *Acta Psychologica* 26, pp. 22–63.

1972 "Some relationships between body motion and speech. An analysis of an example." In: *Studies in Dyadic Communication*. Ed. by Aron Wolfe Siegman and Benjamin Pope. Elmsford, NY: Pergamon Press. Chap. 9, pp. 177–210.

1980 "Gesticulation and Speech: Two Aspects of the Process of Utterance." In: *The Relationship of Verbal and Nonverbal Communication*. Ed. by Mary Ritchie Key. Vol. 25. Contributions to the Sociology of Language. The Hague: Mouton Publishers, pp. 207–227.

1994 "Do Gestures Communicate? A Review." In: *Research on Language and Social Interaction* 27.3, pp. 175–200.

2000 "Language and Gesture: Unity or Duality?" In: *Language and Gesture*. Ed. by David McNeill. Cambridge, UK: Cambridge University Press. Chap. 2, pp. 47–63.

2004 *Gesture: Visible Action as Utterance*. Cambridge, MA: Cambridge University Press.

Kipp, Michael, Michael Neff and Irene Albrecht

2007 "An Annotation Scheme for Conversational Gestures: How to economically capture timing and form." In: *Journal on Language Resources and Evaluation – Special Issue on Multimodal Corpora* 41.3–4, pp. 325–339.

Kita, Sotaro and Asli Özyürek

2003 "What Does Cross-Linguistic Variation in Semantic Coordination of Speech and Gesture Reveal?: Evidence for an Interface Representation of Spatial Thinking and Speaking." In: *Journal of Memory and Language* 48.1, pp. 16–32.

Kita, Sotaro, Asli Özyürek, Shanley Allen, Amanda Brown, Reyhan Furman and Tomoko Ishizuka
2007 "Relations between syntactic encoding and co-speech gestures: Implications for a model of speech and gesture production." In: *Language and Cognitive Processes* 22.8, pp. 1212–1236.

Kittler, Friedrich
2004 "Geschichte der Kommunikationstechniken." In: *Semiotik: ein Handbuch zu den zeichentheoretischen Grundlagen von Natur und Kultur*. Ed. by Roland Posner, Klaus Robering, and Thomas A. Seboek. Vol. 4. Handbücher zur Sprach- und Kommunikationswissenschaft. Berlin: de Gruyter. Chap. XV. Ausgewählte Gegenstände der Semiotik / Selected Topics of Semiotics, pp. 3345–3357.

Koons, David B., Carlton J. Sparrell and Kristinn R. Thorisson
1993 "Integrating simultaneous input from speech, gaze, and hand gestures." In: Menlo Park, CA, USA: American Association for Artificial Intelligence, pp. 257–276.

Kopp, Stefan, Kirsten Bergmann and Ipke Wachsmuth
2008 "Multimodal Communication from Multimodal Thinking–Towards an Integrated Model of Speech and Gesture Production." In: *International Journal of Semantic Computing* 2.1, pp. 115–136.

Kopp, Stefan, Bernhard Jung, Nadine Lessmann and Ipke Wachsmuth
2003 "Max – A Multimodal Assistant in Virtual Reality Construction." In: *Künstliche Intelligenz* 4, pp. 11–17.

Kopp, Stefan, Timo Sowa and Ipke Wachsmuth
2004 "Imitation Games with an Artificial Agent: From Mimicking to Understanding Shape-Related Iconic Gestures." In: *Gesture-Based Communication in Human-Computer Interaction, International Gesture Workshop, Genua, Italy April 2003*. Ed. by A. Camurri and G. Volpe. LNAI 2915. Springer, pp. 436–447.

Kopp, Stefan, Paul Tepper and Justine Cassell
2004 "Towards Integrated Microplanning of Language and Iconic Gesture for Multimodal Output." In: *Proceedings of the International Conference on Multi-modal Interfaces (ICMI'04)*. ACM Press, pp. 97–104.

Kopp, Stefan and Ipke Wachsmuth
2012 "Artificial interactivity." In: *Handbook of Technical Communication*. Ed. by Alexander Mehler, Laurent Romary, and Dafydd Gibbon. Berlin/Boston: De Gruyter.

Kranstedt, Alfred, Stefan Kopp and Ipke Wachsmuth
2002 "MURML: A Multimodal Utterance Representation Markup Language for Conversational Agents." In: *Proceedings of the Workshop Embodied conversational agents – let's specify and evaluate them!, held at the First Int. Joint Conference on Autonomous Agents & Multi-Agent Systems*. Bologna, Italy.

Krauss, Robert M., Yihsiu Chen and Rebecca F. Gottesmann
2000 "Lexical Gestures and Lexical Access: A Process Model." In: *Language and Gesture*. Ed. by David McNeill. Cambridge, UK: Cambridge University Press. Chap. 13, pp. 261–283.

Kröller, Jürgen
1997 "Chemical Channels." In: *Semiotik: ein Handbuch zu den zeichentheoretischen Grundlagen von Natur und Kultur*. Ed. by Roland Posner, Klaus Robering, and Thomas A. Seboek. Vol. 1. Handbücher zur Sprach- und Kommunikationswissenschaft. Berlin: de Gruyter. Chap. II. Gegenstand I: Aspekte der Semiose – Kanäle, Medien und Kodes / General Topics I: Aspects of Semiosis – Channels, Media, and Codes, pp. 306–316.

Kubina, Petra and Andy Lücking
2012 "Barrier-free communication." In: *Handbook of Technical Communication*. Ed. by Alexander Mehler, Laurent Romary, and Dafydd Gibbon. Berlin/Boston: De Gruyter.

Kurtenbach, Gord and Eric A. Hulteen
1990 *Gestures in human-computer communication*. In: The art of human-computer interface design, pp. 309–317.

Landwehr, Klaus
1997 "Der optische Kanal." In: *Semiotik: ein Handbuch zu den zeichentheoretischen Grundlagen von Natur und Kultur*. Ed. by Roland Posner, Klaus Robering, and Thomas A. Seboek. Vol. 1. Handbücher zur Sprach- und Kommunikationswissenschaft. Berlin: de Gruyter. Chap. II. Gegenstand I: Aspekte der Semiose – Kanäle, Medien und Kodes / General Topics I: Aspects of Semiosis – Channels, Media, and Codes, pp. 288–294.

Latoschik, Marc E.
2001a "A General Framework for Multimodal Interaction in Virtual Reality Systems: PrOSA." In: *The Future of VR and AR Interfaces – Multimodal, Humanoid, Adaptive and Intelligent. Proceedings of the workshop at IEEE Virtual Reality 2001, Yokohama, Japan*. Ed. by W. Broll and L. Schäfer. Sankt Augustin: GMD Report No. 138, GMD-Forschungszentrum Informationstechnik GmbH, pp. 21–25.
2001b "Multimodale Interaktion in Virtueller Realität am Beispiel der virtuellen Konstruktion." PhD thesis. University of Bielefeld.
2002 "Designing Transition Networks for Multimodal VR-Interactions Using a Markup Language." In: *Proceedings of the IEEE fourth International Conference on Multimodal Interfaces*. ICMI 2002. Pittsburgh, USA, pp. 411–416.

Latoschik, Marc E., Martin Fröhlich, Bernhard Jung and Ipke Wachsmuth
1998 "Utilize Speech and Gestures to Realize Natural Interaction in a Virtual Environment." In: *Proceedings of the 24th Annual Conference of the IEEE Industrial Electronics Society*. Vol. 4. IECON'98. IEEE, pp. 2028–2033.

Lauth, Codrina, Bettina Berendt, Albrecht Schmidt and Bastian Pfleging
2012 "Ubiquitous computing." In: *Handbook of Technical Communication*. Ed. by Alexander Mehler, Laurent Romary, and Dafydd Gibbon. Berlin/Boston: De Gruyter.

Lazar, Jonathan, Aaron Allen, Jason Kleinman and Chris Malarkey
2007 "What Frustrates Screen Reader Users on the Web: A Study of 100 Blind Users." In: *International Journal of Humancomputer Interaction* 22.3, pp. 247–269.

Lee, Jina, Stacy Marsella, David R. Traum, Jonathan Gratch and Brent Lance
2007 "The Rickel Gaze Model: A Window on the Mind of a Virtual Human." In: *Proceedings of the 7th international conference on Intelligent Virtual Agents*. Ed. by Catherine Pelachaud, Jean-Claude Martin, Elisabeth André, Gérard Chollet, Kostas Karpouzis, and Danielle Pelé. IVA'07. Berlin, Heidelberg: Springer-Verlag, pp. 296–303.

Lehmberg, Tim and Kai Wörner
2007 "Annotation Standards." In: *Corpus Linguistics. An International Handbook of the Science of Language and Society*. Ed. by Anke Lüdeling and Merja Kytö. Vol. 1. Handbooks of Linguistics and Communication Science. Berlin and New York: De Gruyter. Chap. 22, pp. 484–501.

Lessmann, Nadine, Alfred Kranstedt and Ipke Wachsmuth
2004 "Towards a Cognitively Motivated Processing of Turn-Taking Signals for the Embodied Conversational Agent Max." In: *AAMAS 2004 Workshop Proceedings: "Embodied Conversational Agents: Balanced Perception and Action"*. New York, pp. 57–64.

Levelt, Willem J. M.
1989 *Speaking: From Intention to Articulation*. Cambridge, MA: MIT Press.

Levine, James L.
1981 *An eye-controlled computer*. Tech. rep. Research Report RC-8857. IBM Thomas J. Watson Research Center, Yorktown Heights, NY.
1984 "Performance of an eyetracker for office use." In: *Computers in Biology and Medicine* 14.1, pp. 77–79, 81–89.

Levinson, Stephen C.
2006 "On the Human 'Interaction Engine'." In: *Roots of Human Sociality: Culture, Cognition and Interaction*. Ed. by N. J. Enfield and Stephen C. Levinson. Wenner-Gren International Symposium Series. Oxford: Berg. Chap. 1, pp. 39–69.

Liew, Allan Wee-Chung and Shilin Wang
2008 *Visual Speech Recognition: Lip Segmentation and Mapping*. Information Science Reference-Imprint of: IGI Publishing.

Loehr, Daniel
2007 "Apects of Rhythm in Gesture in Speech." In: *Gesture* 7.2, pp. 179–214.

Luccio, Riccardo
1997 "Geschichte der Kommunikationstechniken." In: *Semiotik: ein Handbuch zu den zeichentheoretischen Grundlagen von Natur und Kultur*. Ed. by Roland Posner, Klaus Robering, and Thomas A. Seboek. Vol. 1. Handbücher zur Sprach- und Kommunikationswissenschaft. Berlin: de Gruyter. Chap. II. Gegenstand I: Aspekte der Semiose – Kanäle, Medien und Kodes / General Topics I: Aspects of Semiosis – Channels, Media, and Codes, pp. 345–357.

Lücking, Andy
2011 "Prolegoma zu einer Theorie ikonischer Gesten." PhD thesis. Universität Bielefeld.

Lücking, Andy, Kirsten Bergmann, Florian Hahn, Stefan Kopp and Hannes Rieser
2010 "The Bielefeld Speech and Gesture Alignment Corpus (SaGA)." In: *Multimodal Corpora: Advances in Capturing, Coding and Analyzing Multimodality*. 7th International Conference for Language Resources and Evaluation (LREC 2010). Malta, pp. 92–98.

Lücking, Andy, Alexander Mehler and Peter Menke
2008 "Taking Fingerprints of Speech-and-Gesture Ensembles: Approching Empirical Evidence of Intrapersonal Alignmnent in Multimodal Communication." In: *LonDial 2008: The 12th Workshop on the Semantics and Pragmatics of Dialogue (SEMDIAL)*. King's College London, pp. 157–164.
Martell, Craig, Chris Osborn, Jesse Friedman and Paul Howard
2002 "FORM: A Kinematic Annotation Scheme and Tool for Gesture Annotation." In: *Proceedings of Multimodal Resources and Multimodal Systems Evaluation*. MITRE. Gran Canaria, pp. 15–22.
Martin, Jean-Claude
2005 "Analysis and synthesis of cooperation between modalities." In: *Proceedings of the 11th International Conference on Human-Computer Interaction*. HCI International 2005. Las Vegas, Nevada, USA.
Martin, Jean-Claude and Dominique Béroule
1995 "Temporal codes within a typology of cooperation between modalities." In: *Artificial Intelligence Review* 9 (2), pp. 95–102.
McClave, Evelyn
1994 "Gestural Beats: The Rhythm Hypothesis." In: *Journal of Psycholinguistic Research* 23.1, pp. 45–66.
McNeill, David
1992a *Hand and Mind – What Gestures Reveal about Thought*. Chicago: Chicago University Press.
1992b *Hand and Mind: What Gestures Reveal about Thought*. Chicago: University of Chicago Press.
2000 "Catchments and Contexts: Non-Modular Factors in Speech and Gesture Production." In: *Language and Gesture*. Ed. by David McNeill. Cambridge, UK: Cambridge University Press. Chap. 15, pp. 312–328.
Meier, Uwe, Rainer Stiefelhagen, Jie Yang and Alex Waibel
2000 "Towards unrestricted lip reading." In: *International Journal of Pattern Recognition and Artificial Intelligence* 14.5, pp. 571–586.
MESA Imaging AG
2006 *Homepage*. http://www.mesa-imaging.ch/ retrieved September 2011.
Microsoft Kinect SDK
2010 *Homepage of the Microsoft Kinect SDK*. online. Retrieved September 2011 from http://www.xbox.com/de-de/kinect/.
Moller, Peter
1997 "The Electric and Magnetic Channels." In: *Semiotik: ein Handbuch zu den zeichentheoretischen Grundlagen von Natur und Kultur*. Ed. by Roland Posner, Klaus Robering, and Thomas A. Seboek. Vol. 1. Handbücher zur Sprach- und Kommunikationswissenschaft. Berlin: de Gruyter. Chap. II. Gegenstand I: Aspekte der Semiose – Kanäle, Medien und Kodes / General Topics I: Aspects of Semiosis – Channels, Media, and Codes, pp. 316–325.
Moore, Chris and Philip J. Dunham (eds.)
1995 *Joint Attention: Its Origins and Role in Development*. Hillsdale, New Jersey: Lawrence Earlbaum Associates, Inc.
Müller, Cornelia
1998 *Redebegleitende Gesten. Kulturgeschichte – Theorie – Sprachvergleich*. Vol. 1. Körper – Kultur – Kommunikation. Berlin: Berlin Verlag.

Murphy, Keith M.
2003 "Building Meaning in Interaction: Rethinking Gesture Classifications." In: *Crossroads of Language, Interaction, and Culture* 5, pp. 29–47.
Nakano, Yukiko I., Gabe Reinstein, Tom Stocky and Justine Cassell
2003 "Towards a model of face-to-face grounding." In: *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, pp. 553–561.
Neagle, R. J., K. Ng and R. A. Ruddle
2004 "Developing a Virtual Ballet Dancer to Visualize Choreography." In: *Proceedings of the AISB 2004 on Language, Speech & Gesture for Expressive Characters*. AISB'04, pp. 86–97.
Noma, Tukasa and Norman I Badler
1997 "A virtual human presenter." In: *Proceedings of the IJCAI Workshop on Animated Interface Agents: Making Them Intelligent*. Vol. 45. Citeseer, p. 51.
Nöth, Winfried
2000 *Handbuch der Semiotik*. 2nd ed. Stuttgart: Metzler.
Oh, Alice, Harold Fox, Max Van Kleek, Aaron Adler, Krzysztof Gajos, Louis-Philippe Morency and Trevor Darrell
2002 "Evaluating look-to-talk: a gaze-aware interface in a collaborative environment." In: *CHI '02 extended abstracts on Human factors in computing systems*. Minneapolis, Minnesota, USA: ACM Press, pp. 650–651.
Oviatt, Sharon
1999 "Ten Myths of Multimodal Interaction." In: *Communications of the ACM* 42.11, pp. 74–81.
Oviatt, Sharon, Antonella DeAngeli and Karen Kuhn
1997 "Integration and synchronization of input modes during multimodal human-computer interaction." In: *Referring Phenomena in a Multi-media Context and their Computational Treatment*. Association for Computational Linguistics, pp. 1–13.
Paaß, Gerhard
2012 "Document classification, information retrieval, text and web mining." In: *Handbook of Technical Communication*. Ed. by Alexander Mehler, Laurent Romary, and Dafydd Gibbon. Berlin/Boston: De Gruyter.
Peacocke, Christopher
1987 "Depiction." In: *The Philosophical Review* 96.3, pp. 383–410.
Pearl, Judea
1985 "Bayesian Networks: A Model of Self-Activated: Memory for Evidential Reasoning." In: *Proceedings of the 7th Conference of the Cognitive Science Society*. Computer Science Department, University of California, pp. 329–334.
Peirce, Charles Sanders
1867 "On a New List of Categories." In: *Proceedings of the American Academy of Arts and Sciences Series*. Vol. 7, pp. 287–298.
1965 *Collected Papers of Charles Sanders Peirce*. Ed. by Charles Harthshorne and Paul Weiss. repr. from 1932. Cambridge, MA.
Penfield, Wilder and Herbert Jasper
1954 *Epilepsy and the Functional Anatomy of the Human Brain*. 2nd ed. Oxford, UK: Little, Brown & Co.

Pfeiffer, Thies and Marc E. Latoschik
2004 "Resolving Object References in Multimodal Dialogues for Immersive Virtual Environments." In: *Proceedings of the IEEE Virtual Reality 2004*. Ed. by Yasushi Ikei, Martin Göbel, and Jim Chen, pp. 35–42.

Pfeiffer-Lessmann, Nadine and Ipke Wachsmuth
2008 "Toward alignment with a virtual human – achieving joint attention." In: *KI 2008: Advances in Artificial Intelligence*. Ed. by A. R. Dengel, K. Berns, and T.M. Breuel. LNAI 5243. Berlin: Springer, pp. 292–299.

Poslad, Stefan
2009 *Ubiquitous Computing Smart Devices, Environments and Interactions*. Sussex, UK: John Wiley & Sons Ltd.

Posner, Roland
1997 "Pragmatics." In: *Semiotik: ein Handbuch zu den zeichentheoretischen Grundlagen von Natur und Kultur*. Ed. by Roland Posner, Klaus Robering, and Thomas A. Seboek. Vol. 1. Handbücher zur Sprach- und Kommunikationswissenschaft. Berlin: de Gruyter. Chap. I. Systematik / Systematics, pp. 219–246.

Posner, Roland, Klaus Robering and Thomas A. Seboek (eds.)
1997 *Semiotik: ein Handbuch zu den zeichentheoretischen Grundlagen von Natur und Kultur*. Vol. 1. Handbücher zur Sprach- und Kommunikationswissenschaft. Berlin: de Gruyter.

Poyatos, Fernando
1975 "Cross-Cultural Study of Paralinguistic "Alternants" in Face-to-Face Interaction." In: *Organization of Behavior in Face-to-Face Interaction*. Ed. by Adam Kendon, Harris M. Richards, and Mary Ritchie Key. The Hague, The Netherlands: Mouton, pp. 285–314.

Prasov, Zahar and Joyce Y. Chai
2008 "What's in a Gaze?: The Role of Eye-Gaze in Reference Resolution in Multimodal Conversational Interfaces." In: *Proceedings of the 13th international conference on Intelligent user interfaces*. IUI '08. Gran Canaria, Spain: ACM, pp. 20–29.

Prillwitz, S., R. Leven, H. Zienert, T. Hanke and J. Henning
1989 *HamNoSys Version 2.0: Hamburger Notationssystem für Gebärdensprachen. Eine Einführung*. SIGNUM.

Quek, Francis K. H., David McNeill, Robert K. Bryll, Susan Duncan, Xin-Feng Ma, Cemil Kirbas, Karl E. McCullough and Rashid Ansari
2002 "Multimodal human discourse: gesture and speech." In: *ACM Trans. Comput.-Hum. Interact.* 9.3, pp. 171–193.

Quintilian, Marcus Fabius
1. Century *Institutio Oratoria*. http://thelatinlibrary.com/quintilian/quintilian.institutio.shtml.

Rabiner, Lawrence R.
1989 "A tutorial on hidden Markov models and selected applications in speech recognition." In: *Proceedings of the IEEE*, pp. 257–286.

Recanati, François
2004 *Literal Meaning*. Cambridge, MA: Cambridge University Press.

Rieser, Hannes
2008 "Aligned Iconic Gesture in Different Strata of MM Route-Description." In: *LonDial 2008: The 12th Workshop on the Semantics and Pragmatics of Dialogue (SEMDIAL)*. King's College London, pp. 167–174.
Ruiter, Jan Peter de
2000 "The Production of Gesture and Speech." In: *Language and Gesture*. Ed. by David McNeill. Cambridge, UK: Cambridge University Press. Chap. 14, pp. 284–311.
2004 "On the Primacy of Language in Multimodal Communication." In: *Proceedings of workshop on multimodal corpora*. Lisbon.
San Agustin, Javier, Henrik Skovsgaard, John Paulin Hansen and Dan Witzner Hansen
2009 "Low-cost gaze interaction: ready to deliver the promises." In: *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems*. CHI '09. Boston, MA, USA: ACM, pp. 4453–4458.
Schmidt, Robert F. and Gerhard Thews
1997 *Physiologie des Menschen*. 27th ed. Berlin, Heidelberg, and New York: Springer.
Schneiderman, B. and C. Plaisant
2009 *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. 5th. Addison Wesley.
Scully, Jack and Ernie Blood
1986 *Ascension Technology Corporation*. retrieved April 2010.
Searle, John R.
1978 "Literal meaning." In: *Erkenntnis* 13 (1), pp. 207–224.
Shannon, Claude Elwood
1948 "A Mathematical Theory of Communication." In: *Bell System Technical Journal* 27, pp. 379–423, 623–656.
Shannon, Claude Elwood and Warren Weaver
1949 *The Mathematical Theory of Communication*. Illinois, US: University of Illinois Press.
Sonesson, Göran
1997 "The Ecological Foundations of Iconicity." In: *Semiotics Around the World: Synthesis in Diversity. Proceedings of the Fifth International Congress of the IASS*. Ed. by Irmengard Rauch and Gerlard F. Carr. Berkeley, 1994. Berlin and New York: de Gruyter, pp. 739–742.
1998 "That There Are Many Kinds of Iconic Signs." In: *Visio* 1.1, pp. 33–54.
2009 "Here comes the semiotic species: reflections on the semiotic turn in the cognitive sciences." In: *Symbolic transformations. The mind in movement through culture and society.* Ed. by Brady Wagoner. Cultural Dynamics of Social Representation. London: Routledge, pp. 38–58.
Sowa, Timo
2005 "Understanding Coverbal Iconic Gestures in Shape Descriptions." PhD thesis. Universität Bielefeld, Technische Fakultät, AG Wissensbasierte Systeme.
2006 *Understanding coverbal iconic gestures in shape descriptions*. Akademische Verlagsgesellschaft Aka: Berlin.
Sparrell, Carlton Jr. and David B. Koons
1994 "Interpretation of coverbal depictive gestures." In: *AAAI Spring Symposium Series*. Stanford University, pp. 8–12.

Sperber, Dan and Deirdre Wilson
1986 *Relevance: Communication and Cognition*. Oxford: Blackwell.
Starker, India and Richard A. Bolt
1990 "A gaze-responsive self-disclosing display." In: *CHI '90: Proceedings of the SIGCHI conference on Human factors in computing systems*. Seattle, Washington, United States: ACM, pp. 3–10.
Stokoe, William C., Dorothy C. Casterline and Carl G. Croneberg
1965 *A Dictionary of American Sign Language on Linguistic Principles*. Washington, D. C.: Gallaudet College Press.
Streeck, Jürgen
2009 *Gesture Craft: The Manu-Facture of Meaning*. Vol. 2. Gesture Studies. John Benjamins.
Strube, Gerhard and Gerda Lazarus
1997 "Der akutische Kanal." In: *Semiotik: ein Handbuch zu den zeichentheoretischen Grundlagen von Natur und Kultur*. Ed. by Roland Posner, Klaus Robering, and Thomas A. Seboek. Vol. 1. Handbücher zur Sprach- und Kommunikationswissenschaft. Berlin: de Gruyter. Chap. II. Gegenstand I: Aspekte der Semiose – Kanäle, Medien und Kodes / General Topics I: Aspects of Semiosis – Channels, Media, and Codes, pp. 294–300.
Stührenberg, Maik
2012 "Foundations of markup languages." In: *Handbook of Technical Communication*. Ed. by Alexander Mehler, Laurent Romary, and Dafydd Gibbon. Berlin/Boston: De Gruyter.
Sun, Yong, Fang Chen and Vera Chung
2006 "QuickFusion: multi-modal fusion without time thresholds." In: *Proceedings of the 2005 NICTA-HCSNet Multimodal User Interaction Workshop*. Australian Computer Society, Inc., pp. 51–54.
Ten Kate, J. H., Edward E. E. Frietman, W. Willems, B. M. Ter Haar Romeny and E. Tenkink
1979 "Eye-Switch Controlled Communication Aids." In: *Proceedings of the 12th International Conference on Medical & Biological Engineering*, pp. 19–20.
Thórisson, Kristinn R.
1996 "Communicative humanoids: a computational model of psychosocial dialogue skills." PhD thesis. Massachusetts Institute of Technology.
1997 "Gandalf: an embodied humanoid capable of real-time multi-modal dialogue with people." In: *Proceedings of the first international conference on Autonomous agents*. AGENTS '97. Marina del Rey, California, United States: ACM, pp. 536–537.
Thurlow, Crispin, Laura Lengel and Alice Tomic
2004 *Computer Mediated Communication. Social Interaction and the Internet*. London, UK: SAGE Publications Ltd.
Tomasello, Michael
2008 *Origins of Human Communication*. Cambridge, MA: MIT Press.
Torres, Obed, Justine Cassell and Scott Prevost
1997 "Modeling Gaze Behavior as a Function of Discourse Structure." In: *Paper presented at the First International Workshop on Human-Computer Conversation*.

Traum, David and Jeff Rickel
2002 "Embodied agents for multi-party dialogue in immersive virtual worlds." In: *Proceedings First Int. Joint Conference on Autonomous Agents and Multiagent systems*, pp. 766–773.

Tuite, Kevin
1993 "The Production of Gesture." In: *Semiotica* 93.1/2, pp. 83–105.

Turvey, M. T. and Claudia Carello
1985 "The Equation of Information and Meaning from the Perspectives of Situation Semantics and Gibson's Ecological Realism." In: *Linguistics and Philosophy* 8, pp. 81–90.

Vermaak, J., M. Gangnet, A. Blake and P. Perez
2001 "Sequential Monte Carlo fusion of sound and vision for speaker tracking." In: *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*. Vol. 1. IEEE, pp. 741–746.

Vicon Motion Systems
1984 *Homepage*. http://www.vicon.com, retrieved April 2010.

Virtual Technologies (Virtex)
1991 *CyberGlove*.
http://www.virtex.com/~virtex/products/hw_products/cyberglove.html.

Waibel, Alex and Kai-Fu Lee
1990 *Readings in speech recognition*. Morgan Kaufmann Publishers Inc.

Waisman, Friedrich
1951 "Verifiability." In: *Logic and Language*. Ed. by Anthony Flew. Gregg Revivals Reprint 1993. Vol. 1. 2 vols. Hampshire, UK: Blackwell, pp. 117–144.

Walther, Joseph B.
1996 "Computer-Mediated Communication. Impersonal, Interpersonal, and Hyperpersonal Interaction." In: *Communication Research* 23.1, pp. 3–43.

Weidenmann, Bernd
1997 "Multicodierung und Multimodalität im Lernprozess." In: *Information und Lernen mit Multimedia. Lehrbuch für Studium und Praxis*. Ed. by Ludwig J. Issing and Paul Klimsa. Weinheim: Psychologie Verlags Union. Chap. 4, pp. 65–84.

Wharton, Tim
2009 *Pragmatics and Non-Verbal Communication*. Cambridge, NY: Cambridge University Press.

Wiltshire, Anne
2007 "Synchrony as the underlying structure of gesture: the relationship between speech sound and body movement at the micro level." In: *In Between Wor(l)ds: Transformation and Translation*. Ed. by Robyn Loughnane, Cara Penry Williams, and Jana Verhoeven. Vol. 6. Postgraduate Research Papers on Language and Literature. Melbourne: School of Languages and Linguistics, The University of Melbourne, pp. 235–253.

Winter, Andreas, Bernt Kullbach and Volker Riediger
2002 "An Overview of the GXL Graph Exchange Language." In: *Software Visualization*. Ed. by Stephan Diehl. Vol. 2269. Lecture Notes in Computer Science. Springer Berlin and Heidelberg, pp. 528–532.

Wittenburg, Peter
2007 "Preprocessing Multimodal Corpora." In: *Corpus Linguistics. An International Handbook of the Science of Language and Society*. Ed. by Anke Lüdeling and Merja Kytö. Vol. 1. Handbooks of Linguistics and Communication Science. Berlin and New York: De Gruyter. Chap. 31, pp. 664–685.

Wundt, Wilhelm
1911 *Völkerpsychologie. Eine Untersuchung der Entwicklungsgesetze von Sprache, Mythus und Sitte*. Vol. Erster Band: Die Sprache. Erster Teil. Leipzig: Wilhelm Engelmann.

Zhang, Nevin Lianwen
1993 "A Computational Theory of Decision Networks." PhD thesis. The University of British Columbia.

19. Barrier-free communication

Petra Kubina and Andy Lücking

1. Introduction

1.1. The notion “barrier-free” in technical communication

Under unconstrained conditions, communication exchange is carried out on various communication channels, namely the optical channel (Landwehr, 1997), the acoustic channel (Strube and Lazarus, 1997), the chemical channel (Kröller, 1997), the thermic channel (Brück, 1997), the tactile channel (Heuer, 1997), and the electric and magnetic channels (Moller, 1997). Spoken language is transmitted via the acoustic channel, written language via the optical one. The tactile channel is used for nonverbal communication means like handshaking or conversational touching (Lomranz and Shapira, 1974). Information related to temperature, taste and smell are accessed via thermic, and electric and magnetic channels, respectively.

Humans are equipped with sense organs that facilitate perception of signals from the various channels: the eyes facilitate seeing (visual channel), the ears facilitate hearing (acoustic channel), receptors within the nose and the mouth (tongue) facilitate smelling and tasting (chemical channel), dermal receptors facilitate sensation of heat (thermic channel) and touch (tactile channel) (see e.g., Schmidt and Thews, 1997).[1]

On the production side, the main means for outputting signs are the vocal tract for vocalisations, including spoken language, and motor behaviour, most prominently, of the hands, for any sorts of imprints, be it written language or touch. Part of the latter production means are also, for instance, eye movements or body posture. However, in the context of barrier-free technical communication, such nonverbal cues can be put aside due to their marginal role in contributing linguistic meaning (see Allwood and Ahlsén 2012, Martin and Schultz 2012, and Lücking and Pfeiffer 2012 all in this volume for multimodal communication).

Against this background, the term “barrier” can be defined as follows: a *barrier* is any obstacle that inhibits the transmission of information of one or more channels. Two further refinements are achieved by the following questions:

1. Where is the *source* of the barrier? It is inside or outside of a communicator?
2. What is the *direction* of the barrier? It is an obstacle for perception or production?

As to the first question, the obstacle that constitutes a barrier for communication can be inside or outside of a communicator:

- *Barriers inside of communicators.* The main focus of barrier-free communication is on people that due to some cognitive disability or some disability of a sense organ are not able to communicate via a certain channel. The barriers therefore are "inside" of the communicator, either because of an innate disability like cerebral palsy or because of an acquired disease like a stroke or a brain injury (Carter, 1996). An essential part of research in this area is to develop technical communicative assistance for impaired communicators, also known as *Augmentative and Alternative Communication* (AAC).
- *Barriers outside of communicators.* Barriers can also be imposed by the communication situation. Manifold aspects of the specific physical communication setting may impose barriers on the different communication channels. For example, the visual channel is nearly unusable in the dark, the thermic channel is ineffective for aquanautic communication, and the tactile channel is inoperative at distance. Thus, barrier-free communication does not only concern technical assistance for impaired people, it also applies to situative constraints outside the communicators.

In particular the latter class of barriers outside of communicators illustrates that barrier-free communication is not limited to disabled users. This point is explicitly addressed in the *Guidance on accessibility for human-computer interfaces* published by the *International Organization for Standardization* (ISO): "Although 'accessibility' typically addresses users who have a disability, the concept is not limited to disability issues." (ISO/TS 16071: 2).

The second question clarifies whether a sense organ or a production means is affected by a barrier. Although there are various communication channels for communication, each of which can be obstructed for some reason, the main focus of barrier-free communication is on spoken and written language: "Augmentative communication means any method of communicating that supplements the ordinary methods of speech and handwriting, where these are impaired." (Millar and Scot, 2003: 3). For this reason, the main focus of barrier-free communication is on barriers for producing or receiving spoken, written and signed language, viz. obstructions of visual and acoustic sense organs on the one hand, and restrictions of motor behaviour on the other hand.

Notwithstanding the first distinction of barriers inside vs. barriers outside of communicators, we focus on communicative challenges for people with disabilities (i.e., inside barriers), since they constitute the main and often the single explicitly named target group for barrier-free applications.

In short, this article deals with barrier-free communication means for barriers inside communicators, that is, for impaired users, that affect both the production and the reception side mainly of written and spoken language.

In this line, augmenting or substituting written, spoken or signed language by means of other physical forms of communication, can be seen as supporting barrier-free communication (for an introduction into *Augmentative Communication* in this sense see Wilken, 2006).

In this article, however, we deal with AAC (*Augmentative and Alternative Communication*) as technical or electronic communication aids. The beginning of the technical AAC can be traced back into the 1950s when the first assistive devices were developed (Glennen, 1996: 3–21). Since then, according to Schweinberger (2009: 77) applications for AAC have accomplished three milestones:

1. communication boards with graphical symbols,
2. acceptance of sign language as a natural, full-fledged language,
3. special writing devices and environmental controlling systems for people with physical impairments.

One central term in the context of AAC is, of course, *accessibility*. According to the ISO *Guidance on accessibility for human-computer interfaces*, accessibility is "usability of a product, service, environment or facility by people with the widest range of capabilities." (ISO/TS 16071: 2). This means, that a document accessible by one target group is not necessarily accessible by another group. In order to achieve maximal accessibility, a document should be "readable" via various channels. This definition is essential from the point of view of software applications, including web browsers: *the task of AAC devices is to provide at least one alternative access to some given content.*

The related term "usability" (coined by Jacob Nielsen 2000), however, "usability" does not necessarily capture the specific needs of people with disabilities. We take the terms "accessibility" (as mostly used in the context of information technologies) and the term "barrier-free" to fill this gap. Both terms are used interchangeably here.

The concept of "universal design" tries to unify the concepts "barrier-free", "accessible", and "usable". Universal design denotes a design concept according to which technical devices as well as products and artificial environments are to be designed in such a way that they are usable by as many people as possible without further aids.[2] However, the term "universal design" as well as its associated concept is not yet established.

A final remark on the body of source material. Applications for and research on AAC are widely scattered. Although there are some handbooks on this general topic (Glennen, 1996; Seok et al., 2010; Stephanidis, 2009; Wilken, 2006) and on more specific selected topics (Edman, 1992; Millar, 1997; Hersh and Johnson, 2008), much information is spread via websites, either of companies or of public institutions or organizations. For this reason, our references include a great share of URLs. In addition, we occasionally also draw on information which stems from personal communication with users of barrier-free technology. One of the authors (P.K.) discussed several issues of AAC on the LinAccess mailing list in order to figure out the users' opinions, experiences, and preferences. These discussions can be looked up after having subscribed on the list.

1.2. Outline

It is not possible to present all hardware and software products available in the field of assistive technology for communication. The German rehabilitation database *REHADAT*[3] alone lists 2314 items in their class No. 22 on *Aids for Communication and Information* that, according to the ISO classification on *Assistive products for persons with disability*, are classified as "assistive products especially produced, or generally available, for persons with disability" (ISO 9999:2011). Instead, we will give an overview of a selected group of technical aids with special focus on web technologies. We cover applications and methods in the area of technical AAC. In addition, we also deal with problematic cases in which current applications are not successful to a workable degree. The article is structured in the following way:

Section 2 deals with the main aspects of barrier-free communication. Technical aids for communication applied in the WWW are discussed in Section 2.1. In this context, Section 2.2 deals especially with free software, since it has specific advantages and is for the most part overlooked in the literature. The use of complex language poses a barrier in itself for various target groups. We take up this linguistic barrier in Section 3.1. In Section 3, we present some selected disability-specific applications. Introducing of sample applications is done according to the following scheme: for each disability

1. the character of the disability is described;
2. the kind of technical aid developed for the case in question is identified and exemplified by a list of applications available; and
3. directions of research are outlined.

In this vein, we distinguish the following broad classes of disabilities: Section 3.2 discusses technical aids for the blind and visually impaired people. In Section 3.3, we deal with barrier-free communication for the deaf and hard of hearing people. Section 3.4 introduces aids for the deaf-blind and Section 3.5 for people with motor impairments. Thus, sections 3.2 to 3.4 cover the perception side of AAC, whereas Section 3.5 deals with production in accordance with the scheme outlined above. Finally, in Section 4 we summarize the article and give an outlook for future work in improving barrier-free communication.

2. Main directions

We identify four technological main directions for barrier-free communication: firstly, barrier-free communication within the WWW (Section 2.1), including semantic and social web applications (Sections 2.1.2 and 2.1.3); secondly, barrier-free communication and free software (Section 2.2); thirdly, accessibility and PDF documents (Section 2.3); and fourthly, barrier-free applications from the point of view of the developer (exemplified by Java in Section 2.4).

2.1. Barrier-free communication in the World Wide Web

According to a study conducted on behalf of *Aktion Mensch*, the largest private German organization for advancing disabled people, people with disabilities regard internet applications as helpful for overcoming handicap-specific barriers (Cornelssen and Schmitz, 2008). In particular interactive web applications (as part of Web 2.0) were rated positively. Accordingly, the 671 disabled people that have been interviewed as part of the study use the WWW to an above-average extent.

However, the study also revealed that especially language poses a larger barrier than commonly assumed – and it does so for people with and without disabilities. Due to the comparatively small competence for written language that results out of various disabilities, handicapped users regard many online texts as too diffcult and therefore only take part in web communication to a lessened extent compared to other users.

In spite of multiple diffculties, the study also shows that people with disabilities are highly creative when facing a barrier. This observation points to the fact that the range of possible disabilities is very large and that it is hardly feasible to create a single barrier-free environment that fits all possible accessibility problems. Rather, there are a couple of specialized devices that try to circumvent a specific barrier. In concert, these devices aim towards the general goal of achieving the greatest possible accessibility, as formulated by the *Web Accessibility Initiative* (WAI). The WAI represents the executing power responsible for the development of standards for barrier-free WWW. In the following Section 2.1.1 we briefly introduce the WAI, and subsequently focus on the overall issues related to barrier-free web applications.

2.1.1. The Web Accessibility Initiative (WAI) and standardization

The *Web Accessibility Initiative* was initiated in February of 1997 as a "Research and Development Working Group" within the *World Wide Web consortium* (W3C)[4]. The goal of WAI is to develop standards and guidelines for the overall accessibility of web contents for various target groups by means of us-

ability technologies. Alongside people with disabilities, WAI serves widely spread groups of interest, for example, business communities, non-business organizations, academic groups and governmental institutions.
The technologies dealt with in different WAI projects include, but are not limited to, authoring systems (eLearning systems, Wikis, graphical source code editors), evaluation tools and user agents (for example, web browsers, assisting technologies), as well as interactive web applications.
The WAI, amongst others, publishes *The Web Content Accessibility Guidelines* (WCAG),[5] which, in its most recent version WCAG 2.0 formulate the following four principles whose compliance helps to achieve barrier-free applications (see also at WCAG 2.0):

Perceivable "Information and user interface components must be presentable to users in ways they can perceive." (http://www.w3.org/TR/WCAG20/#perceivable, accessed 18.01 2012).

Operable "User interface components and navigation must be operable." (http://www.w3.org/TR/WCAG20/#operable, accessed 18.01. 2012).

Understandable "Information and the operation of user interface must be understandable." (http://www.w3.org/TR/WCAG20/#understandable, accessed 18.01. 2012).

Robust "Content must be robust enough that it can be interpreted reliably by a wide variety of user agents, including assistive technologies." (http://www.w3.org/TR/WCAG20/#robust, accessed 18.01. 2012).

Each of these principles is elaborated by a couple of guidelines. The guidelines in turn are assigned with criteria of success. These criteria of success are subdivided into three conformity constraints.[6] The higher the level of conformity, the more demanding it is to successfully fulfill the respective constraint. Besides these levels of conformity, there are additional global conformity constraints documented extensively for the developers' understanding in order for them to fulfill the guidelines. Especially for large projects with multiple content levels and languages involved, it is nearly impossible to achieve a fully compliant barrier-free representation. For such cases WCAG 2.0 offers optional conformity explanations that can be used by the website provider not only to make the conformity status of the site explicit but also to make possible hindrances visible. An example for hindrances could be dynamic, non-standard compliant contents offered by a third party which are embedded in the website.

In order for developers to test the usability of their projects in terms of freedom of barriers, WAI also provides a wide range of training materials for developers, for example the *Implementation Plan for Web Accessibility*[7] or the *WAI Resources on Developing Web Accessibility Training and Presentations*[8].

2.1.2. Semantic Web and accessibility

Semantic Web, a term coined by Berners-Lee et al. (2001), collectively refers to collaborative work that aims at a merging of different interfaces and technologies (for example, ontologies, representation formats) in order to provide a better access to related knowledge, especially by including *semantic* content of web pages. In other words:

> The goal of Semantic Web research is to transform the Web from a linked document repository into a distributed knowledge base and application platform, thus allowing the vast range of available information and services to be more effectively exploited. (Horrocks, 2007: 1)

In this section we investigate how far this idea holds for barrier-free communication. First, we focus on *markup languages*. Second, we give a short overview of some research directions from 2004 to 2011 working on optimization of accessibility by means of Semantic Web technologies.

2.1.2.1. Semantic Markup for barrier-free web access

Semantic markup in HTML *inter alia* supplies users of screen readers with information about the structure of the content. Already in 2000 "[…] the Web Accessibility Initiative (WAI) joined forces with the W3C HTML Working Group in the design of HTML4.0." (Jacobs et al., 2000). The WAI identifies three fields of work, that are addressed by the initiative, namely (quoted from Jacobs et al., 2000):

1. Unstructured pages, which disorient users and hinder navigation.
2. Abuse of structural HTML elements for purposes of layout or formatting.
3. Heavy reliance on graphical information (for example, images, image maps, tables used for layout, frames, scripts, etc.) with no text alternatives.

HTML is the first layer users are confronted with. Detailed information about the text structure of a text document is required, for instance, by screen-reader users, but this information is not provided by a simple application of the current web standards. Additional structural or dynamic information can be provided by means of HTML markup and CSS. For example, CSS makes it possible to use non-visible navigation headlines,[9] that can be exploited by screen readers. Furthermore, the *Accessible Rich Internet Applications Suite* (WAI-ARIA)[10] provides a specification of an ontology for accessible interfaces. Due to WAI-ARIA information, accessibility frameworks are able to map marked elements onto alternative access solutions. Thus, markup via WAI-ARIA gives static text and dynamic elements additional meaning that can be exploited by technical applications. Future markup via HTML 5 also aims at similar goals.[11]

Going beyond the level of markup, the Semantic Web offers additional means of web accessibility. XML-based work, for example, the *Resource Description Framework* (RDF) or *Web Ontology Language* (OWL), implements a second layer, a layer to which users usually have no direct contact. Rather, XML-based languages and ontologies are used for machine processing and can therefore be exploited by applications that are related to accessibility. The gain of Semantic Web frameworks for barrier-free communication is already incorporated in the concept of the Semantic Web, as is pointed out by Horrocks (2007: 5) who explicitly mentions the "benefit to users with cognitive or sensory impairments".

In the following Section, we present some examples of Semantic Web technologies that are employed in the context of web accessibility.

2.1.2.2. Semantic Web applications assisting barrier-free communication.

Four research areas in the domain of Semantic Web applications and barrier-free communication can be distinguished:

- developing web accessibility evaluation tools (Abou-Zahra, 2005);
- developing web authoring tools to ease the development or editing of accessible websites, in particular for users with no or minor programming skills (Harper and Yesilada, 2007);
- customization and personalization of web resources (Yesilada et al., 2008);
- supporting the development of applications in the afore-mentioned areas by means of the Social Web (Appelquist et al., 2010).

The fact that websites generally lack accessibility is a well known problem (see, e.g., Romano Jr., 2003). In addition to the development of evaluation tools and authoring tools, tools that enable the user to render web pages directly and transform them into accessible formats is an alternative approach (cf. Seeman, 2003; Harper and Yesilada, 2007; Yesilada et al., 2007).

An additional challenge is the preparation of web-based resources that focus on special needs. Lopes et al. (2009) develop an extensive framework, the *Semantic Accessibility Assessment Framework* (SAAF), based on various ontologies. According to Lopes et al. (2009: 2) there are several approaches utilizing "ontological concepts and taxonomies for disabilities". These ontologies focus on different aspects that are combined in SAAF. SAAF should provide a generic and extensible means accounting for the dynamics of newly emerging standards and web technologies.

A view that focuses more on customization is represented by Sohrabi (2010), whose framework can be customized to fit individual preferences. In a related way, Rico et al. (2009), pursue the approach of personalized web applications. Their approach is to create user profiles adapted to the individual needs of a user.

2.1.3. Social Web and accessibility

The *Social Web* or *Web 2.0* is characterized by interactive technologies of collaborative production of content, supported by social software (see Waltinger and Breuing 2012 in this handbook). Web 2.0 not only inherits its web accessibility problems from the WWW, it also brings along some genuine barriers. According to Cornelssen and Schmitz (2008), disabled people struggle mostly with CAPTCHAs (*Completely Automated Public Turing Test to tell Computers and Humans Apart*), user guidance, language, and inconsistency.

However, there are a couple of applications and research projects settled in the context of accessible Social Web, for example:

– The *DictaSign* project[12] (Efthimiou et al., 2009) aims to improve interaction by means of sign language (see also Section 3.3.2).
– *Social Media For All* (SOMFA) accounts for the multimedia-based processing of social media content and its final retrieval by means of TV-media (Borrino et al., 2009). Users who do not have access to the internet "regardless of their age or their possible technological, geographical, sociological, or physical limitations" (Borrino et al., 2009: 54) can gather information from social media in terms of audio or video output. SOMFA does not modify the source code of the website but collects social media contents independently of the underlying standards (e.g., RSS) and prepares them for presentation on television.
– *HearSay3/VXMLSurfer*[13] is a non-visual web browser designed especially for visually impaired, blind and deaf-blind users (Ramakrishnan et al., 2004). The goal is to recognize and support multi-lingual environments, personalized and collaborative labeling (the main feature of social web applications), and to facilitate access to dynamic web contents. The HearSay3 browser segments web pages into semantically meaningful units (Borodin et al., 2008). The automatic recognition of web contents uses collaborative labeling to label the elements not tagged in HTML. For example, many query forms are submitted by clicking on a "Go" button. This button does not have to be tagged in HTML. In this case, however, the button can be labeled by another user. This "social tag" can then be correctly interpreted by HearSay3. Users can define their own labels or retrieve labels provided by other users from remote repositories. This example shows how social web technologies are used to support accessibility.

2.2. Barrier-free communication and free software

The concept of free software has been pushed forward by Richard Stallman and the GNU project. It has to kept separate from “Freeware”, “Open Source”, as well as “Public Domain”. Free software according the GNU project is published under the *GNU General Public License* (GPL)[14]. The GPL rests on the following four levels of “freedom” (quoted from http://www.gnu.org/philosophy/free-sw.en.html, as of 19.01.2012):

- The freedom to run the program, for any purpose (freedom 0).
- The freedom to study how the program works, and change it so it does your computing as you wish (freedom 1).
- Access to the source code is a precondition for this. The freedom to redistribute copies so you can help your neighbor (freedom 2).
- The freedom to distribute copies of your modified versions to others (freedom 3). By doing this you can give the whole community a chance to benefit from your changes. Access to the source code is a precondition for this.

Meanwhile, a wide range of assistive technologies is implemented as free software, for example:

- *Audio Desktop Reference Implementation and Networking Environment* (ADRIANE)[15], a complete non-graphical desktop system for blind people;
- the *LinAccess* project[16], which is concerned with the overall distribution of barrier-free software;
- the *VizAudio* project[17], which optimizes desktop accessibility for deaf people;
- the KDE Accessibility project[18];
- the GNOME accessibility project[19].

KDE and GNOME are themselves free software projects, namely the two most well known graphical desktop environments for free operating systems.

Since free software is independent from proprietary formats and software companies, publishing technological applications under the GPL is a highly sustainable solution (Grassmuck, 2004: 11).

Nevertheless, conflicts arise in everyday work. Not all application scenarios have free alternatives. Furthermore, legal provisions can entail additional diffculties. We will exemplify these diffculties by the example of PDF documents.

2.3. Barrier-free communication and PDF documents

The initial goal of the *Portable Document Format* (PDF), created in the early 1990s, was to provide an operating system independent format for visualizing and printing documents.[20] As a result, there was seen no need to support automatic retrieval of PDF documents (for example, via screen readers). Later, in

2005, with the emergence of the PDF/A-1a standard, more restricted criteria based on the US Government, *Section 508 of the Rehabilitation Act*[21], were elaborated to ensure access to documents for users of assistive technologies. This development made it necessary to provide tools for creating PDF/A-1a and converting non-standard compliant PDF to PDF/A-1a.

PDF documents are currently used for the exchange of text files as well as of complex interactive documents. Registration forms or tax assessments are common examples of the interactive use of PDF enhanced with other applications. For this reason, PDF is also of importance for barrier-free communication.

The requirements for a PDF/A-1a document are very strict. To this end, the following minimum features should be met to produce an accessible PDF document (BIK-Projekt, 2010):

– create a tagged-PDF (logical structure markup) document,
– define the document's main language and its encoding,
– create the document's title and visualize it via PDF-reader,
– account for a correct order of tabs and the reading direction (for example, for double column texts or tables),
– create useful bookmarks,
– provide alternative text for images,
– guarantee the text's consistency in linear layout,
– special requirements on activated security mechanisms.

In the following we discuss chances and limitations of creating and editing barrier-free PDF documents by means of free software.

Reading barrier-free PDF documents with free software PDF readers based on the *Gimp Toolkit* (GTK)[22] offer the possibility of reading meta-tags from PDFs created via OpenOffice.org or pdfLaTeX. Tagged PDFs can be interpreted by the free screen readers *Orca*[23] and *SUE* (*Screenreader & Usability Extensions*)[24]. Via the linaccess mailinglist, we asked a testee to evaluate the barrier-free PDF reading tools. The testee returned three results: (1) Evince does not work together with Ocra, (2) Okular, since being based on qt, is not even able to interact with Ocra, and (3) the testee was not able to access PDF contents when using Okular. In everyday usage, the testee used Acrobat Reader on Windows machines, while converting PDF documents into plain text by means of *pdftotext*[25] on Linux machines. However, the latter approach presupposes that the PDF does not contain non-text elements like tables or figures (Sebastian Andres, p.c.). For more information, including non-free software, see Heuwinkel (2003).

Creating PDF documents OpenOffice.org[26] has a built-in export-function into the PDF/A-1a format since version 2.4. In order to maintain editability of an

OpenOffice document, OpenOffice.org recommends to combine the use of the PDF-import-plug-in with the option to create PDFs in a hybrid PDF format. Hybrid PDF contains the embedded ODF (*Open Document Format*) document within the PDF file. Note that objects like tables or figures cannot be exported as barrier-free elements. A guide for creating barrier-free PDF documents by means of various programs and tools is provided by Hellbusch (2005).
Darvishy et al. (2010) developed a tool that structures MS-Word[27] documents in such a way that they can be exported as barrier-free PDF documents. There is no such tool available so far for OpenOffice.org or LibreOffice[28].
In March 2010 the free-of-charge tool PDF-Accessibility-Checker (PAC)[29] has been released for checking and ensuring technical accessibility of PDF documents. While it is free to download and use, this tool is not, however, an open-source software product. Currently, there is no open-source software tool available for testing or editing the barrier-free features of a PDF document.

PDF and platform independence While there are PDF *readers* for any software platform, interpreting meta-tags (for example, from a PDF/A-1a file) for accessibility content is facilitated by devices specific to operating systems. On Windows systems, PDF meta-tags are processed by the *Microsoft Active Accessibility* (MSAA) interface[30] and then transmitted to the corresponding output device.
On Linux-based platforms using the GNOME desktop there is the GNOME accessibility project [31] which provides the *Assistive Technology Service Provider Interface* (AT-SPI) and the *Accessibility Toolkit* (ATK).

In sum, in spite of their advantages in terms of sustainability, free PDF accessibility alternatives usually have less functionality than their proprietary counterparts. In addition, barrier-free PDF applications are hardly platform independent.

A platform independent environment is provided by the Java programming language, which is available for various operating systems. We list Java's accessibility devices in the following section.

2.4. The Java programming language and accessibility

Java offers a wide range of devices to develop barrier-free applications (Heinrich, 1998). Especially in the area of communication technologies there are various interfaces:

- *Java Communication API* [32] (e.g., for Smart Cards, embedded systems, point-of-sale devices, financial devices, fax, modems, display terminals and robotic equipment),

- *Java Telephony API* (JTAPI)[33] (provides telephony call control),
- *Java Speech API* [34] (for speech synthesis and speech recognition),
- *Java Message Service* (JMS).[35]

Java provides two methods to develop barrier-free applications: a) the Java Access Bridge (Zukowski, 2005), or Java SWING, and b) the SWT API[36]. The two variants differ in the range of possibilities offered to enhance applications.

Java Access Bridge/SWING. The Java Access Bridge serves as a mediator between the Java application and the assistive technology (screen-reader) on the particular operating system. All SWING *graphical user interface* (GUI) elements can be made accessible via the Access Bridge. However, the effort is higher since the developers have to explicitly name all SWING elements they use (see Oikonomou et al. (2010) for a listing of tools available in support of developers aiming to design barrier-free applications for the web and more). Furthermore, the Access Bridge has to be updated simultaneously with the *Java Runtime Environment* (JRE).

SWT classes. The SWT classes do not require the Access Bridge as the mediator, they communicate directly with the native accessibility interface of the operating system. This fact ties the application bound to the particular operating system; all system specific elements need to be known in advance, otherwise the application might not be accessible on this system.

The following basic aspects should be considered by the developer (independently of the programming language used) – cf. the Java tutorial on *How to Support Assistive Technologies*[37]:

- GUI elements are to be accessible by keyboard and/or by mouse;
- GUI elements are to be structured in a reasonable way in order to facilitate a barrier-free keyboard usage;
- large data sets are to be partitioned in a meaningful way;
- keyboard short cuts are to be set in a reasonable way (for example, an official input form contains about 30 input fields, out of which only 10 are mandatory. The ‘OK’ button is placed last at the bottom of the input form. Without short cuts, a blind user having input the mandatory information would have to tab 20 times before reaching the end of the form. A reasonable short cut would be to bypass the optional input fields.).

Validators that check for accessibility issues support the realization of Java-based barrier-free applications. In order to abstract over platform-specific Java constraints, validators test applications by means of screen readers that are independent of the Java method used for implementing the applications. SWING-based applications can be tested via “Monkey” or “Ferret”. Monkey validates the component tree and the access to the accessibility objects. Ferret uses the

Java Accessibility Utilities and supports, for example, the testing of different screen reader modes (Zukowski, 2005). To test SWT-based applications the *AccExplorer*[38] or *UISPY*[39] (both for Windows) can be used.

In general, Java offers a useful set of tools needed to develop barrier-free applications. However, the platform independence is not always supported due to system specific realizations. In any event, the particular operating system's possibilities and the understanding of the specific user group's needs (with its strengths and limitations) are indispensable for developing reliable barrier-free applications.

3. Selected handicap-specific applications

Barriers inside communicators arise out of some malfunction of a sense organ or a production device – see Section 1.1 above. In this section, we address the different disability-specific approaches to accessibility. In subsequent sections we will firstly focus on perception directed barriers, namely in terms of learning impaired, blind, deaf, and deaf-blind target groups. Secondly, production directed barriers are dealt with in a section on physically handicapped target groups. Each section first describes the disease with its specific barriers, presents the tools available to overcome the barriers, and outlines the directions of research in this area.

3.1. Barrier-free communication (not only) for people with cognitive or learning difficulties

Compared to barriers imposed due to visual or auditory impairments of users, accessibility problems of people with learning or cognitive difficulties are less considered by software developers (Bohman, 2004). As noted in Section 2.1, users with these difficulties may have problems in understanding the language used on websites. Accordingly, the technological task set by these linguistic barriers is to make the websites' language as clear and easy to read as possible.

3.1.1. Broader projects

In addition to mere software- or hardware-based applications, there are some broader projects designed for people with cognitive or learning difficulties that include some social embedding. One such framework is the learning platform *"on-line"*.[40] The *on-line* platform has been founded in 2007 as part of a European educational program. The *online* homepage provides, for example, information about how to learn, techniques of how to use a computer, an easy to use e-mail client and an all-over layout according to special needs of cognitive-disabled.

Another software project aiming at needs for people with learning difficulties is *Boardmaker*.[41] The Boardmaker software draws on the symbol library of Mayer-Johnson LLC and facilitates the realization of icon-based, easy-to-read communication media, for instance, for autism therapy or as part of barrier-free websites. For an example of the use of Boardmaker in barrier-free applications see the library services project *Across the Board* of the library of Leeds.[42]

In this section, however, we focus on means of reducing linguistic barriers in technical communication. First, we introduce an approach to reducing language complexity making the content more accessible. Second, we will discuss projects aiming to alleviate barriers for people in particular from poor countries.

3.1.2. Applications and directions of research

3.1.2.1. Easy-to-read

Within the *Inclusion Europe* project,[43] which is supported by the European Commission, the *European Easy-to-Read Guidelines* have been developed in 1998 (Frexhoff et al., 1998). "*Easy-to-read*" stands for conceptual text production according to the special needs of people with learning difficulties or literacy problems. The goal was to facilitate the access to official and legal texts, dictionaries, historical and cultural writing for people who otherwise were not able to take part in this kind of communication.

The European Guidelines (Frexhoff et al., 1998) give advice for how texts in *Easy-to-Read* should be created in a list of six phases of the publication process, both conceptually and with regards to content. Phase No. 4 explicitly addresses text production for people with learning difficulties: "Check whether people with learning disabilities can understand your draft. Ask people with learning disabilities to read your document before it is printed." (Frexhoff et al., 1998: 14). Taken together, texts that have been produced according to the *Easy-to-Read* guidelines can generally be characterized by four features (Frexhoff et al., 1998: 8), namely "the use of a simple, straightforward language", "only one main idea per sentence", "the avoidance of technical language, abbreviations and initials", and "a clear and logical structure." websites that adhere to the *Easy-to-Read* standards of text production are marked by the *Easy-to-Read* logo, which is displayed on the right.

More extensive recommendations for readability are edited by Inclusion Europe under the heading "European Standards for Making Information Easy to Read and Understand". These standards concern the preparation of written information, electronic information (for example, websites or CD-ROMs), video and audio information in such a way that they are easy to process (cf. Inclusion Europe, 2009).

3.1.2.2. Technologies compensating illiteracy

The *Simputer*[44] is a Linux-based *Personal Digital Assistant* (PDA) developed in 2001. The device was developed to assist illiterate people by offering them a connection to the internet. Available at a low price, the device supports multi-user functionality by means of a readable and writable smart-card access.

The user interface of the Simputer, *Malacca*, is based on the *Information Markup Language* (IML).[45] IML was designed to support various web applications in order to provide access for illiterate people and to facilitate interoperability between HTML and WML.[46] Malacca's graphical interface allows the user to operate the Simputer by means of graphical input as well as by speech input.

The Simputer has been developed especially for the third world: it shall ensure that illiteracy is no longer a barrier that hinders using a computer. However, the Simputer remained widely unaccepted by its target group (Pal and Fonseca, 2003).

The idea to help people (in particular, children) from poor countries by improving their chances in education by means of a personalized system is pushed forward by the project "One Laptop per Child" (OLPC).[47] Despite country-specific problems, OLPC is an ongoing worldwide project – according to the project's website, "over 2.5 million children and teachers have xo laptops".[48]

3.2. Barrier-free communication for the blind and visually impaired people

The *World Health Organization* (WHO) estimated in May of 2009 that the number of visually impaired people is about 269 million and that there are approximately 45 million blind people, out of which 87 % live in the developing world.[49] The WHO distinguishes the following four levels of visual impairment:

1. normal vision,
2. moderate visual impairment,
3. severe visual impairment,
4. blindness.

Depending on the severity of the disease, the third level may be further subdivided (Slawinski, 2005). In this article, we concentrate on aids for blind people and people with a visual impairment who are still able to retrieve a standard text with a 6-fold zoom (compare Figure 1).[50]

3.2.1. Problem description

The point in time when a person becomes blind influences the development of their linguistic and auditory skills (see for example Wakefield et al., 2006; Stevens and Weaver, 2005). In addition, the acceptance of technical communication aids is besides, for instance, personal attitudes also influenced by the du-

Figure 1. A website of the size of 1024 × 768 pixels (left) compared to the same website in a 4-fold zoom (right) in the Firefox browser.

ration of seeing experiences made before becoming blind (for example, Braille and Braille displays on PCs are less preferred by lately blinded people).

3.2.2. Tools

3.2.2.1. Screen reader

Screen readers for accessibility process data from the operating system or from applications (for example, via MSAA or *IAccessible2* on Windows, and via AT-SPI or ATK on Linux – see also Section 2.3). This data is sent to an alternative output device like speech output or Braille display. Screen readers work in two different modes: the tracking mode allows to make a selection following the user's focus, the review mode allows one to navigate through the environment (Kochanek, 1994).

Successful transmission of data to the user depends not only on the screen reader, but also on the delivery of data from applications to the reader's interface. Processing of complex graphical interfaces still poses a challenge for screen readers. Depending on the output (speech output or Braille display) different models of visualization are implemented. Speech output is realized by means of selection models (representation of contents by relevance), structure models (objects and their relations represented as a tree) or a combination of both.

3.2.2.2. Speech output

Generally, visually impaired users face two kinds of speech output, distinguished by the method involved in speech synthesis (see Martin and Schultz 2012 in this handbook on techniques of speech synthesis):

- rule-based speech synthesis (fully synthesized output, speech output sounds rather artificial or unnatural);
- data-based speech synthesis (also called digital speech synthesis); pho-

nemes are produced by a human speaker with the goal to achieve an authentic pronunciation.

The user's preference mostly depends on the application context (for detailed information on user preferences regarding screen readers see the three surveys administered by WebAIM (*Web Accessibility in Mind*)[51]). Rule-based speech synthesis is preferred for accomplishing information tasks or for everyday work. Proficient screen-reader users can deal with high speech rates. Data-based speech synthesis is hard to understand at high speech rates. Natural voices are generally preferred, for example, for listening to narratives.

3.2.2.3. Screen magnifier

A screen magnifier is a tool for enlarging the display on the monitor – see Figure 2 for an example. Screen magnifiers can be configured in various ways. Possible customizations include, but are not limited to, changing the color, form, or size of the cursor, setting the resolution of the underlying reticule, specifying zoom options, color modifications of the whole display, and optional speech output. Note that a screen magnifier is not to be confused with a magnifying glass; the latter enlarges selected portions of the display, while the former augments the whole display.

Figure 2. Screen magnifier of the screen reader *Orca.*

3.2.2.4. Assistive technology by means of Braille

In its "written" (embossed) form, each letter of the Braille alphabet consists of six printed dots on embossed paper. PC Braille displays on the other hand consist of eight dots in order to map the extended ASCII alphabet of 256 characters.

Figure 3 exemplifies embossed Braille by means of the term “barrier-free”, where the black dots represent the relief-like texture that can be read with the finger tips.

Although Braille could in principle be an input means, Braille displays are primarily used as output devices. Many Braille applications also support audio controlling (e.g., of the cursor) (Hersh and Johnson, 2008: 143).

It is important to note that the users of Braille displays have to have enough haptic capability in their fingers. Especially elder people can have problems with the tactile sense caused by illnesses (e.g., diabetes).

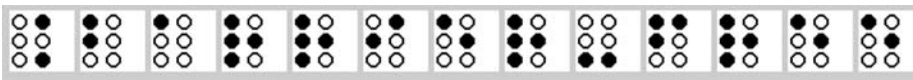


Figure 3. “Barrier-free” written in Braille (generated on http://www.blista.de/bd/braille/index.php on March 17, 2011).

3.2.2.5. Haptic displays

Haptic displays make visual information accessible through the tactile channel. The general procedure is to produce a texture on some medium that can be sensed with (typically) the fingers. In addition, these haptic displays can be augmented by acoustic information. The range of haptic displays is exemplified by means of the following examples of three kinds of applications (see Edman (1992) and Millar (1997) for comprehensive overviews):

Displays on microcapsule paper. Tactile maps are produced with prepared paper (see e.g. Edman, 1992: 82f.). White paper is coated with millions of microscopic capsules. The display is then printed in black on the prepared paper. When heated, only the capsules beneath printed black areas expand, giving rise to a “relief” representing the original, printed display. The relief structure can be recognized via touching. One advantage of microcapsule displays is that additional information like verbal descriptions can easily be added, for instance by means of printing of Braille letters.

Computer-controlled pin displays. The graphical display on a computer monitor is simplified by an image reduction process and prepared for output on a touchable display of a certain resolution (Edman, 1992: 100f.) – the electronically steered display of the GWP from Figure 4 uses a screen made out of 24 × 16 pins.

Tactile web browsing. Based on information encoded in HTML 5, CSS or JavaScript (cf. Section 2.1.2), rendering applications process web documents and prepare a non-visual output thereof. The output can either be a tactile display like Braille text (by means of some computer-generated display, see above) or synthesized voice (see e.g. Rotard et al., 2008).

Figure 4. The electronic pin display *Graphic Window Professional* (GWP) displaying the "@" sign (figure taken from https://www.handytech.de/index.php?id=26&L=0, accessed June 1, 2011).

3.2.3. Applications and directions of research

3.2.3.1. Digital Accessible Information System (DAISY)

DAISY[52] is an XML-based format for synchronizing audio and text, which is used for creating digital audio books in particular for blind people (ANSI/NISO Z39.86–2002). The format is designed to provide audio books with meta information that allows users to navigate through it. DAISY books can be played either on DAISY players or on computers that are equipped with DAISY reader software. Turning to production, simple documents that comply to the DAISY format can be produced by means of extensions for word processors (e.g., by means of *odt2daisy* in OpenOffice.org or *Save-as-DAISY* for Microsoft Office).[53]

A DAISY book usually consists of the following parts (for more details see Brzoza, 2008; Eberius, 2008):

1. one or more audio files (containing texts verbalized by a human speaker),
2. an XML file containing the whole or the parts of the text,
3. a synchronization file that connects text segments to their corresponding sections of the audio file by means of specific markup,
4. a file containing the navigation information.

The DAISY format is widely used in the domain of accessibility documents. The following tools exemplify current applications and research projects:

- The *gh* player can deal with DAISY audio books that include mathematical content.[54] Mathematical expression are represented in terms of MathML. “The *gh* player presents DAISY book content, including MathML, simultaneously in three sensory modes: audibly, visually, and tactile. The player supports both a fallback presentation of MathML and an advanced native MathML presentation.” (Leas et al., 2009: 3). The fallback presentation in this case is a rendered image.
- Export into DAISY formats from an Office document by means of word processor extensions allows one to easily create DAISY books out of text files (Strobbe et al., 2010).
- Developing a DAISY online standard: accounting for an online connection of a DAISY-based service and a player. The goal of the online standard is to access DAISY books directly from the library via streaming. The DAISY online standard includes a range of features, for example, for security of data (data transfer via https), for service models (loan handling), and for selection of DAISY audio books (via a mobile phone, a website or a DAISY player itself) (McMullen and Fitzpatrick, 2010).
- Preparing DAISY books with a word processor is usually limited to a single author. In order to allow for multi-authored documents, Bernier and Burger (2010) developed a Wiki-based environment for producing DAISY texts by means of a simplified markup language.

3.2.3.2. Sonification

The term *sonification* denotes the acoustic non-speech representation of non-acoustic data. In barrier-free applications, acoustic signals are used within the framework of *Auditory Information Seeking Actions* (AISAs), which utilizes verbal and non-verbal audio signals within a task-oriented scenario. The non-verbal signals are sometimes called “earcons” in analogy to “eyecons”, that is, “icons” (Blattner et al., 1989).

Sonification is employed in the following research projects:

- The tool *iSonic*, developed for exploring geo-referenced data, represents the five layers of a choropleth map by means of five different violin pitches (Zhao et al., 2008). Other navigation cues, for instance, when a user leaves the operative map, are also sonically coded. Zhao et al. (2008) also present experiments with seven blind users and report a gain in efficiency of solving tasks via AISAs using *iSonic*.
- Chen et al. (2006) employed sonification techniques in their *AudioBrowser*, a barrier-free PDA. *AudioBrowser* was designed to be controllable by visually impaired users. “The user is guided in moving about the touch screen and making selections via speech and non-speech audio to indicate the

user's location on the touch screen." (Chen et al., 2006: 9). This multimodal navigation assists the user in finding the right "virtual segments" on the touch screen, that are connected with commands and actions. The authors also conducted a usability study in comparing the usage behavior of sighted and visually impaired users. They found that visually impaired users are able to operate *AudioBrowser* with roughly the same error rate as sighted users in terms of both, correct operations and correct answering of a questionnaire.

3.2.3.3. Haptic applications

Besides the auditory channel, the tactile channel is the main communication means in case of an inoperative visual channel. This means for the output side of technical applications, that information usually transmitted by graphical displays have to be transformed into a touchable structure. For the input side this amounts to applications that are controlable by gesture input. *Inter alia*, the following haptic applications have been developed in this context:

– *Haptic/audio-based exergaming for visually impaired individuals.* Exergamings are movement games where the user plays with full physical engagement, and not just with the handling of a joystick (Sinclair et al., 2007). This use of video games as part of an exercise serves both, fitness and fun – see, for instance, *Kinect*.[55] As they are basically video games, exergames rely heavily on visual displays and, hence, exclude visually impaired users. In order to facilitate exergaming also for partially sighted and blind exergamers, Morelli (2010) developed the tennis game *VITennis*. The name is derived from *visually impaired tennis* and is developed for the WiiMote controller. It gives blind users auditory feedback supporting their spatial orientation. Tactile feedback informs the user when to raise the arm to perform a hit. The project is still under development.
– *Haptic Rendering of Visual Data for the Visually Impaired.* Moustakas et al. (2007) devised an algorithm to convert urban 3D maps into force fields. Digitized three-dimensional video scenes are first converted into 3D models that are superimposed with a force field giving rise to a tangible "resistor texture". Users, especially visually impaired ones, can use the haptic signals of these 3D force field models in order to navigate through the maps and distinguish and identify streets, fences and facades.
– *Tactile Pens.* Tactile pens provide a combination of tactile graphics (cf. Section 3.2.2.5) and acoustic information. Tactile maps are equipped with contact points. If a tactile pen is pressed on such a contact point, the user can retrieve additional auditory messages. Tactile pens are used in various application areas. For instance, the *Tag It Guide*[56] provides a tactile-auditory

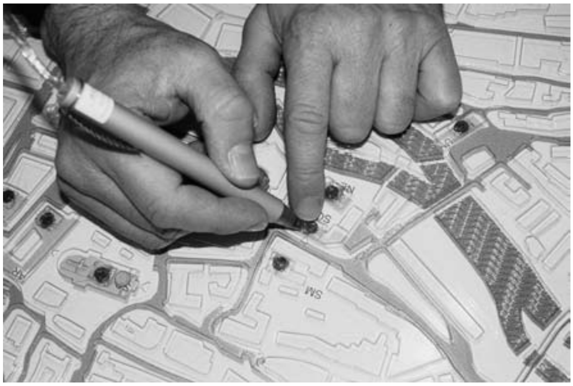


Figure 5. Snapshot of the *Tag-It Guide*, an audio tactile map of the Nürnberg city center (Germany) (http://www.dlinfo.de/content/tagitguide.php, March 17, 2011).

map of Nürnberg (Germany) – see Figure 5 for an illustration. The talking tactile pen offered by *Touch Graphics*[57] is not only a learning aid for the natural sciences, but also comes with an audio-tactile Sudoku game.

3.2.3.4. Barrier-free WWW for blind and visually impaired people

The development of the WWW helps blind and visually impaired people to achieve independence. Listening to the news, reading newspapers, booking a flight, or participating in online banking: all these applications are accessible for blind people, too – at least when freedom from barriers is accomplished. However, the *Aktion-Mensch* study (Cornelssen and Schmitz, 2008) shows that user-specific behavior reflects particular forms of disability: in contrast to blind users, visually impaired users tend to tolerate barriers (since they are not completely cut off from web-based communication by them). Blind users on the other hand organize themselves in associations and protest against insufficient accessibility. In addition, there also seems to be a difference in the primary goals of blind people and visually impaired ones for using the WWW: while the former mainly use it to acquire information, the latter use it to cultivate social contacts (Cornelssen and Schmitz, 2008).

According to Hellbusch and Probiesch (2011), barrier-free web rests on "7 pillars", which stand for the topics accessible web design has to take into account, namely 1) text orientation (to distinguish text from images, audio etc. by annotating multimedia files with additional text), 2) contrast and colors (e.g., of foreground text and background), 3) scalability of text, 4) linearization (readability of the text in an incremental order even without activated layout elements like CSS), 5) device-independence and dynamic content (navigation by keyboard and mouse), 6) understandability (different ways of presenting content for different users), and 7) structured contents (e.g., by usage of logical markup).

The most general advice for the developer, however, is to take the perspective of a blind person. Conformity to web standards is only half the battle. A misleading overall structure or texts that are hard to understand can hinder not only blind users. The developer should always be up-to-date by taking the newest technological developments into account. Optimally, the developer should stay in contact with the actual target groups of users in order to be able to be acquainted with their actual needs.

3.3. Barrier-free communication for the deaf and hard of hearing

3.3.1. Problem description

According to the *World Federation of the Deaf* (WFD) there are about 70 million deaf people worldwide.[58] With regard to technical communication, three groups of hearing-impaired people are to be distinguished, since communicative abilities can be limited on various levels (cognitive, social, emotional) depending on the severity and/or duration of the disease (cf. Linnartz, 2003), namely:

- the deaf, that are people that either were born deaf or have become deaf long before the first language acquisition process ended;
- the hard of hearing, that are people with a reduced hearing ability, that are nonetheless able to recognize speech and other sounds, perhaps with the help of a hearing aid;
- the late-deafened, this refers to people that acquired their being hard of hearing or being deaf after the age of 3 years.

Modern technologies can largely enhance the deaf people's independence and quality of life. The WFD developed eleven guiding principles in support of accessibility and technological aids for deaf people.[59] The guidelines capture the full range of tasks needed "in order to enable [a deaf person's] full participation in society" (principle No. 9 of the WFD guidelines), from technological design principles to the support of initiatives concerned with topics related to barrier-free technical communication for the hearing-impaired.

Worth mentioning from the point of view of speech technology are the 2306 devices (as of December 19, 2011) listed within class No. 22 of the Assistive Device Groups *Assistive products for communication and information* (ISO 9999:2011), to be found at http://www.rehadat.de/rehadat/Reha.KHS. In addition to these tools, there is still need for further web-based accessibility technology, since when retrieving written texts, hearing-impaired people, in particular the deaf, face the problems of too complex texts, of audio files without subtitles, of incorrect automatic translations of texts, or of the lack of anonymity. Especially the latter problem affects chat communication, since a deaf person needs to use a webcam in order to communicate via lip reading or sign language. Virtual avatars (see Section 3.3.3) can help to overcome this limitation. The avatars receive sign language input from the user and reproduce it to the communication partner in the chat so that the user remains anonymous.

To offer speech-technological aids for deaf people means to take the perspective of this target group, in particular, to accept sign language as the target group's language system. Accordingly, a technological aid specific to the needs of the hearing impaired is a sign language screen reader (Ghoul and Jemni, 2009). A sign language application can, at least to a certain degree, draw on devices developed for spoken language. The reason is that first language acquisition of sign language exhibits the same mechanisms as that of spoken language (Leuninger, 2000). For example, sign languages also have dialects and varieties (Mally, 1993), i.e., there is no single universal, but rather a couple of (more or less) international sign languages. Moreover, gestures and speech are processed partly in the same regions of the brain (Petitto et al., 2000).

In this section, we will focus on two aspects of technological applications for barrier-free communication of the deaf and the hard of hearing. The mother tongue of users of this group usually is a sign language. A minimal sketch of sign languages is given in Section 3.3.2. Afterwards, we go over applications in the area of dynamic, interactive displays and web content in Sections 3.3.3 and 3.3.4. Additionally, we will look at cochlear implants (Section 3.3.4.6).

3.3.2. A very brief introduction into sign language

There is a general distinction between the *standard sign languages* and the so called *manually coded languages* (MCLs). MCLs do not represent independent language systems, they rather fully adapt the grammar of the underlying verbal language. Standard sign languages, in contrast, are full-fledged, natural, complex, visually retrieved languages. Hands, head, face and body are used as articulators. Following Valli and Lucas (2000), sign languages are to be described on the four standard linguistic levels:

Phonology Sign languages have a phonological structure in the sense that the elements of a gesture are organized into smallest contrastive features (Stokoe, 1960). Each sign of a sign language can be represented by four "phonological" feature values, namely hand-shape, the orientation of the hand, location of the gesture within gesture space, and movement (Liddell, 1984; Liddell and Johnson, 1989). Auto-segmental features are represented on different levels independent of each other. Movements form the nuclei or the core part of a gesture syllable.

Figure 6. The term "barrierefrei" (the German word for *barrier-free* written in fingerspelling. The letter-handshape correspondences are drawn according to http://www.taubblindenwerk.de/ fingeralphabet.html.

Morphology Sign language gestures are made up of "morphemes", that is, of smallest meaningful units. Morphological composition of gestures accounts for the morphological processes of inflection, derivation and compound formation. A sign language verb can agree with its subject or its object, which is generally called subject-object agreement. Agreement is also encoded by means of morphemes. Verbs can be modified by means of adverbs or classifiers (where the latter is a certain kind of predicate which denotes a *class* of objects). Perhaps the most intriguing property of sign language morphology is its utilization of space and the location of gestures within gesture space in a meaningful way (Emmorey, 1999). Most gestures are monosyllabic, including the polymorphic ones.

Syntax The sentence structure of a sign language is spanned as a (perhaps very complex and nested) predicate-object pattern in a three-dimensional gesture space using pronominal references to referents placed within the space. Sentence mode is expressed by means of syntactic facial expressions or prosodic cues.

Semantics Like any natural language, a sign language has a set of lexicalized signs with lexicalized meaning (i.e., a lexicon), and some syntactic means to build larder meaning units out of the lexical ones.

In addition to the autonomous language levels mentioned above, a sign language has fingerspelling at its disposal. Fingerspelling is a translation of the written (Latin) alphabet – see Padden and Gonsalius (2003) for a historical review on fingerspelling. Signers use the fingerspelling alphabet in order to communicate about things for which there is no ready means in their sign language, for instance, for spelling foreign or proper names, or for signing abbreviations. Figure 6 shows the term "barrierefrei" *barrier-free* written in fingerspelling.

3.3.3. Methods

The methods presented in this section sketch the historical development of interactive (web) applications for deaf or hearing-impaired people. These applications aim to fulfill the following tasks:

– translation from spoken language into sign language gestures (speech-to-sign language),
– translation of written texts into sign language (text-to-sign language),
– translation from speech to written text (speech-to-text, for example, by means of speech recognition, *Computer Assisted Notetaking* (CAN)[60] or *Communication Access*, or *Computer Aided Real-Time Translation* (CART)[61]; see also section 3.3.4.1),
– interactive editing of sign language texts (e.g., in Wikis),
– providing subtitles for online audios and videos.

In the following Sections 3.3.3.1 and 3.3.3.2 it is described how these tasks are accomplished by means of motion capturing techniques and sign language notation, respectively.

3.3.3.1. Motion capturing

Motion capturing is a method to generate three-dimensional computational models that record human movements. Although not primarily developed for applications for the hearing-impaired, motion capturing techniques can also be used for recording and producing gestures of sign languages. In this line of usage, motion capturing was applied within the *Virtual Signing, Capture, Animation, Storage and Transmission* project (ViSi-CAST, 2000–2003; Bangham et al., 2000; Kennaway, 2002). ViSiCAST integrates avatars into browser plug-ins of web applications in order to translate texts into sign language. Future application scenarios include face-to-face transactions (e.g., communication with officials), Web2.0 (e.g., chat communication), and TV (e.g., watching films in sign language) (Schulmeister, 2001). However, nowadays most speech technology projects avoid motion capturing since it is a very labor-intensive and expensive approach.

3.3.3.2. Sign language notation

A standard notation system for sign languages does not exist so far. However, there is a range of transcription and annotation systems. Löffler (2004: 16–27) distinguishes the following three basic approaches to the notation of sign languages (examples are added by the authors):

1. *Verbal-Language-Oriented Sign Writing* as a mediator between verbal and sign languages (for example, finger alphabets),
2. *Scientific Transcription Systems* (e.g., *HamNoSys*, the *Hamburger Notationssystem* for sign languages[62] (Prillwitz et al., 1989), and
3. *Full-Fledged Writing Systems* for the everyday use (for example, *Sutton Sign Writing*[63]).

The above-listed notation systems are designed to be human readable. There is a wide range of annotation frameworks to retrieve sign languages automatically. In the following, we firstly focus on transcription systems that differ greatly in the form and complexity of their descriptive apparatus. Secondly, markup languages used for automatic retrieval of sign language are presented.

Transcription Systems

Notational systems for transcribing sign languages are distinguished according to their area of use into transcription systems for the everyday use and transcription systems for scientific and documentary purposes.

1. *The everyday use*: sign language writing by means of *SignWriting*. SignWriting is a component of the *International Movement Writing Alphabet (IMWA)*[64], initiated by Valerie Sutton in 1975. SignWriting uses the *International SignWriting Alphabet* (ISWA) that allows one to represent movements and mimics in terms of highly stylized hand shapes. SignWriting, thus, transcribes sign language by means of symbols, that are abstract representations of gestures (da Rocha Costa and Pereira Dimuro, 2001).
   The basic unit of SignWriting is called *sign-box*. A sign-box is a combination of graphical and schematic symbols that represent head, hand and body shapes or movements to express mimics (Papadogiorgaki et al., 2005). SignWriting version SSS-2004 contains about 25.372 symbols, which are divided into 8 categories, 10 groups, 50 elements, 5 variations, 6 fillings and 16 rotations (Aznar et al., 2006: 1). With this descriptive inventory, SignWriting allows to transcribe 40 of the international sign languages (cf. www.signwriting.org).
   At the beginning, sign written texts had to be created manually. *Shorthand*[65], SignWriting stenography, facilitates a quite rapid process of transcription (as appropriate, e.g., for dictations). Today, sign written texts can be used by print media as well as in digital communication. There is a wide range of technical means in support of sign written communication, for example, the international online dictionary *SignPuddle*.[66] The dictionary can be searched for words, lists of gestures or symbols. As an example, Figure 7 shows the technical gesture for the search term "Linguistik" (*linguistics*) in the dictionary of German sign language. Search terms are displayed in sign writing and the written form of the particular verbal language. The *SignWriting* character set according to ISO 15924:2004 (No. 095, Code *Sgnw*[67]) is available under the SIL open font license.[68] For more information see *Sutton's SignBankSite*[69].
2. *Scientific Application*: the *Hamburger Notationssystem* (HamNoSys). Based on *StokoeNotation* (Stokoe et al., 1965), is a transcription system

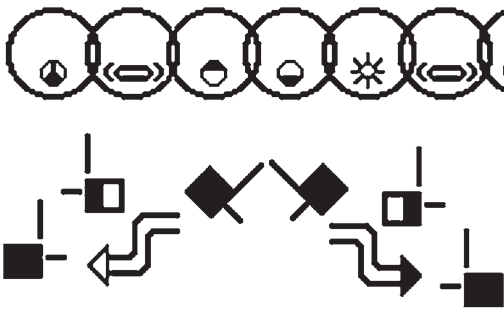


Figure 7. The concept of /linguistics/ as a subject.

from American Sign Language into verbal language. HamNoSys provides a "phonetic" notation system, which allows to manually transcribe gestures (Prillwitz et al., 1989). It is freely available since 1989 and it is used primarily in academic contexts.

HamNoSys uses an inventory of about 200 symbols. Most of them can be uniquely assigned to a single class of hand shapes. Where applicable, the form of a symbol is iconically motivated, that is, to resemble the corresponding gesture.[70]

HamNoSys signs encode one- or two-handed gestures, head position, body posture, and gaze as well as different locations, movements and enacted shapes. Basic signs can be combined to capture more complex gestural structures, so that in principle all possible fine grained phonetic differences can be described. Examples of some basic hand shapes and their representation in HamNoSys are shown in Figure 8.

HamNoSys notations are used in theoretical and applied research. They represent a basis for automatic retrieval of sign language by means of *Signing Gesture Markup Language* (SiGML – see the subsequent section).

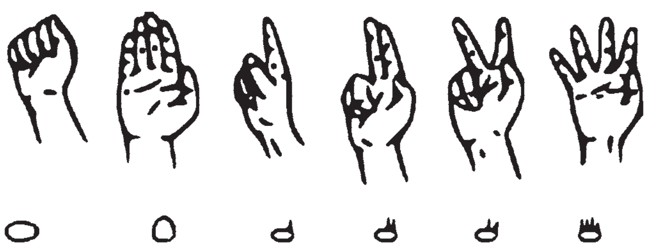


Figure 8. Six basic hand shapes (upper row) and their HamNoSys transcription (taken from www.sign-lang.uni-hamburg.de).

Markup Languages

Text technological applications for sign languages as well as gesture recognition and production systems presuppose some machine readable representation of gestures. To this end, two markup languages have been developed.

- *SignWriting Markup Language (SWML)*. SignWriting Markup Language (SWML) (da Rocha Costa Antônio Carlos and Pereira Dimuro, 2002) is an XML-based markup language based on SignWriting developed in 2001 in order to account for archiving, exchange and retrieval of sign written texts and dictionaries. SWML accounts for interoperability between SignWriting and compatible applications. Furthermore, SWML allows a deaf person to access the web using their native language. Based on SWML three application were developed: SWML-D for the Dutch *WoordenBoek* online dictionaries[71], SWML-S for SignPuddle online dictionaries (see above) and SWDB for UCPel Linguistic SignWriting Database (da Rocha Costa and Pereira Dimuro, 2001).[72] For further reading on SWML see Papadogiorgaki et al. (2004).
- *Signing Gesture Markup Language (SiGML)*. SiGML is an XML markup language derived from the HamNoSys notation. SiGML was developed within the ViSiCAST project (see Section 3.3.3 above). In ViSiCAST, SiGML served as an interface between generation of content and of avatars (Glauert, 2002).
  SiGML provides elements to annotate manual and non-manual gestures. These annotations are implemented as sequences of so-called "signing units". A "signing unit is an explicit gestural definition for a single sign, but it may also be a direct definition of avatar animation parameters, or an indirect reference to another SiGML document." (Elliott et al., 2004: 98 f.). An additional characteristic of SiGML is its extensibility to further standards like SMIL and XMT (MPEG-4) (Elliott et al., 2004).

3.3.4. Applications and directions of research

Technical applications that aid a barrier-free communication for the hearing-impaired involve a visual display of sign languages. There are two kinds of applications: on the one hand, recordings of signed texts by a human signer make the text accessible for deaf users. On the other hand, synthetic sign language gestures produced by an avatar can be employed in a range of technical interfaces. Both kinds of applications are introduced in more detail in the following Sections 3.3.4.1 and 3.3.4.2.

Figure 9. The website of the libraries of the city of Hamburg (Germany) offer sign language videos for information and navigation (http://www.buecherhallen.de/dgs/beispiel.html).

3.3.4.1. Video mediated communicative aids – natural sign language videos

Natural sign language video recordings are a communicative aid in various contexts of application. For instance, public websites can make their content accessible for the deaf by means of recordings of signers – see Figure 9 for an example of sign language videos used as part of the online service of libraries. However, utilizing sign language videos still poses technological challenges in the area of the WWW. The reasons are, according to Kennaway et al. (2007), high production and data storage costs, large data size of the videos in connection with possibly low data rate internet connections, and the need for a server architecture that allows to transmit videos via streaming.

Natural sign language videos are nevertheless preferred for communication for personal and technical reasons (Kennaway et al., 2007: 4). In the following, we will describe some current applications. These research projects are still under development, and thus, not yet widely applied. Applications can be partitioned according to whether they rely on sign language recordings or allow for video mediated face-to-face communication.

Sign language recordings

Technologically aided mutual translations between recordings of signed texts and written texts facilitate access to text-based services also for signers. The following systems implement slightly different approaches to this kind of augmentative communication:

- *Communication Access/Aided Real-Time Translation* (CART) is a tool that supports the translation of gestures into notations in the role of subtitles. The subtitles are transmitted onto displays in front of deaf or hard of hearing users.[73] Application areas of CART are for instance: conversations, school lessons, live shows such as sport events, parliament debates, conferences or services.[74] The subtitles can be installed for single users or for a whole group of users. A human interpreter enters the subtitles into a shorthand machine using shorthand notation (see above). The machine translates shorthand into subtitles in real time. This method enables a fast information transmission and allows users to follow the presentation without delay. The main problem of this method is the time pressure of shorthand input and the dependency of the recipient from the interpreter. The time pressure forces the interpreter to quickly decide, select and structure the incoming information, which ultimately leads to a shortening of the information (Wagner, 2005).
- The *Signing Question and Answer Tool (SQ?!AT)* enables deaf users to use web services by means of requests formulated in sign language. On the user side, SQ?!AT requires a webcam, on the web service side, cooperation with a SQ?!AT provider is required. SQ?!AT works as follows: users record their gestures with their webcams and send the recordings directly to a translation service offered by the provider (in the case of the German SQ?!AT implementation this is the *Gebärdenwerk* ("gesture factory")[75]). The gesture video is translated into text and sent to the requested web service. The answer of the web service is translated into a sign language video and sent back to the user – see http://www.dgs-filme.de/Flashplayer/sqat.htm for a signed introduction into SQ?!AT.

Face-to-face communication

Technological transmission means of realtime video data, as used for video telephony, can also be employed for signed communication. Here, two kinds of systems have to be distinguished: on the one hand, there are systems that just provide a video telephony platform, like UbiDuo. On the other hand, there are applications that are equipped with built-in speech-to-sign language and sign language-to-speech translators, which therefore allow for the interaction of signers and non-signers. An example for the latter approach is PLMT. Both example applications are described in more detail hereinafter.

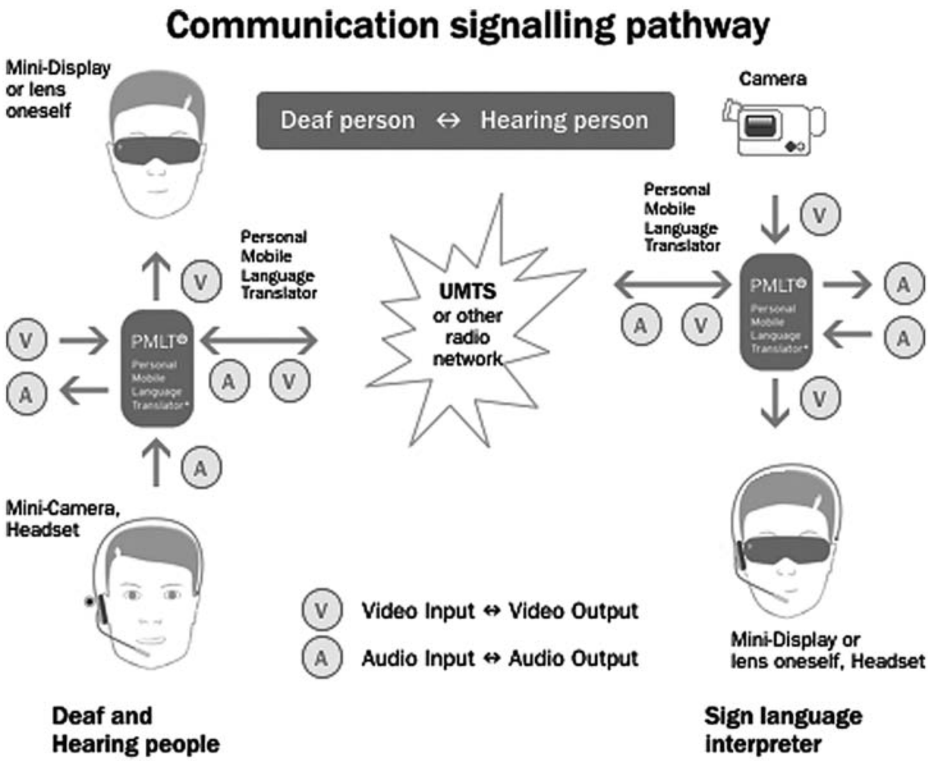


Figure 10. Graphical depiction of the architecture of PMLT (taken from http://www.global-communication-premium.de/loesung.en.html, March 17, 2011).

– *UbiDuo.* The UbiDuo is a twin notebook that is widely used in the USA. This notebook is equipped with wireless internet access. Its battery operation makes it portable. The double screen and two keyboards enable a simultaneous face-to-face interaction.[76] The two devices are connected and communication exchange takes place via chat.
  UbiDuo is supported by the National Institute of Health in the US and is used in various official institutions such as hospitals, schools, or governmental organizations.
– *PMLT (Personal Mobile Language Translator).* The PMLT is the hardware component of a mobile communicative aid. A special characteristic of PMLT is that it facilitates face-to-face communication between signers and non-signers (typically this means deaf and hearing people). The mobile communication process, including speech-to-sign language and sign language-to-speech translation in a signer/non-signer interaction setting, proceeds in the following steps (see Figure 10 for a graphical scheme):[77]

1. contacting the translation center, the so called *relay service* (RS), via SMS from the PMLT of the deaf person;
2. sign language is transmitted via video streaming to the RS by means of a mini camera placed on the headset of the hearing person;
3. translation of sign language into speech and transmission (audio streaming) from the RS to the headset of the hearing person;
4. transmission of speech (audio-streaming) from the headset of the hearing person to the RS;
5. translation of speech into sign language and transmission (video streaming) from the RS to the video eyewear of the deaf person.

Optional add-on functionality allows one to guide a blind person via his mini camera.

3.3.4.2. Video mediated communication – virtual avatars

Regarding the needs of individual users (e.g., anonymity) and the problems with respect to natural sign language videos (e.g, transmission rate), virtual avatars provide a viable alternative. According to Kennaway et al. (2007: 3), signing avatars have the following advantages:

– Gestures can be entered directly, making the complex procedure of motion capturing superfluous.
– Continuity is secured, since any gesture can be presented to any of the avatars.
– The influence of dialects of a sign language becomes irrelevant, since the spelling of the content of a sign language video can easily be edited without a complete or partial re-recording.
– The speed of the internet connection and the size of the hard disk do not play a crucial role: animations can be rendered by means of the software installed on the client computer, whereas only the annotations have to be transmitted.
– Data transmission is automatically adapted to the client's hardware environment. The user can shift the perspective on the avatar, and change the avatars if desired.

From the second half of the 20th century on, many research projects have been dealing with artificial animations of sign language – see the overview given by Kennaway et al. (2007). However, a particular short-coming of these avatars is that their rather inexpressive mimics and body movements give rise to misinterpretations of gestures. To this end, the eSIGN avatar *Virtual Guido* (*VGuido*) has been endowed with highly animated mimic facilities (Elliott et al., 2008). This section, however, deals with the generation of gestures proper. We distinguish between dictionary-based generations of animation on the one hand (Section 3.3.4.3) and applications of markup languages on the other hand (Sections 3.3.4.4 and 3.3.4.5).

3.3.4.3. Dictionary-based applications

There are two prevalent dictionary-based sign language animators: the one, *DictaSign*, has been developed in the context of Web 2.0; the other, *Signing Science Dictionary*, has been devised for the classroom.

– *DictaSign.*[78] DictaSign lets signing users take part in the Web 2.0, allowing them, amongst others, to create anonymous sign language video news in forums and to collaboratively edit videos in Wikis. Efthimiou et al. (2009: 80) describe the workflow of DictaSign as follows: "Users make their contributions via webcams. These are recognized by the sign language recognition component and converted into a linguistically informed internal representation, which is used to animate the contribution with an avatar, and to translate it into the other respective three sign languages." Additionally, sign-language-to-sign-language translators (such as "Google Translator") are supported.
– *Signing Science Dictionary.*[79] The Signing Science Dictionary (SSD) provides an interactive three-dimensional sign language dictionary to be used for educational purposes in the classroom. It therefore serves not only as a resource for teachers and pupils, but also for parents. The virtual human technology on which SSD's gesture animation is based, is delivered by *Vcom3D*.[80] Figure 11 presents a screenshot of an Vcom3D/SSD avatar in action.

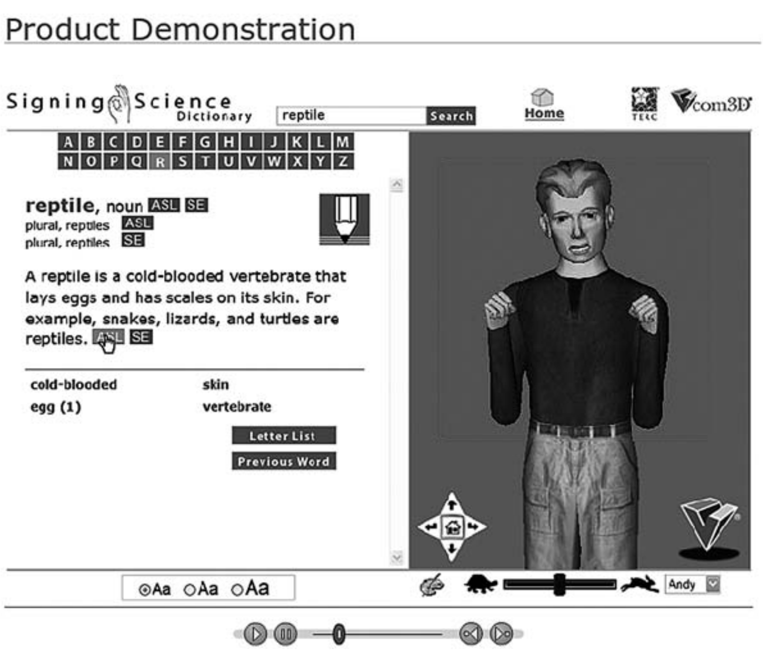


Figure 11. Screenshot of a *Vcom3D* avatar enacting the "reptile" gesture from the *Signing Science Dictionary.*

3.3.4.4. *Applications of HamNoSys*

HamNoSys was developed as a notation system for generating linguistic resources (Hanke, 2001: 195–201) – see also Section 3.3.3.2 above. Applications based on HamNoSys are, according to Hanke (2001), lexica and transcription databases such as iLex (Hanke and Storz, 2008), a combined lexicon and transcription database. Further applications of HamNoSys include specialized sign lexica (for example, terminological sign lexica for carpenters).[81] Recent applications include the coverage of further sign languages like the Estonian (Paabo et al., 2009) and the Greek (Karpouzis et al., 2007) ones.

3.3.4.5. *Applications of SignWriting*

Wöhrmann (2005) adapted the SignWriting system (see Section 3.3.3.2) to the German sign language and successfully applied it in teaching. However, in everyday life in Germany, SignWriting is still used rarely. One exception is the adult education center of Kaarst-Korschenbroich,[82] whose website allows users to select between written and signed content. The user can select between a HamNoSys and a SignWriting sign language notation. Figure 12 shows the menu written in SignWriting.

Ahmed and Seong (2006) worked on the integration of SignWriting into mobile phones. Their software converts sign language input into written text, allowing, for instance, a hearing recipient to read the message. The authors evaluated their system in terms of its acceptance rate: 88.80 % of the users gave positive feedback. For future work, Ahmed and Seong (2006) identified the need to optimize the keyboard, and to account for an automatic translation of sign language videos (recorded via a mobile phone camera) into text.

Figure 12. Website of the adult education center of Kaarst-Korschenbroich: German menu (left) and the same menu written in SignWriting (right).

As already mentioned above in Section 3.2.3.3, Moustakas et al. (2007) use the SignWriting Markup Language (SWML) (see Section 3.3.3.2) as part of their haptic force field modeling of visual data.

3.3.4.6. Cochlear implants

The cochlear is the part of the inner ear that is responsible for the conversion of sound waves into nerve impulses, which are processed further via the auditory nerve (*Nervus cochlearis*). Deaf people with an unimpaired auditory nerve can be provided with a “hearing prosthesis” known as *cochlear implant* (or the cochlear implant system, hereafter abbreviated as “CI”). The CI is an electronic aid partially implanted beneath the skull and directly connected to the cochlear. An external microphone with speech processor converts the incoming sound into electromagnetic signals which are passed on to the *Nervus cochlearis* – see Figure 13 for an illustration. This way, CI makes speech and sound perceptible for people with an impaired hearing.

Cochlear implants are, however, highly debated within the hearing impaired community. Besides medical reservations, many deaf people argue linguistically and ethically that a sign language, being a full-fledged language, is the

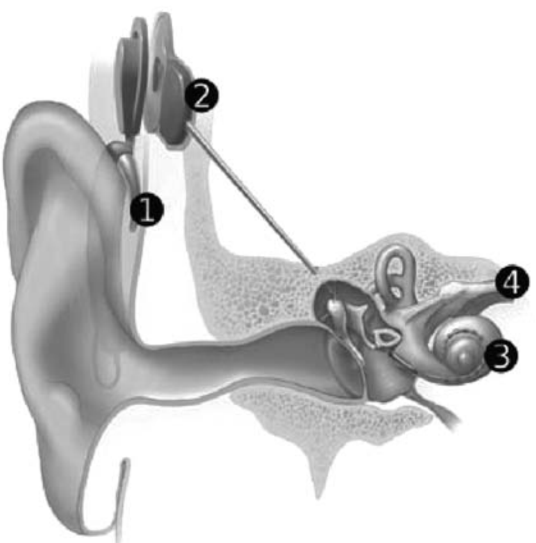


Figure 13. Cochlear implant: (1) is the microphone with sound processor, (2) is the transmitter coil, (3) is the cochlea, and (4) is the *Nervus cochlearis* (*National Institute of Health* (http://www.nih.gov/), public domain; also http://commons.wikimedia.org/wiki/File:Cochlear_implant.jpg, slightly modified by the authors).

natural mother tongue of the deaf. CI is no cure, but rather a prosthesis (on this and related arguments see, for instance, the 2006 statement on CI by the Austrian association of the hearing impaired.[83])

In fact, CIs do not bring about a full reconstruction of hearing abilities. CIs produce sound effects that can be interpreted differently by CI users. Previous experiences and competence in verbal language have a large impact on the ability to recognize language and to distinguish it from noise. Loud background noise often makes the identification of speech impossible (Carroll and Zeng, 2007; Di Nardo et al., 2010).

There is a scientific interest in CIs especially from the areas of linguistics and medical sciences. On the one hand, scientists are working on the technical optimization of the implants and the understanding of their functionality (for example Lai and Yeh, 2009; Eilers et al., 2009; McDonnell et al., 2010). On the other hand, there is an interest in studying the effects on language acquisition after implantation (for example Adi-Bensaid and Most, 2009; Hayes, 2010; Strelnikov et al., 2009).

3.4. Barrier-free communication for deaf-blind people

The causes for deaf-blindness are manifold (Atwood et al., 1992).[84] The *Autosomal Recessive Hereditary Transmitted Disorder*, the so called "Usher Syndrome" (Mets et al., 2000), is of special interest since it is the most frequent source of inherited deaf-blindness. Further causes for deaf-blindness are, for example, Meningitis or Toxoplasmosis (Atwood et al., 1992; Alzheimer, 1970).

3.4.1. Problem description

In most cases, the first symptom of deaf-blindness is a beginning hearing impairment followed by a gradually progressing loss of visual capabilities. Since in case of deaf-blindness the two main communication channels, viz. the visual and the auditory one, are not at disposal, communication takes place via the tactile channel. In general, the typical communication means of a deaf-blind person are:[85]

– the Lormen alphabet (Section 3.4.2.1),
– reading from PC via Braille display,
– the *DeafBlind Communicator* (DBC), that is, a mobile device for exchanging information by means of SMS and Braille display[86],
– finger Braille,
– tactile gestures, for instance, sign language and finger alphabets, that can be read out via touching (*Hands-On Signing*[87]),
– in case of remaining visual abilities, also signing within the deaf-blind person's remaining field of vision comes into question (*Visual Frame Signing*).

In principle, deaf-blind people can make use of Braille Displays or hearing devices like cochlear implants. If gestural communication is limited due to visual impairment, there are alternatives in terms of tactile alphabets (Cooper, 2006). These are described in the following Section 3.4.2.

3.4.2. Applications and directions of research

There are two tactile alphabets, namely Braille (cf. Section 3.2.2.4 above) and Lormen (Section 3.4.2.1). Accordingly, technical applications can be distinguished according to whether they draw on Braille (see Section 3.4.2.2 below) or on Lormen (see Section 3.4.2.1 below).

3.4.2.1. Lormen alphabet

The Lormen alphabet is typed directly into the palm of the deaf-blind person's hand. As an example, Figure 14 shows a Lormen transcription of the English word "barrier-free". Lormen is a translation of verbal language, hence, it is only suitable for deaf-blind people that have acquired verbal language as their first language. A technical application of the Lormen alphabet is applied as patent (Rupp, 2003), however, there seems to be no working tool yet.

Figure 14. The word "barrier-free" written in terms of the Lormen alphabet.

3.4.2.2. Tabli

Tabli[88] is an extended communication module for the *Braille Wave*[89] and *Braille Star 40* [90] devices. The hearing and visually impaired person can enter a text via keyboard. This text appears directly on a screen. The sighted communication partner replies via keyboard, and the answer is received on a Braille display. Hence, Tabli enables face-to-face interaction in real time.

3.5. Barrier-free communication for people with physical impairments

3.5.1. Problem description

Hearing and seeing impairments described in previous sections limit primarily the perception of communicative signals. Physical impairments limit the person in their ability to producing communicative signs. The oral motor functions can be affected by neurological illnesses (e.g., the Guillain-Barré syndrome). Diseases like Parkinson's or age-related tremor hinder a precise control of mouse and keyboard. Paralysis can lead to a disfunctionality of limbs to such an extent that only one operational finger may remain for operating the communicative aids.

3.5.2. Applications and directions of research

Given the wide range of disease patterns of physical impairments, the respective technical applications for augmentative communication pertain to all communication channels. We distinguish the following list of technical aids:

- *Speech aids* (i.e., voice generator, voice amplifier). These devices help people with, for example, laryngeal diseases (laryngectomy, tracheotomy) or dysarthria. Speech aids are hardware components that are placed directly onto the patient's neck. The aids compensate for weak intonation and weak voice capacity.
- *Writing and Drawing aids.* Besides the general problem of operating computer devices (tremor hinders coordination, weaknesses of muscles make mouse clicking difficult, etc.), in particular the complex process of writing texts with a keyboard poses the main challenge. Individually programmable on-screen keyboards with scan functionality, options to hide unused features, speech input systems, and automatically adaptive or predictive text completion are approaches to alleviate these problems. There are text processing tools (like MOMO SK[91]) that unify the above functionalities within a single framework, making them applicable for personal as well as educational use.
- *Aids for nearby communication* (e.g., character sets, symbol sets, communicators, software for nearby communication). Applications for improving nearby communication are communicators with a speech synthesis functionality.

  Communicators mostly produce speech synthesis output or natural recordings. Control devices of communicators are plates or symbol plates. Each image on a plate corresponds to a speech message. Depending on the device, the user activates a message pressing a symbol, using an eye controller, a head mouse (or other mouse variants), digital joysticks or a scan functionality.
- *Alternative mouse controllers*, like the *Hands free interface* for human-computer interaction (Chathuranga et al., 2010), are hardware and software-based implementations. On the software side, aids for mouse key activation (see above) are used. The user only needs to scroll the cursor over the screen and to rest on the selected item – the selected function is executed after a predefined time has passed. Depending on which device substitutes the mouse-controlling hand, mouse controllers are divided into the following groups:

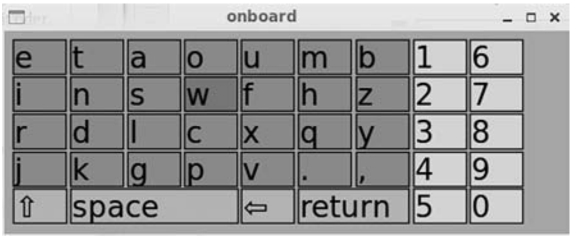


Figure 15. Letter “w” is activated (indicated by a shaded cell) on the on-screen keyboard *on-board* of the Gnome desktop.

  - *Head and/or mimics controllers* (mimics recognition) represent a combination of software and webcam (Mohamed and Koggalage, 2007). Three kinds of control can be distinguished: (1) *Controlling via a contact point:* the user attaches a small reflecting tracking marker on his nose or glasses. The camera tracks the marker and enables the user to operate with the mouse by means of moving and holding the marker. (2) *Controlling via head movements:* the mouse is operated through tracked head movements. No tracking marker is needed. (3) *Controlling via mimics:* this controlling variant is especially suited for applications using a scan mode. Facial motions such as mouth or eyelid movements are assigned to signals that actuate the cursor.[92]
  - *Eye control.* Applications using eye controlling analyze facial features, for example, the distance between the eyes or the size of the pupil (Hori et al., 2004; Ali et al., 2007). First, the program has to learn the user’s typical features which are required for calibrating the application. The cursor is then controlled via eye movements. Mouse functionality is steered by evaluating eyelid movements and resting times of the eyes.
- *On-screen keyboards* are accessed via mouse or mouse-like devices. Mouse-like devices utilize the scan functionality of an on-screen keyboard. In scan mode, the single mnemonics are colored according to rows and columns. If the scanner is in the correct row, the user can confirm the selection via mouse click (in the on-screen keyboard displayed in Figure 15, letter “w” is selected). The scanner goes through the row highlighting each of the characters. If the target character is active, it can be selected by mouse clicking and, for example, inserted into a text. When a software for decoding mouse clicks is used, a mouse click is performed evaluating the user’s eye resting time on a key.

  Various on-screen keyboards differ in their additional functionality. There are, for instance, customizable keyboard layouts, word completions, magnification of characters, acoustic feedback, or speech output.

Figure 16. Dasher: an on-screen keyboard with word prediction.

A special variant of an on-screen keyboard is *Dasher*. Dasher presents a vertically aligned alphabet on the screen. A user can zoom in and select letters via a mouse (or any other compatible pointing devices). Each selection activates suggestions of the most likely letters, letter combinations and texts, which, if selected, are added to the editor at the top of the window. Zooming and selecting this way, texts can be constructed in a comparatively convenient way. The most likely pieces of text with respect to the previous selection appear larger so that they can be selected more easily. A snapshot of a Dasher session is given in Figure 16.

The software is adaptive and can memorize the formulations preferred by the user. Dasher can be actuated by hand, head mouse, eye-tracker, or even neural signals (Felton et al., 2007) and is available under the GPL license for Linux, MacOS X, Pocket PCs and Windows systems.

– *Voice controlling.* Technical aids based on voice controlling range from supporting the writing of texts to the control of the whole computer system. *Via Voice*[93] and *Dragon Naturally Speaking / RealSpeak V2* [94] are, for example, multifunctional tools for text editing and PC controlling. Additional modules allow to access the mouse and to command the web browser. Dictionaries and macro commands facilitate the process of text editing.

3.5.3. Barrier-free web access for people with physical impairments

Technical aids that facilitate the usability of desktops does not simultaneously guarantee a barrier-free access to websites. Users that are bound to the use of keyboards are especially affected. So called *add-ons* and functionalities of particular browsers can be used for augmenting their operability (though not for all user groups). Improving the accessibility of the web is accomplished from two sides: from the web itself (Section 3.5.3.1), and from the browser (Section 3.5.3.2).

3.5.3.1. Techniques in Web Development

The requirements for accessibility on websites differ slightly, depending on country specific guidelines. Requirements for usability of keyboards are listed in WCAG 2.0 within the guideline of usability 2.1, called *Make all functionality available with a keyboard*. How to meet the success criteria for the guideline 2.1 is described at http://www.w3.org/WAI/WCAG20/quickref.

3.5.3.2. Browser extensions

There are accessibility add-ons for the most popular free browsers, viz. Firefox and Opera.

1. The browser *Opera* allows for a completely mouseless access via keyboard by means of built-in short cuts.[95]
2. For the *Firefox* browser, the add-on *mouseless-browsing*[96] offers an extension that enables numerical indexing of all links required for navigation (including tab keys) – see Figure 17. Typing a number on the keyboard activates the corresponding link. This functionality enables a fast access to the desired links. The layout of the links is selected by the web developer. Various options allow the user to customize *mouseless-browsing* for his individual use.

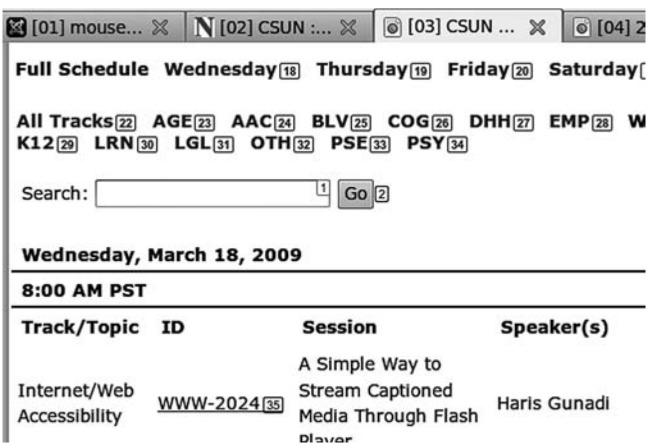


Figure 17. Indexing of links accomplished by the Firefox add-on *mouseless browsing*.

3.6. Neurally controlled devices

Probably the biggest challenge for barrier-free communication are users suffering from the locked-in syndrome, that is, people who are conscious but physically fully paralyzed. Obviously, assisting devices that have to be controlled by hands, voice or even eyes will not work out for this group of users. Instead of interfaces that rely on some sort of muscular activity, so called *brain-computer interfaces* (BCI; Dornhege et al., 2007) have recently proven useful in this case (Kübler and Neumann, 2005; Birbaumer et al., 2008; Daly and Wolpaw, 2008). The idea behind BCI is to use neural activity in order to control some assisting device, for instance, an robotic arm (Hochberg et al., 2012).

A first distinction to draw here is between *invasive* and *non-invasive* BCI applications. Invasive applications are implanted into the grey matter of the brain. An example for invasive BCIs is the “neuroprosthetic” device known as *cochlear implant* (see Section 3.3.4.6). Noninvasive BCI makes use of neural properties that can be measured without surgical intrusion, typically brain waves that can be recorded by means of an *electroencaphalogram* (EEG). The neural control of Dasher described in Section 3.5.2 is an example in this regard.

While the electrodes of an EEG detect electrical signals on the scalp at a rather broad resolution, higher resolution can be achieved when the electrodes are placed directly on the cerebral cortex. This medical imaging variant of EEG is known as *electrocorticography* (ECoG). ECoG has been used as an interface technique for controlling a visual keyboard (Krusienski and Shih, 2011). ECoG is a *semi-invasive* method, since it does not involve a surgical treatment if the cortex, but still requires an opening of the cranium.

Not only the medical procedure underlying in particular invasive BCI calls for an ethical consideration of neurally controlled devices. Other unresolved issues include, for example privacy, the notion of personhood, and responsibility for eventual device-caused errors (Haselager et al., 2009) (Tamburrini, 2009). Discussing such issues and striving for a consent for them has given rise to a new field of ethics, called *neuroethics* (cf. Illes, 2006).

4. Conclusion

The practical needs of handicapped individuals require an integration of assistive aids and state-of-the-art applications. This is an ongoing process, especially in the area of the WWW (see Section 2.1 above). The formulation of guidelines and specifications for barrier-free web-sites and web applications (WAI-ARIA) hardly keep up with the rapid development of new technologies. Thus, the integration of novel tools and assistive technologies is a continuous challenge for the developers.

This article is meant to support developers and readers of an inter-disciplinary community in the understanding of barrier-free communication, the needs of the target groups and the technical aids available today to meet these needs. In addition to the overview given in this article, see the following publications for further reading on selected topics:

– *Audio/video for elderly people* (Mylonakis et al., 2008),
– *Speech processing of the deaf and hard-of-hearing people* (Kasturi, 2006; Pa et al., 2008; Fajardo et al., 2004, 2009),
– *Speech processing and ontologies for blind and visually-impaired people* (Tazawa et al., 2010; Kopeček and Oýlejýek, 2010),
– *Aids for speech-impaired people* (Sanger and Henderson, 2007; Saz et al., 2009; Yakcoub et al., 2008; Biswas and Samanta, 2008; Pfurtscheller et al., 2008),
– *Aids for deafblind people* (Miyagi et al., 2006; Ohtsuka et al., 2010),
– *Speech processing/aids for cognitively-impaired people* (Morris et al., 2010; Nikolova et al., 2010; Chu et al., 2009; Aluísio and Gasperin, 2010; Matausch and Peböck, 2010),
– *Inclusive e-technology for the disabled* (Rodriguez-Ascaso et al., 2010; Bernier et al., 2010; Freire et al., 2009).

5. Acknowledgments

We would like to thank Daniala Furrer for contributing the information about assistive technology for the physical impaired and Sigrid Jakobs for useful information according to the needs of the deaf and hearing impaired. Further, we thank Dirk Kochanek and Heiko Folkerts who supplied the technical details about Java and accessibility. We thank Sebastian Andres, Klaus Knopper, Sven Gukkes and Henning Oschwald who provided the information about Free Software/PDF and accessibility from the perspective of screen reader users, and John Glauert from University of East Anglia who contributed the detailed information on eSign. We thank the reviewer of this article for useful comments. We thank Dirk Kochanek and Vincent Gouws for reviewing the article. We are also grateful to Alexander Mehler, Paul Warner and Nils Diewald for useful comments and fruitful discussions. We thank Hansjörg Lienert (Tag It Guide, Dräger & Lienert, Figure 5) and Norbert Baron (Personal Mobile Language Translator, PMLT, Figure 10) for permitting the use of product images in this article.

6. List of relevant links to institutions and conferences on accessibility

1. UNITED NATIONS, Development and human rights for all: http://www.un.org/disabilities/index.asp
2. G3ict – The Global Initiative for Inclusive ICTs: http://g3ict.org/
3. CSUN – Annual International Technology & Persons with Disabilities Conference: http://csunconference.org
4. W4A – Cross-Disciplinary Conference on Web Accessibility: http://www.w4a.info
5. ICCHP – International Conference on Computers Helping People with Special Needs: http://www.icchp.org/
6. ASSETS – The International ACM SIGACCESS Conference on Computers and Accessibility: http://www.sigaccess.org/conferences/assets/
7. AHG – Annual Accessing Higher Ground Accessible Media, Web and Technology Conference: http://www.colorado.edu/ATconference/
8. AEGIS – Open Accessibility Everywhere; Groundwork, Infrastructure, Standards): http://www.aegis-project.eu/
9. Workshop on Representation and Processing of Sign Languages at LREC – International Conference on Language Resources and Evaluation http://www.lrec-conf.org/
10. ICAD – International Conference on auditory display: http://www.icad.org/
11. Inclusion Europe; Respect, Solidarity, Inclusion for People with Intellectual Disabilities: http://www.inclusion-europe.org/

7. Some awards for barrier-free design

– BIENE award http://biene-award.de/ (BIENE abbreviates "Barrierefreies Internet eröffnet neue Einsichten", which translates to *barrier-free internet opens up new insights*)
– Excellence Through Accessibility Award http://www.nda.ie/eta
– Access Award of the American Foundation of the Blind http://www.afb.org/Section.asp?SectionID=28&TopicID=236
– Universal Design Award http://www.bca.gov.sg/barrierFree/uda.html
– Jodi Awards http://www.jodiawards.org.uk/

Notes

1. We can ignore forms of inner perception like equilibrioception in this context of communication.
2. See http://www.ncsu.edu/project/design-projects/udi/ or http://ftb-esv.de/unides.html for universal design projects (both accessed December 21, 2011).
3. http://www.rehadat.de/rehadat/, as from November 29, 2011. REHADAT is an information system supporting the vocational integration of disabled persons that has been commissioned by the *Bundesministerium für Arbeit und Soziales* (Federal Ministry of Labor and Social Affairs) and was established by the *Institut der deutschen Wirtschaft Köln* (Cologne Institute for Economic Research).
4. See the homepage of the Web Accessibility Initiative at http://www.w3.org/WAI/ (accessed June 18, 2011)
5. See http://www.w3.org/TR/WCAG20/
6. See http://www.w3.org/TR/WCAG20/ #conformance (accessed June 1, 2011).
7. http://www.w3.org/WAI/impl/ (accessed June 1, 2011).
8. http://www.w3.org/WAI/train.html (accessed March 17, 2011.)
9. http://webaim.org/techniques/css/invisiblecontent/
10. http://www.w3.org/WAI/intro/aria.php (accessed June 1, 2011). Note that the current version of WAI-ARIA is still a candidate recommendation (2011–01–18) and it is not yet supported by all browsers and assistive technologies. The current status can be retrieved from http://www.w3.org/standards/techs/aria #w3c _all.
11. See http://dev.w3.org/html5/spec/Overview.html (accessed May 31, 2011.)
12. DictaSign http://www.dictasign.eu/ (accessed June 1, 2011.)
13. http://www.cs.sunysb.edu/~hearsay/ (accessed June 1, 2011.)
14. http://www.gnu.org/licenses/gpl.html (accessed March 17, 2011.)
15. http://knopper.net/knoppix-adriane/ (accessed March 17, 2011).
16. http://www.linaccess.org/ (accessed June 1, 2011.)
17. http://2009.hfoss.org/VizAudio (accessed June 1, 2011.)
18. http://accessibility.kde.org/, March 17, 2011.
19. http://projects.gnome.org/accessibility/, March 17, 2011.
20. http://www.adobe.com/pdf/ (accessed December 1, 2011).
21. http://www.section508.gov/ (accessed June 1, 2011).
22. http://www.gtk.org/ (accessed June 1, 2011).
23. http://projects.gnome.org/orca/ (accessed January 12, 2012).
24. http://sourceforge.net/projects/sue/ (accessed January 12, 2012).
25. *pdftotext* is part of the xpdf utils
26. “OpenOffice.org” is the official name for the domain as well as the software itself: http://www.openoffice.org/ (accessed June 1, 2011).
27. http://office.microsoft.com/en-us/word/
28. www.libreoffice.org/ (accessed December 5, 2011).
28. Available at http://www.access-for-all.ch/ (accessed December 5, 2011).
30. http://msdn.microsoft.com/en-us/library/ms697707.aspx (accessed December 8, 2011).
31. http://projects.gnome.org/accessibility/.
32. http://www.oracle.com/technetwork/java/index-jsp-141752.html (accessed June 1, 2011).
33. http://java.sun.com/products/jtapi/ (accessed June 1, 2011).

34. http://java.sun.com/products/java-media/speech/forDevelopers/jsapi-guide/index.html (accessed June 1, 2011).
35. http://www.oracle.com/technetwork/java/index-jsp-142945.html (accessed June 1, 2011).
36. http://www.eclipse.org/swt/, June 1, 2011.
37. http://docs.oracle.com/javase/tutorial/uiswing/misc/access.html (accessed January 12, 2012).
38. http://msdn.microsoft.com/en-us/library/ms696082 %28v=vs.85 %29.aspx (accessed June 1, 2011).
39. http://msdn.microsoft.com/de-de/library/ms727247.aspx (accessed June 1, 2011).
40. http://www.on-line-on.eu/ (accessed March 17, 2011).
41. http://www.mayer-johnson.com/category/boardmaker-family/(accessed December12,2011).
42. http://www.cilip.org.uk/about-us/medalsandawards/libraries-change-lives/pages/lclafinalist09.aspx #board, June 1, 2011.
43. http://www.inclusion-europe.org (accessed December 12, 2011).
44. http://www.simputer.org/ (accessed December 13, 2011).
45. http://www.simputer.org/simputer/spec/iml2.0/ (accessed March 17, 2011).
46. http://www.xml.com/pub/r/1070 (accessed June 1, 2011).
47. http://one.laptop.org/ (accessed June 1, 2011).
48. See http://one.laptop.org/map,asofDecember13,2011.
49. WHO Fact Sheet No. 282, May 2009, http://www.who.int/mediacentre/factsheets/fs282/en/ (accessed June 1, 2011).
50. We select an example instead of providing different diopter values in order to illustrate the difference visually. The 4-fold zoom is a test criterion in the BITV-test (BITV is the German Barrier-free Information Technology Regulation): http://testen.bitv-test.de/index.php?a=di&iid=1142&s=n (accessed June 1, 2011).
51. http://webaim.org/projects/screenreadersurvey/, http://webaim.org/projects/screenreadersurvey2/, and http://webaim.org/projects/screenreadersurvey3/ (each accessed January 12, 2012).
52. http://www.daisy.org/ (accessed December 15, 2011).
53. http://www.daisy.org/projects/save-as-daisy/ (accessed December 15, 2011).
54. The *gh* player is now called *RedHear PC*, see http://www.gh-accessibility.com/software/readhear_pc (accessed December 15, 2011).
55. http://www.xbox.com/kinect (accessed December 16, 2011.).
56. http://www.dlinfo.de/content/tagitguide.php (accessed December 18, 2011).
57. http://www.touchgraphics.com/research/pen.htm (accessed December 19, 2011).
58. http://www.wfdeaf.org/about, as of December 19, 2011.
59. WFD Policy, Position Paper on Technology & Accessibility 2007, http://www.wfdeaf.org/databank/policies/policy-position-paper-on-technology-accessibility (accessed December 19, 2011).
60. http://www.nchearingloss.org/can.htm (accessed June 1, 2011).
61. http://www.michdhh.org/assistive_devices/cart.html (accessed March 17, 2011).
62. http://www.sign-lang.uni-hamburg.de/dgs-korpus/index.php/hamnosys.html (accessed December 20, 2011).
63. www.signwriting.org (accessed June 1, 2011).
64. http://www.movementwriting.org/ (accessed June 1, 2011).

65. http://www.signwriting.org/lessons/cursive/shorthand/ (accessed June 1, 2011).
66. http://www.signbank.org/signpuddle/ (accessed March 17, 2011).
67. http://unicode.org/iso15924/iso15924-num.html (accessed March 17, 2011).
68. http://scripts.sil.org/cms/scripts/page.php?site_id=nrsi&id=OFL (accessed March 17, 2011).
69. http://www.signbank.org/ (accessed June 1, 2011).
70. http://www.sign-lang.uni-hamburg.de/projekte/hamnosys/hns3.0/hamnosysintrode.html (accessed June 1, 2011).
71. http://www.mijnwoordenboek.nl/ (accessed December 20, 2011).
72. For more information see http://www.signwriting.org/forums/software/archive/softarc15.html and http://www.signwriting.org/forums/software/swml/ (accessed June 1, 2011).
73. See http://cart-info.org/ and http://www.csi.cuny.edu/csa/rcd/staff_faculty.html (both accessed December 20, 2011).
74. See www.netac.rit.edu/downloads/TPSHT_CART.pdf for some application areas of CART in a university faculty (accessed December 20, 2011).
75. http://www.dgs-filme.de/GWHomepage/ (accessed December 21, 2011).
76. http://www.scomm.com/ubiduo (accessed December 20, 2011).
77. See http://www.global-communication-premium.com/pmlt-translator.en.html for a video that shows PMLT in action (accessed December 21, 2011).
78. http://www.dictasign.eu/ (accessed June 1, 2011).
79. http://signsci.terc.edu/dictionary/index.htm (accessed December 21, 2011).
80. http://www.vcom3d.com/ (accessed December 21, 2011).
81. http://www.sign-lang.uni-hamburg.de/tlex/start.htm (accessed March 17, 2011).
82. http://www.vhs-kaarst-korschenbroich.de/ (accessed March 17, 2011).
83. http://www.oeglb.at/?id=LH2006–02-06-3323 (accessed December 21, 2011).
84. See also http://www.deafblind.org.uk/ and http://www.deafandblind.com/eafblind.html (accessed March 17, 2011).
85. http://www.deafsign.com/ds/index.cfm?scn=article&articleID=46 (accessed June 1, 2011).
86. http://www.humanware.com/en-usa/products/blindness/deafblind_communicator/ (accessed June 1, 2011).
87. http://deafblindinternational.org/publications_handson.html (accessed December 21, 2011).
88. https://www.handytech.de/de/normal/produkte/fuer-blinde/tabli/index.html (accessed June 1, 2011).
89. http://www.handytech.de/en/normal/products/for-blind/braille-wave/index.html (accessed June 1, 2011).
90. http://www.handytech.de/en/normal/products/for-blind/braille-star40/index.html (accessed June 1, 2011).
91. http://www.incap.de/index.php?article_id=122&artcat=10&artnr=11337 (accessed June 1, 2011).
92. http://www.cancontrols.com/ (June 1, 2011).
93. http://www-01.ibm.com/software/pervasive/viavoice.html (accessed June 1, 2011).
94. http://www.nuance.com/naturallyspeaking/ (accessed June 1, 2011).
95. http://www.opera.com/browser/tutorials/nomouse/ (accessed March 17, 2011).
96. http://www.mouseless.de/ (accessed March 17, 2011).

References

Abou-Zahra, Shadi
2005 Semantic web enabled web accessibility evaluation tools. In: *Proceedings of the International Cross-Disciplinary Workshop on Web Accessibility*, W4A, pages 99–101. ACM.

Adi-Bensaid, Limor and Tova Most
2009 The acquisition of the prosodic word by children with hearing loss using a cochlear implant. *The Volta Review*, 109(1):5–31.

Ahmed, Aisha Shamsuna and Daniel Su Kuen Seong
2006 SignWriting on mobile phones for the deaf. In: *Proceedings of the 3rd international conference on Mobile technology, applications & systems*, Mobility'06. ACM.

Ali, N., N.Y. Khan and A.S. Imran
2007 Controlling mouse through eyes. In: *International Conference on Emerging Technologies*, ICET'07, pages 179–183.

Allwood, Jens and Elisabeth Ahlsén
2012 Multimodal communication. In: Alexander Mehler, Laurent Romary, and Dafydd Gibbon, editors, *Handbook of Technical Communication*. De Gruyter, Berlin/Boston.

Aluísio, Sandra Maria and Caroline Gasperin
2010 Fostering digital inclusion and accessibility: the PorSimples project for simplification of portuguese texts. In: *Proceedings of the Young Investigators Workshop on Computational Approaches to Languages of the Americas*, NAACL HLT'10, pages 46–53, Stroudsburg, PA, USA. Association for Computational Linguistics.

Alzheimer, Odilo (ed.)
1970 *Lexikon der Medizin*. Ullstein, Berlin.

Appelquist, Daniel, Dan Brickley, Melvin Carvahlo, Renato Iannella, Alexandre Passant, Christine Peray and Henry Story
2010 A standards-based, open and privacy-aware social web, W3C Incubator Group report. http://www.w3.org/2005/Incubator/socialweb/XGR-socialweb-20101206/. W3C Incubator Group. Accessed June 19, 2011

Atwood, Alan A., John Dennis Clarkson and Charlene R. Laba
1992 *Being In Touch: Communication and Other Issues in the Lives of People who*. Gallaudet University Press, Washington, D.C.

Aznar, Guylhem, Patrice Dalle and Clément Ballabriga
2006 Analysis of the differents methods to encode SignWriting in Unicode. In: *Workshop on the Representation and Processing of Sign Language: Lexicographic Matters and Didactic Scenarios*, LREC'06, pages 59–63. Evaluations and Language resources Distribution Agency (ELDA).

Bangham, J.A., S.J. Cox, R. Elliott, J.R.W. Glauert, I. Marshall, S. Rankov and M. Wells
2000 Virtual signing: capture, animation, storage and transmission – an overview of the ViSiCAST project. In: *IEE Seminar on Speech and Language Processing for Disabled and Elderly People*.

Berners-Lee, Tim, James Hendler and Ora Lassila
2001 The semantic web: A new form of web content that is meaningful to com-

puters will unleash a revolution of new possibilities. *Scientific American Magazine*, pages 24–30.

Bernier, Alex and Dominique Burger
2010 A Wiki-based environment for DAISY text production. PDF. Online available at Online-Proceedings of the 2010 Conference on Technology and Persons with Disabilities (CSUN): http://www.letsgoexpo.com/utilities/File/viewfile.cfm?LCID=3950 &eID=80000218. Accessed June 19, 2011

Bernier, Alex, Dominique Burger and Bruno Marmol
2010 Wiki, a new way to produce accessible documents. In: Klaus Miesenberger, Joachim Klaus, Wolfgang Zagler, and Arthur Karshmer, editors, *Proceedings of the 12th international conference on Computers helping people with special needs: Computers for Handicapped Persons, Part I*, pages 20–26, Berlin and Heidelberg. Springer.

BIK-Projekt
2010 BITV-Test, Prüfschritt 11.1.1, Angemessene Formate. http://testen.bitv-test.de/index.php?a=di&iid=1125. Accessed June 19, 2011

Biswas, Pradipta and Debasis Samanta
2008 Friend: A communication aid for persons with disabilities. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 16(2):205–209.

Blattner, Meera M., Denise A. Sumikawa and Robert M. Greenberg
1989 Earcons and icons: Their structure and common design principles. *Human-omputer Interaction*, 4(1):11–4.

Bohman, Paul
2004 Kognitive Behinderungen, Teil 1: Wir wissen immer noch zu wenig und wir tun noch viel weniger. http://www.einfach-fuer-alle.de/artikel/kognitive-behinderungen/. Accessed March 17, 2011

Borodin, Yevgen, Jeffrey P. Bigham, Amanda Stent and I. V. Ramakrishnan
2008 Towards one world web with HearSay3. In: *Proceedings of the international cross-disciplinary conference on Web accessibility*, W4A'08, pages 130–131, New York, NY, USA. ACM.

Borrino, Roberto, Marco Furini and Marco Roccetti
2009 Augmenting social media accessibility. In: *Proceedings of the International Cross-Disciplinary Conference on Web Accessibililty*, W4A'09, pages 54–57, New York, NY, USA. ACM.

Brück, Kurt
1997 Der thermische Kanal. In: Roland Posner, Klaus Robering, and Thomas A. Seboek, editors, *Semiotik: ein Handbuch zu den zeichentheoretischen Grundlagen von Natur und Kultur*, volume 1 of *Handbücher zur Sprach- und Kommunikationswissenschaft*, chapter II. Gegenstand I: Aspekte der Semiose – Kanäle, Medien und Kodes / General Topics I: Aspects of Semiosis – Channels, Media, and Codes, pages 325–330. de Gruyter, Berlin.

Brzoza, Piotr
2008 Multimedia MathReader for Daisy books. In: Klaus Miesenberger, Joahim Klaus, Wolfgang Zagler, and Arthur Karshmer, editors, *Proceedings of the 11th International Conference on Computers Helping People with Special Needs: Computers for Handicapped Persons*, volume 5105 of *Lecture Notes in Computer Science*, pages 875–878, Berlin and Heidelberg. Springer.

Carroll, Jeff and Fang-Gang Zeng
2007 Fundamental frequency discrimination and speech perception in noise in cochlear implant simulationsstar. *Hearing Research*, 231(1–2):42–53.
Carter, Jennifer Hardy (ed.)
1996 *Communication Problems After a Brain Injury or Stroke*. Pritchett & Hull Associates, Inc., Atlanta, GA.
Chathuranga, S.K., K.C. Samarawickramac, H.M.L. Chandima, K.G.T.D. Chathuranga and A.M.H.S. Abeykoon
2010 Hands free interface for human computer interaction. In: *International Conference on Information and Automation for Sustainability*, ICIAFs'10, pages 359–364.
Chen, Xiaoyu, Marilyn Tremaine, Robert Lutz, Jae-woo Chung and Patrick Lacsina
2006 AudioBrowser: a mobile browsable information access for the visually impaired. *Universal Access in the Information Society*, 5(1):4–22.
Chu, Hui-Chuan, Tsung-Yi Chen, Chia-Jou Lin, Min-Ju Liao and Yuh-Min Chen
2009 Development of an adaptive learning case recommendation approach for problem-based e-learning on mathematics teaching for students with mild disabilities. *Expert Systems with Applications: An International Journal*, 36:5456–5468.
Cooper, Holly L.
2006 A brief history of tactile writing systems for readers with blindness and visual impairments. *See/Hear*, 11(2):12–18.
Cornelssen, Iris and Christian Schmitz
2008 http://www.einfach-fuer-alle.de/studie/. Summary of the results of the study "Chancen und Risiken des Internets der Zukunft aus Sicht von Menschen mit Behinderungen" given at the *Aktion-Mensch*-Symposium, Wissenschaftspark Gelsenkirchen. Accessed June 19, 2011
da Rocha Costa, Antônio Carlos and Graçalitz Pereira Dimuro
2001 Supporting deaf sign languages in written form on the web. In: *Poster Proceedings of the Tenth International World Wide Web Conference*, WWW10, page p1011. URL http://www10.org/cdrom/posters/frame.html. Accessed March 17, 2011
da Rocha Costa, Antônio Carlos and Graçaliz Pereira Dimuro
2002 Signwriting-based sign language processing. In: Ipke Wachsmuth and Timo Sowa, editors, *Gesture and Sign Language in Human-Computer Interaction*, volume 2298 of *Lecture Notes in Computer Science*, pages 202–205. Springer, Berlin and Heidelberg.
Darvishy, Alireza, Hans-Peter Hutter, Alexander Horvath and Martin Dorigo
2010 A flexible software architecture concept for the creation of accessible pdf documents. In: Klaus Miesenberger, Joachim Klaus, Wolfgang L. Zagler, and Arthur I. Karshmer, editors, *Proceedings of the International Conference on Computers Helping People with Special Needs: Computers for Handicapped Persons, Part I*, Lecture Notes in Computer Science, pages 47–52, Berlin and Heidelberg. Springer.
Di Nardo, W., A. Scorpecci, S. Giannantonio, F. Cianfrone, C. Parrilla and G. Paludetti
2010 Cochlear implant patients' speech understanding in background noise: effect of mismatch between electrode assigned frequencies and perceived pitch. *The Journal of Laryngology & Otology*, 124(8):828–834.

Eberius, Wolfram
2008 Multimodale Erweiterung und Distribution von Digital Talking Books. Diplomarbeit, Technische Universität Dresden. Online available at: http://daisy.dynamic-designer.de/resources/diplom.pdf, accessed June 19, 2011.

Edman, Polly K.
1992 *Tactile Graphics*. American Foundation for the Blind, New York, NY.

Efthimiou, Eleni, Stavroula-Evita Fotinea, Christian Vogler, Thomas Hanke, John Glauert, Richard Bowden, Annelies Braffort, Christophe Collet, Petros Maragos and Jérémie Segouat
2009 Sign language recognition, generation, and modelling: A research effort with applications in deaf communication. In: Constantine Stephanidis, editor, *Universal Access in Human-Computer Interaction. Addressing Diversity*, volume 5614 of *Lecture Notes in Computer Science*, pages 21–30. Springer, Berlin and Heidelberg.

Eilers, H., S. Baron, T. Ortmaier, B. Heimann, C. Baier, T.S. Rau, M. Leinung and O. Majdani
2009 Navigated, robot assisted drilling of a minimally invasive cochlear access. In: *IEEE International Conference on Mechatronics*, pages 1–6.

Elliott, R., J. Glauert, J. Kennaway, I. Marshall and E. Safar
2008 Linguistic modelling and language-processing technologies for avatar-based sign language presentation. *Universal Access in the Information Society*, 6:375–391.

Elliott, Ralph, John Glauert, Vince Jennings and Richard Kennaway
2004 An overview of the SiGML notation and SiGMLSigning software system. In *Fourth International Conference on Language Resources and Evaluation*, LREC'04, pages 98–104.

Emmorey, Karen
1999 Do signers gesture? In: Lynn S. Messing and Ruth Campbell, editors, *Gesture, speech, and sign*, pages 133–159. Oxford University Press, New York.

Fajardo, Inmaculada, Julio Abascal and José Cañas
2004 The role of working memory and long term memory in deaf users' hypertext navigation: Review of guidelines for web accessibility. In: Christian Stary and Constantine Stephanidis, editors, *User-Centered Interaction Paradigms for Universal Access in the Information Society*, volume 3196 of *Lecture Notes in Computer Science*, pages 320–325. Springer, Berlin and Heidelberg.

Fajardo, Inmaculada, Markel Vigo and Ladislao Salmerón
2009 Technology for supporting web information search and learning in sign language. *Interact with Computers*, 21(4):243–256. doi: 10.1016/j.intcom.2009.05.005.

Felton, E.A., N.L. Lewis, S.A. Wills, R.G. Radwin and J.C. Williams
2007 Neural signal based control of the Dasher Writing System. In: *3rd International IEEE/EMBS Conference on Neural Engineering*, CNE'07, pages 366–370.

Freire, André P., Flávia Linhalis, Sandro L. Bianchini, Renata P.M. Fortes and Maria da Gra§a C. Pimentel
2009 Revealing the whiteboard to blind students: An inclusive approach to provide mediation in synchronous e-learning activities. *Computers & Education*, 54(4): 866–876.

Freyhoff, Geert, Gerhard Hess, Linda Kerr, Elizabeth Menzel, Bror Tronbacke and Kathy Van Der Veken
1998 Make it simple: European guidelines for the production of easy-to-read information for people with learning disability for authors, editors, information providers, translators and other interested persons. PDF. Available online at http://www.inclusion-europe.org/uploads/doc/99.pdf. Accessed June 19, 2011

Ghoul, Oussama and Mohamed Jemni
2009 A sign language screen reader for deaf. In: *Proceedings of the 5th Symposium of the Workgroup Human-Computer Interaction and Usability Engineering of the Austrian Computer Society on HCI and Usability for e-Inclusion*, USAB'09, pages 476–483, Berlin and Heidelberg. Springer.

Glauert, John
2002 SiGML standardisation. PDF. ViSiCAST milestone M5–8; online available at: http://www.visicast.co.uk/members/milestones/M5–8SiGMLStandard.pdf. Accessed June 19, 2011

Glennen, Sharon I.
1996 Introduction to augmentative and alternative communication. In: Sharon I. Glennen and Denise C. DeCoste, editors, *Handbook Of Augmentative And Alternative Communication*, chapter 1, pages 3–19. Singular Publishing Group Inc., San Diego and London.

Grassmuck, Volker
2004 *Freie Software – zwischen Privat- und Gemeineigentum*. Bundeszentrale für politische Bildung, Bonn.

Hanke, Thomas
2001 Auf dem Wege zu Language Resources für Gebärdensprachen. *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München*, 37:191–203.

Hanke, Thomas and Jakob Storz
2008 iLex – a database tool integrating sign language corpus linguistics and sign language lexicography. In: *Proceedings of the 3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora*, LREC'08, pages 64–67. Sixth International Language Resources and Evaluation.

Harper, Simon and Yeliz Yesilada
2007 Web authoring for accessibility (WAfA). *Web Semantics: Science, Services and Agents on the World Wide Web*, 5:175–179.

Hayes, Heather
2010 *Spelling in deaf children who wear cochlear implants*. PhD thesis, Washington University.

Heinrich, Mark
1998 Methoden und Techniken des nicht-visuellen Zugangs basierend auf Java. Diplomarbeit, Universität Stuttgart, Fakultät Informatik, Germany. URL http://elib.uni-stuttgart.de/opus/volltexte/1999/237/pdf/237.pdf. Accessed June 19, 2011

Hellbusch, Jan Eric
2005 Gestaltung barrierefreier PDF-Dokumente. *Information – Wissenschaft & Praxis*, (8). Special Issue *Barrierefreiheit im Internet*.

Hellbusch, Jan Eric and Kerstin Probiesch
2011 *Barrierefreiheit verstehen und umsetzen. Webstandards für ein zugängliches und nutzbares Internet*, chapter Säulen der Barrierefreiheit, pages 21–27. dpunkt Verlag, Heidelberg.
Hersh, Marion and Michael A. Johnson (eds.)
2008 *Assistive Technology for Visually Impaired and Blind People*. Springer, London.
Heuer, Herbert
1997 Der taktile Kanal. In: Roland Posner, Klaus Robering, and Thomas A. Sebeok, editors, *Semiotik: ein Handbuch zu den zeichentheoretischen Grundlagen von Natur und Kultur*, volume 1 of *Handbücher zur Sprach- und Kommunikationswissenschaft*, chapter II. Gegenstand I: Aspekte der Semiose – Kanäle, Medien und Kodes / General Topics I: Aspects of Semiosis – Channels, Media, and Codes, pages 300–306. de Gruyter, Berlin.
Heuwinkel, Roland
2003 PDF-Dokumente – lesbar für alle. http://www.einfach-fuer-alle.de/download/pdf_barrierefrei.pdf. Accessed June 19, 2011
Hori, J., K. Sakano and Y. Saitoh
2004 Development of communication supporting device controlled by eye movements and voluntary eye blink. In: *Annual International Conference of the Engineering in Medicine and Biology Society*, IEMBS'04, pages 4302–4305. IEEE.
Horrocks, Ian
2007 Semantic web: the story so far. In: *Proceedings of the international cross-disciplinary conference on Web accessibility*, W4A'07, pages 120–125. ACM.
Inclusion Europe
2009 Information for all – european standards for making information easy to read and understand. PDF. URL http://www.inclusion-europe.org/LLL/documents/Information\%20for\%20all.pdf. Accessed June 19, 2011
ISO/TS 16071
2003 *Ergonomics of human-system interaction – Guidance on accessibility for human-computer interfaces*. International Organization for Standardization. Revised by ISO 9241–171:2008.
ISO 15924:2004
2004 Information and documentation – Codes for the representation of names of scripts. International Organization for Standardization.
ISO 9999:2011
2011 Assistive products for persons with disability – classification and terminology. International Organization for Standardization.
Jacobs, Ian, Judy Brewer and Daniel Dardailler
2000 WAI resource: HTML 4.0 accessibility improvements. WAI makes strides towards universal web accessibility with HTML 4.0. http://www.w3.org/WAI/References/HTML4-access. Web Accessibility Initiative. Accessed May 31, 2011
Karpouzis, K., G. Caridakis, S.-E. Fotinea and E. Efthimiou
2007 Educational resources and implementation of a Greek sign language synthesis architecture. *Computer Education*, 49:54–74.

Kasturi, Kalyan S.
2006 *Signal processing strategies for better melody recognition and improved speech understanding in noise for cochlear implants*. PhD thesis, University of Texas at Dallas.

Kennaway, Richard
2002 Synthetic animation of deaf signing gestures. In: *Revised Papers from the International Gesture Workshop on Gesture and Sign Languages in Human-Computer Interaction*, GW'01, pages 146–157, London, UK.

Kennaway, Richard, J. R. W. Glauert and I. Zwitserlood
2007 Providing signed content on the internet by synthesized animation. *ACM Transactions on Computer-Human Interaction*, 14(3). Article No. 15.

Kochanek, Dirk
1994 Designing an offscreen model for a GUI. In: Wolfgang Zagler, Geoffrey Busby, and Roland Wagner, editors, *Computers for Handicapped Persons*, volume 860 of *Lecture Notes in Computer Science*, pages 89–95. Springer, Berlin and Heidelberg.

Kopeček, Ivan and Radek Oýlejýek
2010 Annotating and describing pictures – applications in e-learning and accessibility of graphics. In: *Computers for Handicapped Persons: Part I*, volume 6179 of *Lecture Notes in Computer Science*, pages 124–130. Springer, Berlin and Heidelberg.

Kröller, Jürgen
1997 Chemical channels. In: Roland Posner, Klaus Robering, and Thomas A. Seboek, editors, *Semiotik: ein Handbuch zu den zeichentheoretischen Grundlagen von Natur und Kultur*, volume 1 of *Handbücher zur Sprach- und Kommunikationswissenschaft*, chapter II. Gegenstand I: Aspekte der Semiose – Kanäle, Medien und Kodes / General Topics I: Aspects of Semiosis – Channels, Media, and Codes, pages 306–316. de Gruyter, Berlin.

Lai, Wei-Dian and Fu-Kuang Yeh
2009 Modeling performances of virtual channels for cochlear implant systems. In: *Proceedings of the WRI World Congress on Computer Science and Information Engineering*, pages 351–355, Washington, DC, USA. IEEE.

Landwehr, Klaus
1997 Der optische Kanal. In: Roland Posner, Klaus Robering, and Thomas A. Seboek, editors, *Semiotik: ein Handbuch zu den zeichentheoretischen Grundlagen von Natur und Kultur*, volume 1 of *Handbücher zur Sprach- und Kommunikationswissenschaft*, chapter II. Gegenstand I: Aspekte der Semiose – Kanäle, Medien und Kodes / General Topics I: Aspects of Semiosis – Channels, Media, and Codes, pages 288–294. de Gruyter, Berlin.

Leas, Dennis, Neil Soiffer and Masakazu Suzuki
2009 Word, math, and DAISY: It all adds up. DOC. Available through the Online Proceedings of the The 24th Annual International Technology and Persons with Disabilities Conference (CSUN'09) at http://www.letsgoexpo.com/utilities/File/viewfile.cfm?LCID=3308&eID=80000144. Accessed June 19, 2011

Leuninger, Helen
2000 Mit den Augen lernen: Gebärdenspracherwerb. In: Hannelore Grimm, editor, *Enzyklopädie der Psychologie, Bd. 3: Sprachentwicklung*, pages 229–272. Hogrefe, Göttingen.

Liddell, Scott K.
1984 Think and believe: Sequentiality in american sign language. *Language*, 60(2):372–399.

Liddell, Scott K. and Robert E. Johnson
1989 American sign language: The phonological base. *Sign Language Studies*, 64:195–277.

Linnartz, Christine
2003 Gehörlose können doch lesen …? http://www.einfach-fuer-alle.de/artikel/gehoerlos/. Accessed June 19, 2011

Löffler, Conny
2004 *Die Verschriftung der Gebärdensprache*, volume 7 of *SIGNA Beiträge zur Signographie*. Denkmalschmiede Höfgen, Edition Waechterpappel, Leipzig.

Lomranz, J. and A. Shapira
1974 Communicative patterns of self-disclosure and touching behavior. *Journal of Psychology: Interdisciplinary and Applied*, 88(2):223–227.

Lopes, Rui, Konstantinos Votis, Luís Carriço, Dimitrios Tzovaras and Spiridon Likothanassis
2009 Towards the universal semantic assessment of accessibility. In: *Proceedings of the ACM symposium on Applied Computing*, pages 147–151, New York, NY, USA. ACM.

Mally, Gertrud
1993 *Redewendungen in der Deutschen Gebärdensprache*. Signum, Seedorf.

Martin, Jean-Claude and Tanja Schultz
2012 Multimodal and speech technology. In: Alexander Mehler, Laurent Romary, and Dafydd Gibbon, editors, *Handbook of Technical Communication*. De Gruyter, Berlin/Boston.

Matausch, Kerstin and Birgit Peböck
2010 EasyWeb – a study how people with specific learning difficulties can be supported on using the internet. In: Klaus Miesenberger, Joachim Klaus, Wolfgang Zagler, and Arthur Karshmer, editors, *Computers for Handicapped Persons*, volume 6179 of *Lecture Notes in Computer Science*, pages 641–648. Springer, Berlin and Heidelberg.

McDonnell, Mark D., Anthony N. Burkitt, David B. Grayden, Hamish Meffin and Alex J. Grant
2010 A channel model for inferring the optimal number of electrodes for future cochlear implants. *IEEE Transactions on Information Theory – Special Issue on Information Theory in Molecular Biology and Neuroscience*, 56:928–940.

McMullen, Declan and Donal Fitzpatrick
2010 Combining web services and DAISY for the production and delivery of audio tactile diagrams. In: *Proceedings of the 12th International Conference on Computers Helping People with Special Needs: Computers for Handicapped Persons, Part I*, ICCHP'10, pages 12–19.

Mets, Marilyn B., Nancy M. Young, Arlene Pass and Janice B. Lasky
2000 Early diagnosis of Usher syndrome in children. *Transactions of the American Ophthalmological Society*, 98(2):37–45.

Millar, Sally and Janet Scot
2003 What is augmentative and alternative communication? In: H Galloway and S. V. Millar, editors, *Augmentative Communication in Practice*, chapter 1, pages 3–12. University of Edinburgh, CALL Centre, Edinburgh, UK.

Millar, Susanna
1997 *Reading by Touch*. Routledge, London, UK.

Miyagi, M., M. Nishida, Y. Horiuchi and A. Ichikawa
2006 Analysis of prosody in finger Braille using electromyography. In: *Conf Proc IEEE Eng Med Biol Soc*, volume 1, pages 4901–4904.

Mohamed, A.W. and R. Koggalage
2007 Control of mouse movements using human facial expressions. In: *Proceedings of the 3rd International Conference on Information and Automation for Sustainability*, ICI-AFS'07, pages 13–18.

Moller, Peter
1997 The electric and magnetic channels. In: Roland Posner, Klaus Robering, and Thomas A. Seboek, editors, *Semiotik: ein Handbuch zu den zeichentheoretischen Grundlagen von Natur und Kultur*, volume 1 of *Handbücher zur Sprach- und Kommunikationswissenschaft*, chapter II. Gegenstand I: Aspekte der Semiose – Kanäle, Medien und Kodes / General Topics I: Aspects of Semiosis – Channels, Media, and Codes, pages 316–325. de Gruyter, Berlin.

Morelli, Anthony
2010 Haptic/audio based exergaming for visually impaired individuals. *SIGACCESS Accessibility and Computing*, 96:50–53. doi: http://doi.acm.org/10.1145/1731849.1731859.

Morris, Robert R., Connor R. Kirschbaum and Rosalind W. Picard
2010 Broadening accessibility through special interests: a new approach for software customization. In: *Proceedings of the 12th international ACM SIGACCESS conference on Computers and accessibility*, AS-SETS '10, pages 171–178, New York, NY, USA. ACM.

Moustakas, Konstantinos, Georgios Nikolakis, Konstantinos Kostopoulos, Dimitrios Tzovaras and Michael G. Strintzis
2007 Haptic rendering of visual data for the visually impaired. *IEEE Multimedia*, 14(2): 62–72.

Mylonakis, Vasileios, John Soldatos, Aristodemos Pnevmatikakis, Lazaros Polymenakos, Alex Sorin and Hagai Aronowitz
2008 Using robust audio and video processing technologies to alleviate the elderly cognitive decline. In: *Proceedings of the 1st international conference on PErvasive Technologies Related to Assistive Environments*, PE-TRA'08, pages 1–8, New York, NY, USA. ACM.

ANSI/NISO Z39.86–2002
2002 *Specifications for the Digital Talking Book*. National Information Standards Organization.

Nielsen, Jakob
2000 *Designing Web Usability*. New Riders, Indianapolis, IN.

Nikolova, Sonya, Marilyn Tremaine and Perry R. Cook
2010 Click on bake to get cookies: guiding word-finding with semantic associations. In: *Proceedings of the 12th international ACM SIGACCESS confer-*

ence on Computers and accessibility, ASSETS'10, pages 155–162, New York, NY, USA. ACM.

Ohtsuka, S., S. Hasegawa, N. Sasaki and T. Harakawa
2010 Communication system between deaf-blind people and non-disabled people using body-braille and infrared communication. In: *7th IEEE Consumer Communications and Networking Conference*, CCNC'10, pages 1–2.

Oikonomou, Theofanis, Konstantinos Votis, Dimitrios Tzovaras and Peter Korn
2010 Designing and developing accessible Java swing applications. In: *Proceedings of the 12th International Conference on Computers Helping People with Special Needs: Computers for Handicapped Persons, Part I*, ICCHP'10, pages 186–188.

Pa, Judy, Stephen M. Wilson, Herbert Pickell, Ursula Bellugi and Gregory Hickok
2008 Neural organization of linguistic short-term memory is sensory modality-dependent: Evidence from signed and spoken language. *Cognitive Neuroscience*, 20(12):2198–2210.

Paabo, Regina, Monika Foedisch and Liivi Hollmann
2009 Rules for Estonian sign language transcription. *Trames*, 13(4):401–424.

Padden, Carol A. and Darline C. Gonsalius
2003 How the alphabet came to be used in a sign language. *Sign Language Studies*, 4(1):10–33.

Pal, Joyojeet and Rodrigo Fonseca
2003 Bringing devices to the masses: a comparative study of the Brazilian Computador Popular and the Indian Simputer. PDF. Online available at University of California: http://www.icsi.berkeley.edu/pubs/bcis/bringingdevices03.pdf. Accessed June 19, 2011

Papadogiorgaki, Maria, Nikos Grammalidis, Nikos Sarris and Michael G. Strintzis
2004 Synthesis of virtual reality animations from SWML using MPEG-4 body animation parameters. In: *Workshop on the Representation and Processing of Sign Languages on the occasion of the Fourth International Conference on Langauge Resources and Evaluation*, pages 259–266.

Papadogiorgaki, Maria, Nikos Grammalidis, Dimitrios Tzovaras and Michael G. Strintzis
2005 Text-to-sign language synthesis tool. In: *Proceedings of the 13th European Signal Processing Conference*, Eusipco'05.

Petitto, Laura Ann, Robert J. Zatorre, Kristine Gauna, E. J. Nikelski, Deanna Dostie and Alan C. Evans
2000 Speech-like cerebral activity in profoundly deaf people processing signed languages: Implications for the neural basis of human language. *PNAS*, 97(25):13961–13966. doi: 10.1073/pnas.97.25.13961.

Pfurtscheller, Gert, Gernot R. Möller-Putz, Reinhold Scherer and Christa Neuper
2008 Rehabilitation with brain-computer interface systems. *Computer*, 41:58–65.

Prillwitz, S., R. Leven, H. Zienert, T. Hanke and J. Henning
1989 *HamNoSys Version 2.0: Hamburger Notationssystem für Gebärdensprachen. Eine Einführung*. SIGNUM.

Ramakrishnan, I. V., Amanda Stent and Guizhen Yang
2004 Hearsay: enabling audio browsing on hypertext content. In: *Proceedings of the 13th international conference on World Wide Web*, pages 80–89, New York, NY, USA. ACM.

Rico, Mariano, David Camacho and Corcho Oscar
2009 Personalized handling of semantic data with MIG. *20th International Workshop on Database and Expert Systems Applications*, 0:64–68.

Rodriguez-Ascaso, Alejandro, Erik Zetterström, Martin Böcker, Helge Hüttenrauch, Michael Pluke and Matthias Schneider
2010 Inclusive e-services for all: identifying accessibility requirements for upcoming interaction technologies. In: *Proceedings of the 12th International Conference on Computers Helping People with Special Needs: Computers for Handicapped Persons, Part I*, ICCHP'10, pages 135–138.

Romano Jr., Nicholas C.
2003 Customer relationship management for the web-access challenged: Inaccessibilityof the fortune 250 business web sites. *International Journal of Electronic Commerce*, 7(2):81–117.

Rotard, Martin, Christiane Taras and Thomas Ertl
2008 Tactile web browsing for blind people. *Multimedia Tools and Applications*, 37(1):53–69.

Rupp, Thomas
2003 Device for transmission of signs and characters from a data-processing system to a deaf-blind person. http://www.freepatentsonline.com/EP0809223B1.html. Patent EP0809223. Accessed June 19, 2011

Sanger, T.D. and J. Henderson
2007 Optimizing assisted communication devices for children with motor impairments using a model of information rate and channel capacity. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 15(3): 458–468.

Saz, Oscar, Shou-Chun Yin, Eduardo Lleida, Richard Rose, Carlos Vaquero and William R. Rodríguez
2009 Tools and technologies for computer-aided speech and language therapy. *Speech Communication*, 51:948–967.

Schmidt, Robert F. and Gerhard Thews
1997 *Physiologie des Menschen*. Springer, Berlin and Heidelberg and New York, 27 edition.

Schulmeister, Rolf
2001 ViSiCAST: Virtual signing, capture, animation, storage and transmission – übersetzung in und Generierung von virtueller Gebärdensprache im Fernsehen und Internet. http://www.zhw.uni-hamburg.de/pdfs/ViCa.PDF. Accessed June 19, 2011

Schweinberger, Birgit
2009 Methoden der Unterstützten Kommunikation für Menschen mit Autismus – Analyse und Vergleich anhand von dreizehn Kriterien. Diploma thesis, University of Wien. Online available at: http://othes.univie.ac.at/4824/1/2009-03-08_9806369.pdf. Accessed June 19, 2011

Seeman, Lisa
2003 The semantic web, web accessibility, and device independence. *SIGCAPH Computers and the Physically Handicapped*, 76: 19–20.

Seok, Soonhwa, Edward L. Meyen and Boaventura DaCosta (eds.)
2010 *Handbook of Research on Human Cognition and Assistive Technology: Design, Accessibility and Transdisciplinary Perspectives*. Medical Information Science Reference.

Sinclair, Jeff, Philip Hingston and Martin Masek
2007 Considerations for the design of exergames. In: *Proceedings of the 5th international conference on Computer graphics and interactive techniques in Australia and Southeast Asia*, GRAPHITE '07, pages 289–295, Perth, Australia.

Slawinski, Niki
2005 Das Internet hören und fühlen. Eine qualitative Studie zur "Barrierefreie Informationstechnik-Verordnung – BITV"am Beispiel von blinden Schülern. Magister thesis, Ruhr-Universität Bochum, Fakultät für Philosophie, Pädagogik und Publizistik. published online at http://www.barrierefreies-webdesign.de/spezial/internet-hoeren-und-fuehlen/kommunikations situation.html. Accessed March 17, 2011

Sohrabi, Shirin
2010 Customizing the composition of actions, programs, and web services with user preferences. In: *9th International Semantic Web Conference*, ISWC2010.

Stephanidis, Constantine (ed.)
2009 *The Universal Access Handbook*. CRC Press, Boca Raton, FL.

Stevens, Alexander A. and Kurt Weaver
2005 Auditory perceptual consolidation in early-onset blindness. *Neuropsychologia*, 43(13):1901–1910.

Stokoe, William C.
1960 Sign language structure: An outline of the visual communication systems of the american deaf. Studies in linguistics: Occasional papers (No. 8). Buffalo: Dept. of Anthropology and Linguistics, University of Buffalo.

Stokoe, William C., Dorothy C. Casterline and Carl G. Croneberg
1965 *A Dictionary of American Sign Language on Linguistic Principles*. Gallaudet University Press, Washington, DC.

Strelnikov, K., J. Rouger, P. Barone and O. Deguine
2009 Role of speechreading in audiovisual interactions during the recovery of speech comprehension in deaf adults with cochlear implants. *Scandinavian Journal of Psychology*, 50(5):437–444.

Strobbe, Christophe, Jan Engelen and Vincent Spiewak
2010 Generating DAISY books from openoffice.org. In: Klaus Miesenberger, Joachim Klaus, Wolfgang Zagler, and Arthur Karshmer, editors, *Computers Helping People with Special Needs*, volume 6179 of *Lecture Notes in Computer Science*, pages 5–11, Berlin, Heidelberg. Springer.

Strube, Gerhard and Gerda Lazarus
1997 Der akutische Kanal. In: Roland Posner, Klaus Robering, and Thomas A. Seboek, editors, *Semiotik: ein Handbuch zu den zeichentheoretischen Grundlagen von Natur und Kultur*, volume 1 of *Handbücher zur Sprach- und Kommunikationswissenschaft*, chapter II. Gegenstand I: Aspekte der Semiose – Kanäle, Medien und Kodes / General Topics I: Aspects of Semiosis – Channels, Media, and Codes, pages 294–300. de Gruyter, Berlin.

Tazawa, Naoyuki, Shinichi Totihara, Yukio Iwahana, Atsushi Imai, Nobumasa Seiyama and Toru Takagi

2010 Rapid listening of DAISY digital talking books by speech-rate conversion technology for people with visual impairments. In: Klaus Miesenberger, Joachim Klaus, Wolfgang Zagler, and Arthur Karshmer, editors, *Computers Helping People with Special Needs*, volume 6179 of *Lecture Notes in Computer Science*, pages 62–68. Springer, Berlin and Heidelberg.

Valli, Clayton and Ceil Lucas
2000 *Linguistics of American Sign Language: An Introduction*. Gallaudet University Press, Washington, DC, USA, 3 edition.

Wagner, Susanne
2005 Intralingual speech-to-text-conversion in real-time: Challenges and opportunities. In: *Proceedings of the First Conference on Multidimensional Translation: Challenges of Multidimensional Translation*, MuTra'05.

Wakefield, Claire E., Judi Homewood and Alan J. Taylor
2006 Early blindness may be associated with changes in performance on verbal fluency tasks. *Journal of Visual Impairment & Blindness*, 100(5):306–310.

Waltinger, Ulli and Alexa Breuing
2012 Internet-based communication. In: Alexander Mehler, Laurent Romary, and Dafydd Gibbon, editors, *Handbook of Technical Communication*. De Gruyter, Berlin/Boston.

WCAG 2.0
2008 Web content accessibility guidelines (WCAG) 2.0. http://www.w3.org/TR/WCAG20/. World Wide Web Consortium, Web Accessibility Initiative. Accessed June 19, 2011

Wilken, Etta (ed.)
2006 *Unterstützte Kommunikation. Eine Einführung in Theorie und Praxis*. Kohlhammer, Stuttgart, 2 edition.

Wöhrmann, Stefan
2005 *Handbuch zur GebärdenSchrift*. Jacobsen, Hamburg, Germany.

Yakcoub, M. S., S.-A. Selouani and D. O'Shaughnessy
2008 Speech assistive technology to improve the interaction of dysarthric speakers with machines. In: *3rd International Symposium on Communications, Control and Signal Processing*, ISCCSP'08, pages 1150–1154.

Yesilada, Yeliz, Robert Stevens, Simon Harper and Carole Goble
2007 Evaluating DANTE: Semantic transcoding for visually disabled users. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 14 (3).

Yesilada, Yeliz, Sean Bechhofer and Bernard Horan
2008 Dynamic linking of web resources: Customisation and personalisation. In: Manolis Wallace, Marios C. Angelides, and Phivos Mylonas, editors, *Advances in Semantic Media Adaptation and Personalization*, volume 93 of *Studies in Computational Intelligence*, pages 1–24. Springer.

Zhao, Haixia, Catherine Plaisant, Ben Shneiderman and Jonathan Lazar
2008 Data sonification for users with visual impairment: A case study with georeferenced data. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 15(1).

Zukowski, John
2005 Accessibility. In: *The Definitive Guide to Java Swing*, pages 805–811. Apress, New York, BY.

20. Artificial Interactivity

Stefan Kopp and Ipke Wachsmuth

1. Introduction

While in the social sciences and humanities interactivity is considered reserved to real persons, the term "interaction" has long been used in its own right in the field of machine studies and applications, such as in "human-computer interaction". As exemplified below, interactivity also attains meaning in the context of artificial systems – *artificial interactivity* – to the extent that technical artifacts can make human observers perceive them as actors in social interaction, that is, exhibit behaviors that can be ascribed agency. The use of interactive systems and the instantaneity of informations they provide cover a wide range of interaction types, from tool-like operation up to the idea of conducting "social" interaction with robots or artificial agents. The present chapter focuses on this latter type of interaction.

From the sociological perspective, social interactions take place when at least two individuals perceive another, perceive the perceptions of the other, share a joint focus of attention, and orient their actions at the other's behavior (cp. Goffman, 1983). Further, understanding others' intentions and representing them as being able to understand intentions are decisive factors in human cooperation and communication. While not all of these conditions may be true for a technical artifact, there is evidence that computers are being regarded as social actors. For instance, Reves and Nass (1996) have observed that people prefer interacting with computers that have identifiable personalities and are able to adapt their style of communication to that of the user. Studies in the area of affective computing – i.e. relating to, arising from, or deliberately influencing emotions – suggest that machines will communicate more effectively with humans if they can perceive and express emotions (Picard 1997). Indeed, many humans show the same variety of emotional responses as in face-to-face interaction between humans when competing against a computer game (Becker et al. 2005), and humanoid agents in a biased "happy" condition elicit more emotionally positive responses in humans than when conducting dialogue in a neutral condition (Von der Pütten et al. 2008). In this vein, the view that humans are "users" of an application has shifted to that of a "partnership" with artificial agents that can be considered able to take initiative as autonomous entities (Negrotti, 2005). This is particularly true for embodied agents that have conversational or cooperative behaviors and allow for the integration of non-verbal cues in social dialogue (Bickmore and Cassell 2005) and, more generally, multimodal embodied communication (Wachsmuth, Lenzen and Knoblich 2008). But how can this artificial interactivity be achieved technically?

In recent times, a variety of technologies have been developed that enable interacting with and by means of embodied artificial agents. On the one hand, through the use of remotely controlled embodied entities or so-called “avatars”, humans are increasingly able to interact with each other in ways and places naturally not possible. Examples include shared virtual worlds like Second Life, Virtual Reality environments for cooperative work, or tele-operated robots in human-robot teams. This requires technology to capture a human’s relevant behaviors, or even goals, and try to transport or recreate them as an embodied simulation of an interactant. For such systems, the most relevant questions concern the recording and recognition of what a human wants to convey, transmitting it in form of an appropriate representation, and reproducing it as intelligible embodied behavior. Communication thereby often subserves the control of tasks that the user carries out in the remote environment (e.g., navigating around or manipulating objects).

On the other hand, technical system can figure as embodied agents that interact and cooperate autonomously with human partners. One prerequisite for this is to make human-machine communication easier and more intuitive. The common approach is to learn from and increase resemblance with human-like face-to-face interaction (Cassell et al. 2000), and effects like those mentioned above indicate the effectivity of this approach. Interaction with embodied agents is nowadays possible in various settings ranging from entertainment systems to serious applications like information presentation agents in health communication, interactive museum guides, web-based agents for customer support, or social robots for care-taking. In such settings, methods and techniques are needed to build embodied agents that can engage in natural interaction protocols including turn-taking, grounding, or multimodal behavior with speech, facial expressions and gestures. The present chapter reviews these techniques from the perspective of both technical communication through avatars and with autonomous embodied agents.

2. Communicating through embodied agents and avatars

Games and virtual worlds allow humans to dive into a simulated environment with an avatar of arbitrary form and to interact with other avatars through the use of text chat, voice, and non-verbal behavior animations. Avatars can further manipulate objects and perform simulated manual tasks. Research has only started to investigate this remote form of interaction mediated through avatars. First evidence shows that avatar-based interaction can, at least to some extent, be used by humans to form relationships and online communities (Ridings and

Gefen 2004). Real-world interaction behavior has been found in the virtual world, for instance spatial social navigating and proxemics (Friedman, Steed and Slater 2007). Nevertheless, the interaction through an avatar and, even more restricting, through a graphical user interface with which the avatar is controlled is too impoverished to create the impression of unrestrained, fluent interaction. Thus, most of the time, human participants in Second Life engage in classical text chat (Weitnauer et al. 2008) while employing their avatars rather for marking presence and social display. This is in line with evidence showing an influence of avatars on perceived social presence, intimateness, co-presence, and emotionally-based trust in computer-mediated communication (Bente et al. 2008).

Research has started to amalgamate interactive virtual worlds with real environments by use of Augmented Reality technology (e.g., AR Second Life project[1]). While promising to increase a human's presence at distant places, this comes with challenges to enable users to control their avatar's body and face like their own, especially for interactive behavior. First work has explored the possibilities to use motion tracking for avatar control. Ideally, subtle facial expressions, the dynamics of body movements, or gaze are transmitted through the avatar to enable simulated social interaction. However, existing systems so far focused only on the control of very basic actions like locomotion (Oshita 2006). Some researchers have responded to this by trying to make avatars smarter. The basic idea is to let the user specify only what she wants to communicate and to employ intelligent agent technology to let the avatar produce supplementary behaviors autonomously. One example of this is the BodyChat system (Vilhjálmsson and Cassell 1998) that analyzes the text a user inputs and automatically animates attention, salutations, turn taking, back-channel feedback or facial expression with the avatar. The techniques used for this stem from work on conversational agents, to which we turn in the remainder of this chapter.

3. Communicating with embodied agents

Computer-generated characters that demonstrate many of the properties of human face-to-face interaction have been developed under the term "Embodied Conversational Agents" (ECA) (Cassell et al. 2000). ECAs aim for a multimodal interaction where verbal and nonverbal modalities of human conversation, i.e. speech, facial displays, hand gestures, or body stance, are used naturally by both interaction partners to (pro-)actively participate and reciprocate in an evolving dialogue. This implies a number of the requirements for interactive agents.

3.1. Requirements

A number of requirements are imposed by the demand for natural language interaction and dialogue with ECAs, e.g., the ability to process speech acts, manage the initiative, or keep track of the grounding state of facts. There has been extensive work on these issues in the realm of spoken language dialogue systems (e.g., Jokinen and McTear 2009). We focus here on the requirements that apply additionally and specifically to the notion of embodied interactive agents. We start with classical ones that have been pointed out repeatedly and then consider issues that are relatively new in the design of ECAs but are crucial for cooperative and social interaction.

Multimodality. Humans have several modalities such as hand gestures, facial displays, eye gaze, intonation, etc. at their disposal and employ them concertedly and with astonishing smoothness to pursue multiple goals in parallel (see Allwood and Ahlsén 2012 as well as Lücking and Pfeiffer 2012 both in this volume). Implementations of interactive artificial agents are required to also incorporate these multiple channels of communication and to understand and produce their natural usage and functions in context. Roughly, one can differentiate between verbal and nonverbal communication channels, the latter of which can be refined into vocal and non-vocal behavior (Wallbott 1994). Verbal and nonverbal modalities afford different ways of encoding meaning with different degrees of adequacy: language is well suited for coding of symbolic information, hand gestures for visuo-spatial information, while all nonverbal modalities can indicate intention to take the turn. Multimodality also implies tight coordination between behaviors. For example, coverbal gestures and language form an integrated system and coordinate temporally as well as with respect to their meaning or pragmatics functions (McNeill 1992; Kopp et al. 2008).

Nonverbal behavior can be differentiated into three functional aspects relevant for embodied interactive agents: Discourse functions are closely related to verbal behavior and work either as complements, supplements or substitutes of speech in conveying content and constructing discourse. Dialogue functions include turn-taking to manage the exchange of speaker and listener roles (e.g. taking the turn by raising a hand, or keeping it by averting eye gaze), as well as feedback signals (e.g. head nods, verbal back-channels) that are vital for the process of jointly establishing information between interaction partners (grounding). Finally, relational or socio-emotional functions refer to the impact of nonverbal behavior on the perception and evaluation of other individuals (Mehrabian and Ferris, 1967). Historically, artificial agents were based on work in speech processing (see Martin and Schultz 2012 in this volume) and dialogue systems and hence focused on discourse and dialogue functions, and socio-emotional aspects started to be investigated in greater detail later.

Timing. The fact that multiple behaviors are produced in parallel raises the issue of temporal relationships in-between them and the time scales of interaction. The synchrony among events within communicative behavior, or the lack thereof, is meaningful in conversation itself. In speech and gesture, temporal proximity of the meaningful part of the gesture (the so-called gesture stroke) and the co-expressive linguistic elements has even been taken as constitutive of speech-gesture correspondence (McNeill 1992). Timing is also crucial in response behavior. Cassell et al. (2000) noted that already the "slightest delay in responding to conversational events may be taken to indicate unwillingness to cooperate or a strong disagreement" (p.34). This implies challenges for the development of artificial agents that are to interact on different time scales, down to the sub-second level, and to deal with synchrony across modalities. ECA architectures, consequently, feature multi-threaded and concurrent processing of multiple input/output signals.

Behaviors and functions. Perhaps the most central assumption made in ECAs to deal with the variety of communicative behaviors and to provide both modularity and a principled way of combining different modalities, is to separate behaviors (e.g., a head nod) from their functions (e.g., signal affirmation). The core distinction proposed by Cassell et al. (2000) was between propositional functions and interactional functions, paralleling the demarcation of discourse and dialogue functions. This distinction has been fundamental in ECA design, in which producing propositional information requires to plan multi-sentence output and to determine the order of presentation of interdependent facts, while processing interactional information rests on a model of the current state of the conversation accounting for the roles of the interactants, turns, discourse structure, or grounding states of information. The mapping between behaviors and functions, however, is highly complex and redundant: many behaviors fulfill multiple functions and many functions get realized by multiple behaviors. In particular, nonverbal behaviors are subtle and inter-dependent and represent a complex communication system that poses a number of open questions. For example, a smile can be perceived rather differently depending on whether the head is tilted or not. Modeling attempts try to approximate this by means of complex taxonomies of functions and behaviors (see below) and mapping rules.

Turn-taking. One crucial requirement concerns an ability for a specific kind of dialogue functions, namely, to engage in fluent human-like turn-taking. Turn-taking refers to the process of managing the floor between interactants (Duncan 1972) and secures the cooperative progress of interaction, independently of the cooperativeness of individual goals (Goffman 1983). ECAs must thus be able to keep track of the conversational status as well as to perceive and produce turn-relevant cues. Classical approaches assume a limited number of states of the conversation (e.g., *NotPresent*, *Present*, *UserTurn*, *OnHold*, *AgentTurn*; cf. Cassell et al. 2000) along with transitions between them tied to distinct, context-dependent interactional functions (e.g., *give-turn*, *want-turn*, *take-turn*) that correspond to behavioral cues. This approach has been confined both in terms of unclear states (e.g., *Gap*, *Overlap;* Lessmann et al. 2004) as well as functions (e.g., *yield-turn*, *hold-turn*). Recent work tries to acquire turn-taking rules either by models of machine learning (Jonsdottir, Thorisson and Nivel 2008) or to model it in conjunction with the underlying understanding and response processes (Kopp et al., 2008a).

Fluent Feedback. Artificial interactivity requires agents to be active in interaction also when not having the turn. That is, they must show engagement and provide feedback or back-channeling (Yngve, 1970) while not speaking. This has been neglected for a long time in human-agent interaction, probably due to a reminiscence of classical human-computer interaction in which interaction is mainly conceived of as command-execution cycles. Realizing that interactivity builds upon tighter reciprocity and responsiveness, also with regard to those aspects along which listeners evaluate what they are receiving (do I perceive? do I understand? do I agree? how do I feel about it?), researchers started to imbue agents with feedback behavior. This work mainly concentrates on feedback-giving, combining incremental input processing with behavior generation (for an overview see Kopp et al., 2008a). Other work has combined this with attempts to model personality traits and to process social user signals that can impact the affective state of the agent and its sensitive listening behaviors (Schröder et al. 2008).

Based on these general requirements, the remainder of this chapter will review methods and techniques for achieving artificial interactivity in embodied agents. We will specifically concentrate on how interactive behavior is processed and how it is produced in such agents. We will first demonstrate how these requirements apply and are met in an example interaction, in which the virtual human *Max* collaborates with a human partner in a Virtual Reality (VR) environment. Afterwards, we will give a more detailed discussion of how communicative behavior can be processed and generated in such interactions.

4. Interaction with the virtual human Max

As one incarnation of an artificial interactive agent, the A.I. group at Bielefeld University has been developing the virtual human Max (Kopp, Jung, Leßmann, Wachsmuth, 2003). In the scenario considered here, realized and implemented in the SFB 360 "Artificial Situated Communicators", Max is visualized in human-size in a VR environment where he assists a human in assembling complex aggregates. Both interactants stand across a table that has a number of parts lying on it. The human user – being equipped with stereo glasses, data gloves, optical position trackers, and a microphone – can address Max and issue commands in natural language along with coverbal gestures, or can directly grasp and manipulate the graphical models to carry out assembly actions (e.g. to insert a bolt into a bar). The agent has the same capabilities, being able to initiate assembly actions or to engage in multimodal dialogue with prosodic speech, gesture, eye gaze, and emotional facial expressions. In this setting, the human user and Max become collaboration partners in a situated interaction, in which Max provides support whenever his partner does not know how to go on constructing an aggregate. Figure 1 shows a typical interaction in which both partners cooperatively solve a construction problem (see Lessmann, Kopp and Wachsmuth 2006 for a more detailed discussion).

This interaction example demonstrates many of the requirements pointed out above:

- *Mixed initiative:* Initiative, seen as the license to induce an intention with the interaction partner, switches flexibly between the interlocutors. While the human has the initiative in the beginning (1) when he poses a goal to achieve, Max assumes initiative in 3 by bringing up a new sub-goal. The human user takes the initiative back in 4 or 12 by performing a request. When having the initiative, Max recognizes that his interaction partner refuses (as in 11) or hesitates to conduct an action (as in between 14 and 15), and he performs the action himself.
- *Goal refinement:* Originally, the user asks for a propeller to be built, defining the overall goal of the assembly procedure. This goal is refined implicitly by the user who employs specific parts in the construction actions 7 and 10. Max recognizes and keeps track of these implicit situational commitments and refines his subsequent explanations accordingly in 11.
- *Failure detection, feedback and repair:* Max gives feedback, e.g., to acknowledge an instruction, to accept a suggestion, or to report back on the occurrence and kind of a construction failure performed by the human (as in 7). In the latter case, Max undoes the action, retraces his assembly explanations, and then re-explains the wrongly performed manipulation in a more elaborated way (now emphasizing the middle hole of the bar in 9).

1 User: “Let us build a propeller.”

2 Max: “Okay.”

3 Max: “First, a bolt needs to be inserted…”

4 User: “Which bolt?”

5 Max: “…in the middle of a bar.”

6 Max: “Any bolt.”

7 User puts the yellow bolt into the first hole of a three-hole bar.

8 Max: “No, that was the wrong hole.”
Max undoes the action.

9 Max: “Again, a bolt needs to be inserted in the middle hole of a bar.”

10 User puts the yellow bolt into the middle hole of a three-hole-bar.

11 Max: “Alright. Now, insert <*pointing gesture*> this bolt into another three-hole-bar.”

12 User: “Do it yourself.”

13 Max: “Okay, I will do it.”

14 Max connects the objects as instructed.

15 Max: “And now the bars need to be turned <*iconic gesture*> crosswise.”

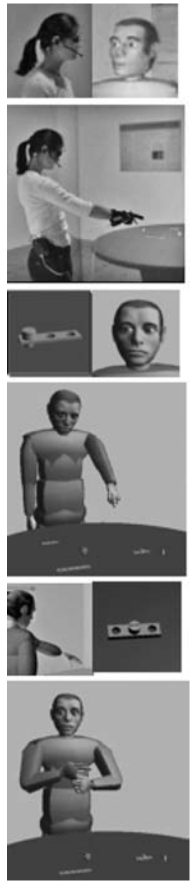


Figure 1. Example of an interaction in which the virtual human *Max* assists the human partner in building a propeller (Lessmann, Kopp and Wachsmuth, 2006).

– *Turn-taking:* Throughout the dialogue, both interactants manage a smooth exchange of turns. In 4, the human interlocutor interrupts Max and rigorously takes the turn to pose a clarification question. Max grants the floor at the next possible point (after 5) and answers the question (6).
– *Engagement and emotional display:* Both agents establish and maintain a connection while jointly undertaking the construction, i.e. they show engagement in their collaboration. Max demonstrates this by committing himself and contributing to the task, and by displaying on his face that even his mood is affected by the outcome of actions.

In sum, interactions like this are characterized by a high degree of interactivity, with frequent switches of initiative and roles of the interactants. The participants reconcile their contributions to the interaction with the need for regulating it by performing multiple activities simultaneously, asynchronously, and in multiple modalities. Moreover, both partners interact as intentional agents with beliefs, goals, and plans to pursue, taking account of the shared activity and the commitments imposed by their collaboration. The reader is referred to (Lessmann, Kopp and Wachsmuth 2006) for a detailed discussion of how perception, cognition and action is modeled and intertwined in Max to allow for this kind of human-agent interaction. Here, we will review the techniques and methods that are required to enable the multimodal communicative behavior, and its processing and generation, respectively, upon which such an interaction rests.

5. Processing communicative behavior

The processing of communicative behavior commonly follows a recognition pipeline leading, coarsely, from sensory data of different input modalities to preprocessing and feature extraction, via segmentation of input streams and recognizing familiar patterns, to multimodal integration and interpretation of observed communication (see Figure 2). In this section, we will discuss different approaches and techniques for preprocessing and multimodal fusion. The data most commonly used for multimodal input to interactive systems are speech and gestural signs, which will be the main focus of this section (for a comprehensive review see Sowa, 2006). Mentioned briefly, there is increasing work on processing eye gaze and facial expressions, for instance, with respect to deriving emotional features or to enable audiovisual speech recognition (Potamianos et al. 2003).

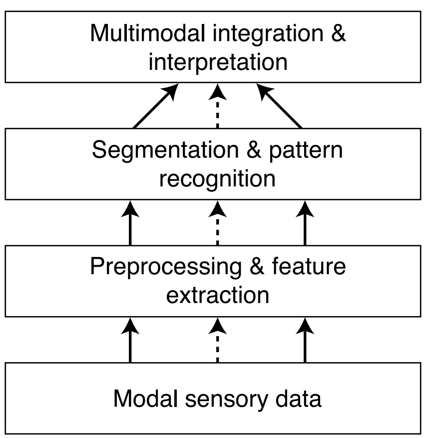


Figure 2. General structure of a pipeline for processing multimodal communicative behavior.

5.1. Capturing modal sensory data and preprocessing

Speech input is commonly picked up by microphones mounted to the user's head/neck, in some cases microphone arrays are used to capture speech input from the distance. The signal is then further processed by speech recognizers. For gestures, the processing chain begins with capturing hand movements, position and orientation, as well as angles of the finger joints for the hand posture. Broadly, capturing is realized either by so-called *contact* methods using data-gloves or markers for a tracking system mounted on the user's hands and arms, or *non-contact* methods based on video or infrared cameras (review see Wu and Huang, 1999). Preprocessing steps like filtering and so-called inverse kinematics are needed to reduce noise in the captured sensory data and to fit an assumed *hand model* or *body model* in the input data, which then allows to read out angles or position vectors. Such a body model represents knowledge about the movement characteristics of human bodies and the constraints this puts on the way a gesture can be performed. This is made available either in the preprocessing stage or during recognition. There are *kinematic models* describing the skeletal structure of joints and limbs, and *dynamic models* which describe motion as the result of the application of forces and torques. Benefits of applying a body model include: estimation of missing sensor information; prediction of movement trajectories; adaptation to the characteristics of an individual user; detection/rejection of impossible configurations or movements (for detail see Sowa, 2006).

5.2. Segmentation and pattern recognition

Once sensory data have been captured, recognition algorithms can be applied. Interaction behaviors are meaningful signals that extend over a period of time. Recognizing them hence entails a *segmentation problem*, which in the case of gesture refers to the problem of filtering out a gesture's expressive phase and determining units of meaning from the continuous stream of data delivered by the sensing devices. Likewise, in speech recognition the input is a stream of sampled sound data, which has to be segmented to deliver discrete words as output. By and large, the data can be described as a discrete *n*-dimensional vector that contains all sensor information at a certain point in time. In the case of different asynchronous data sources as often the case in multimodal interaction, a temporal alignment procedure has to be applied beforehand (Latoschik, 2001a).

The processing approaches to identify a gesture in a time series of such data vectors can be distinguished by the source of knowledge utilized. In *implicit* approaches, this knowledge is obtained via learning from examples; in *explicit* approaches, the recognition knowledge is modeled by the system designer or the user (cf. Sowa, 2006). If the underlying structure of an input is not available, implicit approaches are advantageous, the disadvantage being the potentially large amount of time and data needed for training.

Gesture segmentation is either performed separately as a preprocessing step, or built in as an integral part of the recognition model as, for example, in approaches using *Hidden Markov Model*s (HMM) (Jurafsky and Martin, 2009) (see below). Segmentation in *preprocessing* neglects the gesture's semantics and employs explicit spatiotemporal cues to determine unit borders. For example, Latoschik (2001b) suggested eight general features for segmenting gestural articulation, including action and pause, deviation from rest positions, primitive movement profiles, or internal and external symmetry. Howell and Buxton (1999) describe a vision-based system that computes a measure for movement from the difference between two successive images. A movement amount exceeding a threshold is taken as signal of the beginning of a gesture. Hofmann, Heyer, and Hommel (1998) used the velocity profiles of the hand and the finger joints; a unit of meaning is extracted for each phase in which the absolute hand or joint velocity exceeds a certain threshold. Koons, Sparrell, and Thorisson (1993) used rules to group continuous streams of whole-hand configuration and movement features into so-called *gestlets*. A similar method of grouping was proposed by Wexelblat (1995) where input data is segmented into *frames* based on sign changes of the first derivate in selected features; a new unit is created, for instance, when the movement direction changes from up to down. Harling and Edwards (1997) proposed to use hand tension and relaxation as a gesture segmentation cue; the unit borders are found at the local minima of the tension curve, while the presumably meaningful posture is expressed at the maximum.

In many cases, gesture and speech recognition can be viewed as *pattern classification* task, using standard pattern recognition techniques such as template matching and feature-based recognition (Benoit et al., 2000). Pursuing this paradigm, a given sequence of data vectors is allotted one class in a pre-defined set of classes via a classification function. Template matching refers to choosing the class whose reference feature vector matches an input feature vector, which is computed from the sequence of data vectors, best according to a distance metric. The reference vector may be explicitly provided or implicitly computed via learning. Popular implicit techniques to implement a classification function are *Hidden Markov Models* (HMM) (Jurafsky and Martin, 2009) and *Artificial Neural Networks* (ANN) (Anderson, 1995), both of them subject to training. HMM are well suited for the modeling and classification of time-varying signals and have been applied very successfully to speech recognition (Young, 1996). Gesture recognizers based on HMM may, for example, exploit the spatial configuration of the arms as feature vectors. For instance, Rigoll, Kosmala, and Eickeler (1998) described an HMM system based on difference images of video streams that achieves a recognition rate of about 93 % for 24 isolated gestures. The above-mentioned system by Hofmann et al. (1998) correctly classified 96 % of the gestures of a training set of 100 dynamic gestures. A combination of HMM and ANN techniques was used by Corradini (2002), in which the emission probabilities of an HMM are computed by a partially recurrent ANN. The system recognized five different dynamic gestures from monocular video images with an accuracy of approximately 91 %.

The pattern classification approach towards gesture recognition is widely used, but also criticized as unnatural (e.g., Wexelblat, 1995), mainly because of its restriction to a pre-defined gesture vocabulary, which conflicts with the natural variety of conversational gestures. *Feature-based* methods try to recognize a gesture from a number of different, decisive features instead. Wexelblat (1995) presented an analysis module that processed data from three electromagnetic trackers and CyberGloves, and consists of a body model, signal segmentation, feature detection, a path analysis, and an integration stage. Segments of linear change are combined to proto-features, which are then integrated to more complex features. The prototype implements 30 features and proto-features. The data of all feature detectors are filled into a sequence of frames that describe static and dynamic information of a coherent movement fragment. The final processing stage integrates successive frames if they belong to a higher-order movement segment that was separated by the segmenters. This is similar to the *gestlet* approach adopted by Koons et al. (1993), in which discrete hand posture, orientation, and movement values are bundled and analyzed by a gesture parser. The parser produces a description of the hand motion including the distinction of different movement phases, which can be used to determine a gesture category, e.g. a pointing gesture. Fröhlich and Wachsmuth (1998) proposed a feature-based gesture recognition system using a subset of the sign language notation HamNoSys (Prillwitz et al., 1989) for gesture specification. HamNoSys symbols are used as basic features and fused in an integration stage into a complete gesture according to application-dependent rules. Latoschik (2001a,b) presented a framework that extends the feature-based approach to multimodal input and is particularly suited for real-time interaction in virtual reality systems.

5.3. Multimodal integration and interpretation

Different models and mechanisms are needed to integrate gesture with other modalities, in particular with speech, to a common complex of meaning. Generally, two main subtypes of approaches towards integrating information from different modalities can be distinguished (Oviatt, 2003):

(1) *Feature-level*, or early fusion approaches, relate recognition information from different modalities in a pre-semantic stage; this can be applied if the modalities are strongly coupled such as (acoustic) speech signals and (visual) lip shape information.

(2) *Semantic-level*, or late fusion approaches, independently derive semantic information from different modalities and combine them afterwards to end up with a common meaning specification of a multimodal command or interaction.

A main challenge especially for the latter approaches is the *correspondence problem* (Srihari, 1994). Applied to gesture and speech integration, its central issue is how to correlate chunks of information from one modality to the other. Usually, temporal proximity is employed as a cue to determine corresponding input information. Wachsmuth (1999) demonstrated an integration method based on the idea of communicative rhythms, where a regular pulse defines temporal integration windows. Grammar-based techniques were used to either implicitly consider temporal proximity by linear ordering of input events, or via explicit temporal constraint or integration rules that can pertain to syntactic or semantic features of the input modalities. The grammar-based approach originated from the processing of pointing, or deictic gestures – the gestures used most frequently in multimodal systems for object reference. A cornerstone of multimodal integration was the *Put-That-There* system (Bolt, 1980), where pointing was evaluated whenever indexicals like "this" and "there" are detected in speech (speech-driven integration). The speech-pointing integration mechanism employed for the CUBRICON prototype (Neal and Shapiro, 1991) is based on a multimodal grammar modeled as an Augmented Transition Network (ATN) in which pointing references to objects may appear in noun phrases or locative adverbial phrases. A similar integration approach based on unification grammars, but also able to cope with indirect deictic reference such as pars-pro-toto pointing, was implemented in the XTRA prototype (Wahlster, 1991). In Burger and Marshall's (1993) approach, pointing gestures contribute solely to the attentional model (focus space) as a part of a discourse model, and gestural information from the attentional model was used to supply missing data whenever the speech channel contains indexicals.

A frequently applied semantic-level fusion method is to consider the meaning specification as a frame structure of attribute-value pairs (Waibel et al., 1996): First, the input from each modality is evaluated and produces a partially filled frame. Corresponding frames are then integrated until all attribute slots are filled and the application command is fully specified. Johnston et al. (1997) propose an integration method based on the unification of typed feature structures that allows building up structural hierarchies. In order to cope with the uncertainty of unimodal information sources, Kaiser et al. (2003) extend the unification approach by allowing multiple alternative input and output interpretations tagged by their respective recognition probabilities. The speech and gesture recognizers in their system provide a *n*-best list of the most probable input hypotheses for each modality. The probability of a multimodal integrated feature structure is computed based on the probabilities of the input hypotheses.

An alternative approach based on finite-state machines (FSM), leading to less complex formalisms and better integratability with speech recognition systems, was proposed by Johnston and Bangalore (2001). They model multimodal grammars with FSMs consisting of one input tape for each modality and an output tape containing the meaning of the multimodal speech-act after parsing. A similar, but more flexible approach in which temporal constraints can be expressed explicitly was described by Latoschik (2002), representing multimodal grammars with augmented transition networks enriched by temporal information (tATN). State transitions here model lexical input (words), temporal tests on the presence of gesture attributes, and allow for logical negation and conjunctions of attributes. Chai, Hong, and Zhou (2004) proposed an integration method for reference resolution that encodes each input modality as attributed relational graphs. Nodes represent possible reference objects, while arcs stand for temporal or semantic relations between the referents. Multimodal integration is thus expressed as a graph matching problem. The matching algorithm maximizes the overall fit of the modality-specific graphs.

While many approaches pertain to symbolic and deictic gesture recognition, few approaches have concerned *iconic gestures*, i.e. gestures depicting shape. The pioneering works of Koons et al. (1993) and Sparrell and Koons (1994) were already mentioned above in the context of gesture segmentation and feature processing. In their prototype system, called ICONIC, users could manipulate objects by issuing verbal and gestural commands. A speech-driven interpretation determines suitable places for gesture segments (gestlets) that complement the spoken utterance. A best fit between shape-related descriptions of objects and hand shapes is established by an *iconic mapping* algorithm. Sowa (Sowa 2006; Sowa and Wachsmuth, 2009) extended this to more generic shape-related iconic gestures and to a symmetrical semantic-level fusion with shape-related verbal expressions. Based on a comprehensive corpus of speech-gesture shape descriptions, a formal representation for the semantics of multimodal shape-related expressions was proposed, the *Imagistic Description Tree* (IDT). Operational IDT shape descriptions are derived for the verbal utterance as well as the iconic gesture independently and then integrated using sub-graph unification.

6. Generating communicative behavior

Similar to the recognition pipeline, the generation of communicate behavior is commonly organized in stages (see Figure 3) adopted from natural language and document generation systems (Reiter and Dale 2000). The starting point is a set of communicative goals formed by a dialogue manager to encompass all relevant propositional (discourse) and interactional (dialogue) functions to achieve in the next contribution(s). For example: acknowledge the previous information re-

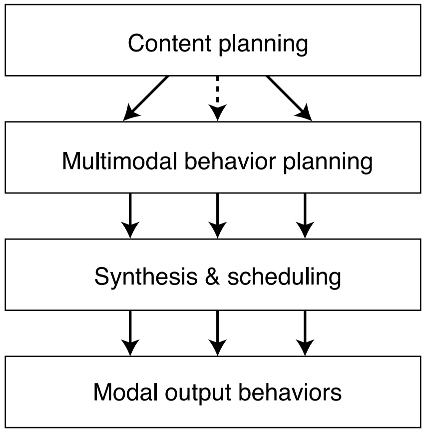


Figure 3. General structure of a pipeline for generating multimodal communicative behavior.

quest; inform the interlocutor about a particular person; keep the turn; reduce social distance. From such a goal set, content planning determines a complete specification of the semantic-pragmatic aspects of the target dialogue act without any reference to physical behavior (unless it is explicitly intended). Behavior planning then chooses a coordinate multimodal surface form that can realize this content in the given context situation, before the respective behaviors are generated by means of speech synthesis and nonverbal behavior animation. This coarse pipeline structure is shared by most existing agents and has even been adopted by current standardization efforts to specify XML-based interface languages between these stages (see also Stührenberg 2012 in this volume): *Function Markup Language* (FML; Heylen et al., 2008) and *Behavior Markup Language* (BML; Kopp et al. 2006). In the following, we will discuss the single stages in more detail, focusing again on the modalities of speech and gesture.

6.1 Content representation and planning

The selection of content and functions for a particular contribution is a highly situation-sensitive task and depends not only on the interlocutor's last contributions but also the overall task domain, the current discourse, or the intentions and goals of the agent. This problem has been tackled in modeling attempts to natural language generation (Reiter and Dale 2000) and dialogue management (cf. Jokinen and McTear 2009). It comprises mainly two steps, content determination and discourse structuring. The result is a tree-like content plan in which basic discourse units (usually propositions) are the leaf nodes and their position in the tree is constituted by discourse relations that hold between them or larger discourse units. Such relations are often some sort of rhetorical relations (cf. *Rhetorical Structure Theory*; Mann and Thompson, 1988). For example, one proposition can mean to *contrast* another proposition, or one larger discourse unit to *elaborate* another unit. Such relations have lately also been used to model the relation between semantics of different modalities (Lascarides and Stone 2009).

Special attention must be given to the problem of planning content for multimodal, multi-functional interactive behaviors. The distribution and coordination of content across modalities is known as *multimodal fission* (Müller et al. 2003) and has been tackled first in systems that generate multimodal or multimedia information presentations. Similar to the integration problem discussed above, fission is usually performed using some sort of multimodal output grammars to form and specify utterance plans (e.g., unification grammars). In doing so, most systems assume a common meaning representation for all behaviors, i.e., a single format to specify the semantics and functions of an entire multimodal utterance. Usually a speech-act-based representation is assumed (e.g. Cassell et al. 2000) building upon an ontology of discourse, dialogue and socio-emotive functions (e.g., Poggi and Pelachaud 2000, Kopp et al. 2005). Poggi and Pelachaud (2000) propose a taxonomy of more general *communicative acts* that combine propositional content with a performative and add aspects like degree of certainty, power relationship, interest in the requested action or provided information, type of social encounter, and affective state. This work has led to representation frameworks like FML (see above) or EmotionML[2] developed to provide a specification of the communicative goals needed for multimodal behavior, as well as the interaction context in which they have to be realized.

Regarding content or semantics representation, the use of a common symbolic formalism (e.g. Cassell, Stone and Yan 2000) proves to be sufficient only for symbolic and indexical signs like words and emblematic or pointing gestures. Iconic gestures, which are used to convey spatial information about objects or events, call for formalisms that enable mental rotation, spatial perspective taking, level of detail reduction, or salient feature determination. Kopp, Bergmann and Wachsmuth (2008) propose a multimodal content representation and simulate the interfacing of propositional knowledge with visuo-spatial knowledge in order to determine the content for speech and iconic gestures. Similar to the semantic representation of iconic gestures (see above, IDT), an analogous imagistic representation format is used for visuo-spatial information which, via so-called *multimodal concepts*, links to the propositional semantics of spatial language.

6.2. Behavior planning and realization

Once content and function specifications have been determined, appropriate behaviors have to be chosen or formulated. For speech, this problem is called *microplanning* and consists of three major sub-tasks (Reiter and Dale 2000):

lexicalization, aggregation, and generating referring expressions. Technical approaches to microplanning can be differentiated into *plan-based* and *template-based*. Template-based models (e.g. Becker 2002) draw from a repository of larger lexico-grammatical templates (e.g., verbal phrase structures or whole sentences) and fill empty slots in them depending on the information at hand. They can be used in limited, static domains for which they prove to be practical and to ease the modeling task. Plan-based approaches (e.g. Stone et al. 2001) utilize a planning algorithm with a search process to generate a vast number of possible grammatically correct realizations and pick from them according to current constraints and needs. The SPUD system (*Sentence Planning Using Descriptions*, Stone et al. 2001) manages a representation of shared and private knowledge to test possible verbal realizations against semantic and pragmatic constraints. Such generic approaches can solve all three sub-tasks at once, but suffer from high computational costs, a need for detailed linguistic resources, and tedious, error-prone modeling. However, there is no principled demarcation between template-based and plan-based approaches – templates can be arbitrarily small, down to the size of single words or even morphemes, and the problem of choosing and combining them can become a full-fledged planning problem (cf. Van Deemter et al. 2005).

Methods for producing nonverbal behavior have predominantly adopted a *lexicon-based approach* (see, e.g., Poggi et al. 1999 for gaze). Behaviors in the lexicon are usually parametrized and annotated with possible functions they fulfill. The BEAT system (Cassell et al. 2001) employed a large number of XML-based mapping rules to pick from the lexicon every behavior that might be possible (over-generalization) and then applied priority rules or filters to resolve resource conflicts and reduce this selection of behaviors. This approach has been adopted widely for all nonverbal behaviors regardless of whether they can be lexicalized or not, e.g., symbolic gestures or stereotypical facial expressions. Behavior lexicons thereby often simply contain the animation of the respective behaviors, modeled by a human designer or acquired by means of motion capture. This affords control over the quality and naturalness of the behavior, but suffers from a lack of flexibility actually needed to adapt a behavior (e.g. form, timing, body parts) to fit a multimodal utterance context or a particular speaker style. Researchers have proposed procedures for modifying behavior animations in correspondence with emotional state, personality, or cultural conventions. Ruttkay (2007) defined different styles for social or ethnical group in a dictionary of meaning-to-gesture mappings with optional modifying parameters to specify the motion characteristics of a gesture. In a similar account Hartmann et al. (2006) investigated the modification of gestures to carry a desired expressive content while retaining their original semantics. Bodily expressivity is defined with a small set of dimensions such as spatial/temporal extent, fluidity or power. These parameters of expressivity can be used in combination to modify gestures.

Similar to the step form pattern-based to feature-based approaches in recognition, generation approaches have started to generate behaviors on the spot by assembling smaller features. The NUMACK system (Kopp, Tepper and Cassell 2004) used empirically investigated patterns of human gesture composition, assuming iconic mappings between visuo-spatial features of referent objects and gestures features like handshape, position, or movement trajectory. The feature-based framework was embedded in a *multimodal microplanner* that could derive and coordinate the form of both natural language and gesture directly for given communicative goals.

Trying to meet the problems of explicit modeling, mainly a lack of heuristic knowledge about nonverbal behavior, another line of research uses *data-driven methods* to simulate gesturing behavior with the help of *implicit models* learned from data. Stone et al. (2004) recombine motion captured pieces with speech samples to recreate coherent multimodal utterances. Units of communicative performance are re-arranged while retaining temporal synchrony and communicative coordination that characterizes peoples' spontaneous delivery. Neff et al. (2008) learn statistical gesture profiles from annotated multimodal behavior. Using the character-specific probabilities, the system takes arbitrary annotated texts as input and produces synchronized discourse gestures and beats. The resulting interactive behaviors succeed in making a virtual character look lively and natural and conform a given performer's style. However, so far data-based approaches are inherently limited by the scope of the recorded behavior and struggle with sparsity of training data that cover a sufficient variety and number of co-occurrences of behavioral forms and functions.

Bergmann and Kopp (2009) present an approach that combines implicit data-based methods with explicit model-based techniques. Their generation network for iconic gestures (Figure 4; left) utilizes a *Bayesian Decision Network* (BDN) that correlates input features (shape properties of the referent object, communicative goal, information structure, previous gesture) with features like gesture occurrence, handedness, or representation technique (e.g., drawing a 2D outline vs. shaping a 3D volume vs. adopting static posture). The relations are defined by means of probability distributions learned from empirical data. Final *decision nodes* employ explicit rules to determine features like hand position, orientation, and handshape from the visuo-spatial representation of the object being referred to. In evaluations (Bergmann, Kopp and Eyssell, 2010), the model reproduced empirically observed gestures that were rated positive and helpful by human interlocutors.

The final stage of the production pipeline is in charge of behavior synthesis and scheduling (see Figure 3). Various kinds of techniques for speech synthesis and animation rendering have been employed at this stage, which is beyond the scope of the present chapter. In general, systems emphasizing lifelikeness and degree of realism use data-based models like unit selection for text-to-speech

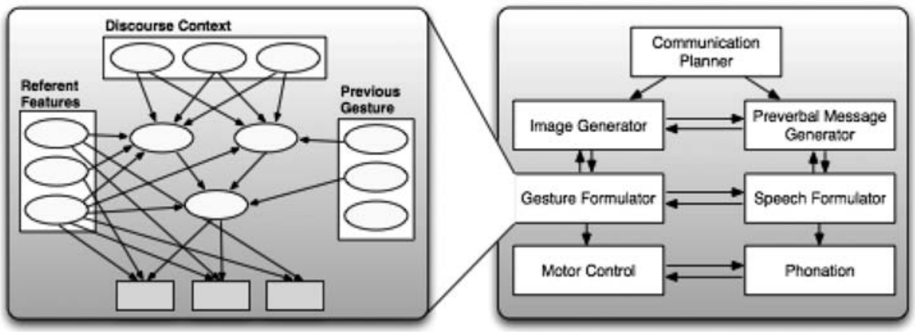


Figure 4. Outline of an interactive behavior production architecture (*right*) and the Bayesian Decision Network used for gesture generation network (*left*) (Bergmann and Kopp 2010).

and motion capturing for nonverbal behavior animation (see, e.g., Stone et al. 2004). Systems that strive for autonomy and flexibility in interaction tend to employ model-based techniques to create, at least, major parts of the behavior autonomously (see, e.g., Kopp and Wachsmuth 2004).

7. Conclusions and prospect of future work

In this chapter we have discussed requirements and techniques for building embodied agents which can be ascribed *artificial interactivity* – the ability to engage in fluent face-to-face communication in which they come across as actors whose behaviors can be ascribed agency. A key requirement for this is the ability to engage in multimodal communication, that is, to perceive, segment and integrate the communicative behaviors of the human interaction partner, to extract their intended functions in a context-sensitive way, and to reciprocate adequately with flexibly generated verbal and nonverbal behaviors. We have reviewed the main stages of the pipelines commonly used for processing and generating the respective behaviors. We have seen that a range of approaches and techniques exist for both tasks. Future work in this field is multifold. One strand of research has started to explore deeply emotional aspects of interaction, enabling agents to recognize a human's affective state and to simulate emotions and emotional expressions themselves (e.g., Becker at al., 2005). This work will continue to bring social aspects into the human-agent interaction and research shows positive effects on agents' acceptance and evaluation (Von der Pütten et al., 2008). Taking this further, another strand of research targets the limited capabilities of technical agents to engage in the dynamic mutual coordinations that can be found in human-human interaction. For example, conversation partners are often found to align linguistically (Pickering and Garrod, 2004), to mimic each other (Chartrand and Bargh, 1999), or to synchronize their timing of behaviors (Bernieri and Rosenthal, 1991). Such adaptations (1) are interactively contingent, i.e. their occurrence is linked to the interactional context including the partner, (2) act in a coordinative fashion between the interactants, and (3) correlate with the communicative success as well as social success of an interaction (Kopp 2010). That is, going beyond the mere necessity of establishing informational common ground, they contribute to a state of *social resonance* which entails affiliation and rapport between the interactants (Lakin et al., 2003). Such coordination phenomena are also considered relevant for interactive humanoid artifacts, which are found to behave as social actors and to evoke similar coordinations in humans (see Kopp 2010 for an overview and discussion). Work in this direction will thus pave the way to systems that can embody artificial interactivity comprehensively, by demonstrating the social and communicative intentions to engage, collaborate, and succeed in their interactions with their human partners.

Notes

1. See http://arsecondlife.gvu.gatech.edu/proj_playspace.html
2. See http://www.w3.org/2005/Incubator/emotion/XGR-emotionml/

References

Allwood, Jens and Elisabeth Ahlsén
2012 Multimodal communication. In: Alexander Mehler, Laurent Romary and Dafydd Gibbon (Eds.): *Handbook of Technical Communication*. Berlin/Boston: De Gruyter.
Anderson, James A.
1995 *An Introduction to Neural Networks.* Cambridge, MA: MIT Press.
Becker, Christian, Helmut Prendinger, Mitsuru Ishizuka and Ipke Wachsmuth
2005 Evaluating affective feedback of the 3D agent Max in a competitive cards game. In: Jianhua Tao, Tieniu Tan and Rosalind Picard (eds.), *Affective Computing and Intelligent Interaction*, 466–473. LNCS 3784. Berlin/Heidelberg: Springer.
Becker, Tilman
2002 Practical, template-based natural language generation with TAG. In: *Proceedings of TAG+6*, 101–104.
Bente, Gary, Nicole C. Krämer and Felix Eschenburg
2008 Is There Anybody Out There? Analyzing the Effects of Embodiment and Nonverbal Behavior in Avatar-Mediated Communication. In: Konijn, Elly A., Sonja Utz, Martin Tanis and Susan B. Barnes (eds.), *Mediated Interpersonal Communication*. Mahwah, NJ: Lawrence Erlbaum.

Benoit, Christian, Jean-Claude Martin, Catherine Pelachaud, Lambert Schomaker and Bernhard Suhm
2000 Audio-visual and multimodal speech-based systems. In: Dafydd Gibbon, Inge Mertins and Roger Moore (eds.), *Handbook of multimodal and spoken dialogue systems: Resources, terminology and product evaluation,* 102–203. Dordrecht, The Netherlands: Kluwer.

Bergmann Kirsten, Stefan Kopp
2009 Increasing Expressiveness for Virtual Agents – Autonomous Generation of Speech and Gesture for Spatial Description Tasks, In: *Proc. of the 8th International Conference on Autonomous Agents and Multiagent Systems* (AAMAS'09), 361–368.

Bergmann, Kirsten, Stefan Kopp and Friederike Eyssel
2010 Individualized Gesturing Outperforms Average Gesturing – Evaluating Gesture Production in Virtual Humans, In: *Proc. of the 10th Conference on Intelligent Virtual Agents,* 104–111. Berlin/Heidelberg: Springer.

Bernieri, Frank J., Robert Rosenthal
1991 Coordinated movement in human interaction. In: Robert S. Feldman and Bernard Rimé (eds.), *Fundamentals of Nonverbal Behavior,* 401–432. Cambridge University Press, New York.

Bickmore, Tim, and Justine Cassell
2005 Social dialogue with embodied conversational agents. In: J. van Kuppevel, L. Dybkjaer, and N. Bernsen (eds.), *Advances in Natural, Multimodal Dialogue Systems,* 23–54. Kluwer Academic, New York.

Bolt, Richard A.
1980 "Put-that-there": Voice and gesture at the graphics interface. *Computer Graphics,* 14(3): 262–270.

Branigan, Holly P., Martin J. Pickering, Jamie Pearson and Janet F. McLean
2010 Linguistic alignment between humans and computers. *Journal of Pragmatics*, 42: 2355–2368.

Burger, John D. and Ralph J. Marshall
1993 The application of natural language models to intelligent multimedia. In: M. T. Maybury (ed.), *Intelligent multimedia interfaces*, 174–196. Cambridge: MIT Press.

Cassell, Justine, Matthew Stone and Hao Yan
2000 Coordination and Context-Dependence in the Generation of Embodied Conversation. In: *Proceedings of the International Natural Language Generation Conference*, 171–178.

Cassell, Justine, Joseph Sullivan, Scott Prevost and Elizabeth Churchill (eds.)
2000 *Embodied Conversational Agents.* Cambridge, MA: MIT Press.

Cassell, Justine, Hannes Vilhjálmsson and Timothy Bickmore
2001 BEAT: the Behavior Expression Animation Toolkit. In: *Proceedings of SIGGRAPH '01*, 477–486.

Cassell, Justine and Timothy Bickmore
2001 A relational agent: a model and implementation of building user trust. In: *Proceedings of CHI'01*, 396–403.

Chai, Joyce Y., Pengyu Hong and Michelle X. Zhou
2004 A probabilistic approach to reference resolution in multimodal user inter-

faces. In: *Proceedings of the 9th International Conference on Intelligent User Interfaces* (IUI), 70–77.

Chartrand, Tanya L. and John A. Bargh
1999 The chameleon effect: the perception-behavior link and social interaction. *Journal of Personality and Social Psychology,* 76(6), 893–910.

Corradini, Andrea
2002 Real-time gesture recognition by means of hybrid recognizers. In: Ipke Wachsmuth and Timo Sowa (eds.), *Gesture and sign language in human-computer interaction*, 34–46. Berlin: Springer.

Duncan, Starkey Jr.
1972 Some Signals and Rules for Taking Speaking Turns in Conversations, in: *Journal of Personality and Social Psychology* 23/2, 283–292.

Friedman, Doron, Anthony Steed, Mel Slater
2007 Spatial Social Behavior in Second Life. In: *Proceedings of the Conference on Intelligent Virtual Agents (IVA'07)*, 252–263. LNAI 4722, Berlin/Heidelberg: Springer.

Fröhlich, Martin and Ipke Wachsmuth
1998 Gesture recognition of the upper limbs – from signal to symbol. In: Ipke Wachsmuth and Martin Fröhlich (eds.), *Gesture and Sign Language in Human-Computer Interaction*, 173–184, LNAI 1371, Berlin/Heidelberg: Springer.

Goffman, Erving
1983 The Interaction Order. *American Sociological Review*, 48, 1–17.

Gratch, Jonathan, Anna Okhmatovskaia, Francois Lamothe, Stacy Marsella, Mathieu Morales, R. van der Werf and Louis-Philippe Morency
2006 Virtual rapport. In: *Proceedings of IVA '06,* 14–27. LNAI 4133. Berlin/Heidelberg: Springer.

Hartmann, Björn, Maurizio Mancini and Catherine Pelachaud
2006 Implementing expressive gesture synthesis for embodied conversational agents. In: Sylvie Gibet, Nicolas Courty and Jean Francois Kamp (eds.), *Gesture in Human-Computer Interaction and Simulation*: 45–55. Berlin/Heidelberg: Springer.

Harling, Philip A. and Alistair D. Edwards
1997 Hand tension as a gesture segmentation cue. In: Philip A. Harling and Alistair D. Edwards (eds.), *Progress in gestural interaction: Proceedings of the Gesture Workshop '96*, 75–87. Berlin/Heidelberg: Springer.

Heylen, Dirk, Stefan Kopp, Stacy Marsella, Catherine Pelachaud and Hannes H. Vilhjalmsson
2008 The Next Step Towards a Functional Markup Language. In: *Proceedings of the COnference on Intelligent Virtual Agents (IVA 2008)*, 270–280. LNAI 5208. Berlin/Heidelberg: Springer.

Hofmann, Frank G., Peter Heyer and Günter Hommel
1998 Velocity profile based recognition of dynamic gestures with discrete Hidden Markov Models. In: Ipke Wachsmuth and Martin Fröhlich (eds.), *Gesture and sign language in human-computer interaction,* 81–95. Berlin/Heidelberg: Springer.

Howell, A. Jonathan and Hilary Buxton
1999 Gesture recognition for visually mediated interaction. In: Annelies Braffort, Rachid Gherbi, Silvie Gibet, James Richardson and Daniel Teil (eds.), *Gesture-based communication in human-computer interaction,* 141–152. Berlin/Heidelberg: Springer.

Jurafsky, Daniel and James H. Martin
2009 *Speech and Language Processing: an Introduction to Natural Language Processing*, Computational Linguistics, and Speech Recognition (2nd Edition). Upper Saddle River: Prentice Hall.

Kristiina Jokinen and Michael McTear
2009 *Spoken Dialogue Systems* (Synthesis Lectures on Human Language Technologies). San Rafael, CA, USA: Morgan and Claypool Publishers.

Johnston, Michael and Srinivas Bangalore
2001 Finite-state methods for multimodal parsing and integration. In: *Proceedings of the ESSLLI Summer School on Logic, Language, and Information*, Helsinki, Finland, August 2001, 1–6.

Johnston, Michael, Philip R. Cohen, David McGee, Sharon L. Oviatt, James A. Pittman and Ira Smith
1997 Unification-based multimodal integration. In: *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics, Madrid*, 281–288.

Jonsdottir, Gudny Ragna, Kristinn R. Thórisson and Eric Nivel
2008 Learning Smooth, Human-Like Turn-taking in Realtime Dialogue. In: *Proceedings of the Conference on Intelligent Virtual Agents (IVA 2008)*, 162–175. Berlin/Heidelberg: Springer.

Kaiser, Ed, Alex Olwal, David McGee, Hrvoje Benko, Andrea Corradini, Xiaoguang Li, Phil Cohen and Steven Feiner
2003 Mutual disambiguation of 3D multimodal interaction in augmented and virtual reality. In: *Proceedings of the 5th International Conference on Multimodal Interfaces (ICMI), Vancouver (BC), Canada*, 12–19.

Koons, David B., Carton J. Sparrell and Kristinn R. Thorisson
1993 Integrating simultaneous input from speech, gaze and hand gestures. In: Mark T. Maybury (ed.), *Intelligent multimedia interfaces*, 257–276. Cambridge, MA: MIT Press.

Kopp, Stefan, Jens Allwood, Elisabeth Ahlsen, Karl Grammer and Thorsten Stocksmeier
2008a Modeling Embodied Feedback in a Virtual Human. In: Ipke Wachsmuth and Günther Knoblich (eds.), *Modeling Communication With Robots And Virtual Humans*, 18–37. LNCS 4930. Berlin/Heidelberg: Springer.

Kopp, Stefan, Kirsten Bergmann and Ipke Wachsmuth
2008 Multimodal Communication from Multimodal Thinking – Towards an Integrated Model of Speech and Gesture Production. *Semantic Computing 2(1)*, 115–136.

Kopp, Stefan
2010 Social Resonance and Embodied Coordination in Face-to-Face Conversation with Artificial Interlocutors. *Speech Communication 52*, 587–597.

Kopp Stefan, Brigitte Krenn, Stacy Marsella, Andrew Marshall, Catherine Pelachaud, Hannes Pirker, Kristinn Thorisson and Hannes Högni Vilhjalmsson, H.
2006 Towards a Common Framework for Multimodal Generation: The Behavior Markup Language. In: *Proceedings of the Conference on Intelligent Virtual Agents (*IVA 2006), 205–217. LNAI 4133. Berlin/Heidelberg: Springer.

Kopp Stefan, Paul Tepper and Justine Cassell
2004 Towards Integrated Microplanning of Language and Iconic Gesture for Multimodal Output. In: *Proceedings of the International Conference on Multimodal Interfaces* (ICMI'04), 97–104.

Kopp, Stefan and Ipke Wachsmuth
2004 Synthesizing Multimodal Utterances for Conversational Agents. *Computer Animation and Virtual Worlds 15(1)*, 39–52.

Lakin, Jessica, Valerie Jefferis, Clara Cheng, Tanya Chartrand
2003 The chameleon effect as social glue: evidence for the evolutionary significance of nonconscious mimicry. *Nonverbal Behavior 27 (3)*, 145–162.

Lascarides, Alex and Matthew Stone
2009 A Formal Semantic Analysis of Gesture, *Journal of Semantics 26(4)*, 393–449.

Latoschik, Marc. E.
2001a A gesture processing framework for multimodal interaction in virtual reality. In: *Proceedings of the 1st International Conference on Computer Graphics, Virtual Reality and Visualization in Africa* (Afrigraph 2001), 95–100.

Latoschik, Marc E.
2001b A general framework for multimodal interaction in virtual reality systems: PrOSA. In: Wolfgang Broll and Léonie Schäfer (Editors): The Future of VR and AR Interfaces – Multimodal, Humanoid, Adaptive and Intelligent. In: *Proceedings of the Workshop at IEEE Virtual Reality 2001, Yokohama, Japan*, 21–25. Sankt Augustin: GMD Report No. 138.

Latoschik, Marc E.
2002 Designing transition networks for multimodal VR interactions using a markup language. In: *Proceedings of the 4th IEEE International Conference on Multimodal Interfaces (ICMI 2002), October 2002, Pittsburgh*, 411–416. IEEE Computer Society.

Leßmann, Nadine, Alfred Kranstedt and Ipke Wachsmuth
2004 Towards a cognitively motivated processing of turn-taking signals for the embodied conversational agent Max. In: *Proceedings of the AAMAS'04 Workshop Embodied Conversational Agents: Balanced Perception and Action*, 57–64.

Lessmann, Nadine, Stefan Kopp and Ipke Wachsmuth
2006 Situated interaction with a virtual human – perception, action, and cognition. In: Gert Rickheit and Ipke Wachsmuth (eds.), *Situated Communication*, 287–323. Berlin: Mouton de Gruyter.

Lücking, Andy and Thies Pfeiffer
2012 Framing multi-modal technical communication. In: Alexander Mehler, Laurent Romary and Dafydd Gibbon (Eds.): *Handbook of Technical Communication*. Berlin/Boston: De Gruyter.

Mann, William C. and Sandra A. Thompson
1988 Rhetorical Structure Theory: Toward a functional theory of text organization. *Text 8(3)*, 243–281.
Martin, Jean-Claude and Tanja Schultz
2012 Multimodal and speech technology. In: Alexander Mehler, Laurent Romary and Dafydd Gibbon (Eds.): *Handbook of Technical Communication*. Berlin/Boston: De Gruyter.
McNeill, David
1992 *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago Press.
Mehrabian, Albert and Susan R. Ferris
1967 Inference of attitudes from non-verbal communication in two channels. *Journal of Consulting Psychology 31*, 248–252.
Müller, Jochen, Peter Poller, and Valentin Tschernomas
2003 A Multimodal Fission Approach with a Presentation Agent in the Dialog System SmartKom. In: *Proceedings of KI 2003: Advances in Artificial Intelligence*, 633–645. LNCS 2821. Berlin/Heidelberg: Springer.
Neal, Jeannette G. and Stuart C. Shapiro
1991 Intelligent multi-media interface technology. In J. W. Sullivan and S. W. Tyler (eds.), *Intelligent user interfaces*, 11–43. New York: ACM Press.
Neff, Michael, Michael Kipp, Irene Albrecht and Hans-Peter Seidel
2008 Gesture modeling and animation based on a probabilistic re-creation of speaker style. *ACM Transactions on Graphics 27(1)*:1–24.
Negrotti, Massimo
2005 Humans and naturoids: from use to partnerships. In: Massimo Negrotti (ed.), *Yearbook of the Artificial Vol. 3*, Cultural Dimensions of the User, 9–15. Bern: Peter Lang.
Oshita, Masaki
2006 Motion-capture-based avatar control framework in third-person view virtual environments. In: *Proceedings of ACE'06 (International Conference on Advances in computer entertainment technology)*, New York, NY, USA: ACM Press, p. 2.
Oviatt, Sharon
2003 Multimodal interfaces. In: Julie A. Jacko and Andrew Sears (eds.), *The human-computer interaction handbook*, 286–304. Mahwah, NJ: Lawrence Erlbaum.
Picard, Rosalind W.
1997 *Affective Computing*. The MIT Press, Cambridge, MA.
Pickering, Martin J. and Simon Garrod
2004 Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences 27*, 169–226.
Poggi, Isabella and Catherine Pelachaud
2000 Performative Facial Expressions in Animated Faces. In: Cassell et al. (eds.), *Embodied Conversational Agents*, 155–188. Cambridge: MIT Press.
Poggi, Isabella, Nicoletta Pezzato and Catherine Pelachaud
2002 Signals and meanings of gaze in animated faces. In: P. McKevitt, S. O'Nuallain & C. Mulvihill (eds.): *Language, Vision and Music*, 271–284, Amsterdam: Benjamins.

Potamianos, Gerasimos, Chalapathy Neti, Guillaume Gravier, Ashutosh Garg and Andrew W. Senior, A.W
2003 Recent advances in the automatic recognition of audiovisual speech. *Proceedings of the IEEE 91*(9), 1306–1326.

Prillwitz, Sigmund, Regina Leven, Heiko Zienert, Thomas Hanke and Jan Henning
1989 HamNoSys version 2.0: Hamburg Notation System for Sign Languages: *An introductory guide (Vol. 5).* Hamburg: Signum Press.

Reeves, Byron and Clifford Nass
1996 *The media equation: How people treat computers, television, and new media like real people and places.* New York: Cambridge University Press.

Reiter Ehud and Robert Dale
2000) *Building Natural-Language Generation Systems*. Cambridge University Press.

Ridings, Catherine M. and David Gefen
2006 Virtual Community Attraction: Why People Hang Out Online. *Journal of Computer-Mediated Communication* 10(1). Online available at http://jcmc.indiana.edu/vol10/issue1/ridings_gefen.html, retrieved on May 31, 2011.

Rigoll, Gerhard, Andreas Kosmala and Stefan Eickeler
1998 High performance real-time gesture recognition using Hidden Markov Models. In: Ipke Wachsmuth and Martin Fröhlich (eds.), *Gesture and sign language in human-computer interaction*, 69–80. Berlin/Heidelberg: Springer.

Ruttkay, Zsófia
2007 Presenting in style by virtual humans. In: Anna Esposito (ed.), *Verbal and Nonverbal Communication Behaviours,* 23–36. Berlin/Heidelberg: Springer.

Schröder, Marc, Elisabetta Bevacqua, Roddy Cowie, Florian Eyben, Hatice Gunes, Dirk Heylen, Mark ter Maat, Gary McKeown, Sathish Pammi, Maja Pantic, Catherine Pelachaud, Björn Schuller, Etienne de Sevin, Michel Valstar, and Martin Wöllmer
2008 Building Autonomous Sensitive Artificial Listeners, IEEE Transactions of Affective Computing, pp. 134–146, October 2011.

Sowa, Timo
2006 *Understanding Coverbal Iconic Gestures in Shape Descriptions*. Berlin: AKA/Infix. (Doctoral dissertation, DISKI series, Vol. 294.)

Sowa, Timo and Ipke Wachsmuth
2009 A computational model for the representation and processing of shape in coverbal iconic gestures. In: Kenny R. Coventry, Thora Tenbrink and John A. Bateman (eds.), *Spatial Language and Dialogue*, 132–146. Oxford: Oxford University Press.

Sparrell, Carlton J. and David B. Koons
1994 Interpretation of coverbal depictive gestures. In: *Proceedings of Intelligent Multi-Modal Multi-Media Interface Systems,* 8–12. AAAI Spring Symposium Series, Stanford University.

Srihari, Rohini K.
1994 Computational models for integrating linguistic and visual information: A survey. *Artificial Intelligence Review 8*, 349–369.

Stone, Matthew, Doug DeCarlo, Insuk Oh, Christian Rodriguez, Adrian Stere, Alyssa Lees, and Chris Bregler
2004 Speaking with hands: Creating animated conversational characters from recordings of human performance. In: *Proceedings of SIGGRAPH* '04, 506–513.

Stone, Matthew, Christine Doran, Bonnie Webber, Tonia Bleam, Martha Palmer
2001 Microplanning with communicative intentions: the SPUD system. *Computational Intelligence 19*, 311–381.

Stührenberg, Maik
2012 Foundations of markup languages. In: Alexander Mehler, Laurent Romary and Dafydd Gibbon (Eds.): *Handbook of Technical Communication*. Berlin/Boston: De Gruyter.

Van Deemter, Kees, Emiel Krahmer and Mariet Theune
2005 Real versus template-based Natural Language Generation: a false opposition? *Computational Linguistics* 31(1), 15–24.

Vilhjálmsson, Hannes H. and Justine Cassell
1998 BodyChat: autonomous communicative behaviors in avatars. In: *Proceedings of the 2nd Annual ACM International Conference on Autonomous Agents*, 269–276. Minneapolis, 1998

Von der Pütten, Astrid, Christian Reipen, Antje Wiedmann, Stefan Kopp and Nicole Krämer
2008 Comparing emotional vs. envelope feedback for ECAs. In: *Proceedings of the Conference Intelligent Virtual Agents (IVA 2008)*, 550–551. LNAI 5208. Berlin/Heidelberg: Springer.

Wachsmuth, Ipke
1999 Communicative rhythm in gesture and speech. In: Annelies Braffort, Rachid Gherbi, Sylvie Gibet, James Richardson and Daniel Teil (eds.), *Gesture-based communication in human-computer interaction,* 277–290. Berlin/Heidelberg: Springer.

Wachsmuth Ipke, Manuela Lenzen and Günther Knoblich
2008 *Embodied Communication in Humans and Machines.* Oxford: Oxford University Press.

Wahlster, Wolfgang
1991 User and discourse models for multimodal communication. In: Joseph W. Sullivan and Sherman W. Tyler (eds.), *Intelligent user interfaces*, 45–67. New York: ACM Press.

Waibel, Alex, Minh Tue Vo, Paul Duchnowski and Stefan Manke
1996 Multimodal interfaces. *Artificial Intelligence Review*, 10, 299–319.

Wallbott, Harald G.
1994 Verhaltensbeobachtung. In: Rolf-Dieter Stieglitz and Urs Baumann, *Psychodiagnostik psychischer Störungen,* 95–106. Serie: Klinische Psychologie und Psychopathologie, Bd. 60. Enke: Stuttgart.

Weitnauer, Erik, Nick Thomas, Felix Rabe and Stefan Kopp
2008 Intelligent agents living in social virtual environments – bringing Max into Second Life. In: *Proceedings of the Conference on Intelligent Virtual Agents (IVA08)*, 552–553. LNAI 5208. Berlin/Heidelberg: Springer.

Wexelblat, Alan D.
1995 An approach to natural gesture in virtual environments. *ACM Transactions on Computer-Human Interaction 2(3)*, 179–200.

Wu, Ying and Thomas S. Huang
1999 Vision-based gesture recognition: A review. In: Braffort, Annelies, Rachid Gherbi, Sylvie Gibet, James Richardson and Daniel Teil (eds.), *Gesture-based communication in human-computer interaction*, 103–115. Berlin/Heidelberg: Springer

Young, Steve
1996 A review of large-vocabulary continuous-speech recognition. *IEEE Signal Processing Magazine,* 45–57.

Yngve, Victor H.
1970 On getting a word in edgewise. In: *Papers from the 6th Regional Meeting of the Chicago Linguistics Society, University of Chicago*, 567–578.

21. Ubiquitous computing

Codrina Lauth, Bettina Berendt, Bastian Pfleging, and Albrecht Schmidt

1. Motivation

Looking at the tremendous technological advancements in the new ways of handling and providing information, we see our society in the middle of a new Era of computation – the *Ubiquitous Computing Era* also called the *3rd Era of Computation*. To this new Era, we intend to dedicate the present chapter. Ubiquitous Computing (in short: ubicomp), as it has been coined by the European research community back in the 1990ies, has clearly initiated the *Post-Desktop Era* of Human-Computer Interaction (HCI) and is still one of the most progressive areas of applied computer science. The technical advances in wireless sensor networks and hardware implementations on the one hand and the rise of a new generation of the Internet (Web 2.0) on the other hand have commonly contributed to the evolution of the *omnipresent* (or ubiquitous) technologies. The ubiquitous user is someone who is connected to all (computing) environments through his/her ubiquitous devices from everywhere and at any time. Ubiquitous or Pervasive Computing is the breakthrough and the response to increasing information needs and consequently a new stage of *user-centred* information delivery and processing. Advanced Internet technologies embedded in mobile and wireless computer devices have already become an integral part of our daily life. The mobile devices like personal digital assistants, smart phones, and tablets include nowadays high-end digital technologies like, e.g., cameras, Internet access, GPS, and gyroscopes in one single device and can easily connect users to websites of interest, social networks, and virtual worlds. At the same time, the TV and broadcasting entertainment industry is offering more complex TV sets that allow Internet access and online digital media entertainment from any location. Especially the younger generations are keen on using these ubiquitous devices without much thinking about it. Thus, the technological development has gone even further. Hundreds of tiny processors and sensors are embedded into our environments and objects that we use every day – household appliances, toys, tools, but also such mundane things as pencils, furniture, and clothes. All these devices can communicate or at least send some kind of information to one another over wireless networks. People are now contributing to the creation of a new Internet, called the "*Internet of Things*" (Botthof 2009, Bullinger 2007). The ubiquitous services and products augment the physical world by new ways of interaction and communication. *Content,* once sent to the Internet, is getting out of the user's control, it is somehow "living" by being re-used and "syndicated"[1], being available in multiple forms like RSS feeds, blogs, wikis, podcasts, geospatial reference, user-generated content like collaborative tagging[2] and instant messaging, MMS, etc. At the same time, ubiquitous technology and user trends pose new challenges on re-defining and modelling *context* that is the essential term for modelling the new forms of physical melting with virtual ways of "ubiquitous interaction". Last but not least, giving an eye to the development of *Tangible User Interfaces* (TUI, Ishii and Ullmer 1997) and more recently *Natural User Interfaces* (NUI), and how haptic or organic computing developments have contributed on designing multi-touch technologies with computationally enhanced surfaces for recognizing objects, all these technical achievements are bringing us to identify new forms of ubiquitous communication marking the Ubiquitous Service-Centric Information Society[3].

After this short motivation (Section 1), we illustrate the beginnings and the classical definition of ubiquitous computing (Section 2.1), and we distinguish it from its neighbouring area called Ambient Intelligence (Section 2.2). In Section 2.3 we discuss current technological, as well as social trends highlighting the *ubiquitous user trends*, since they are rooted in the changed user behaviour using ubiquitous devices and systems. In Section 2.4 and in Section 2.5, we give examples of ubiquitous computing projects and applications. The ubicomp area itself is too broad to cover the whole spectrum of sub areas in one single chapter. Therefore, we will concentrate on a selection of topics of interest that we will illustrate more from the *technological* side (Section 3.1), and from the *methodological* side (Section 3.2), as well as with regard to their *impact on society* (Section 3.3). Section 4 deals with a brief overview of current ubicomp research challenges and in the last part of our chapter (Section 5) we will summarize our ideas and connect the ubicomp perspective with the overall perspective of the handbook by giving potentials and challenges of ubiquitous computing applications for Technical Communication in general.

2. Introduction to ubiquitous computing

2.1. What is ubiquitous computing?

The term ubiquitous computing originates from work at XEROX Parc in Palo Alto, in the late 80ies and early 90ies (Weiser 1991 and Weiser 1999). The initial idea has been formulated by Marc Weiser in his visionary paper from 1991: "*… the most profound technologies are those that disappear. They weave themselves into the fabric of everyday life until they are indistinguishable from it"* (Weiser 1991). Current definitions lean on this idea of making technology invis-

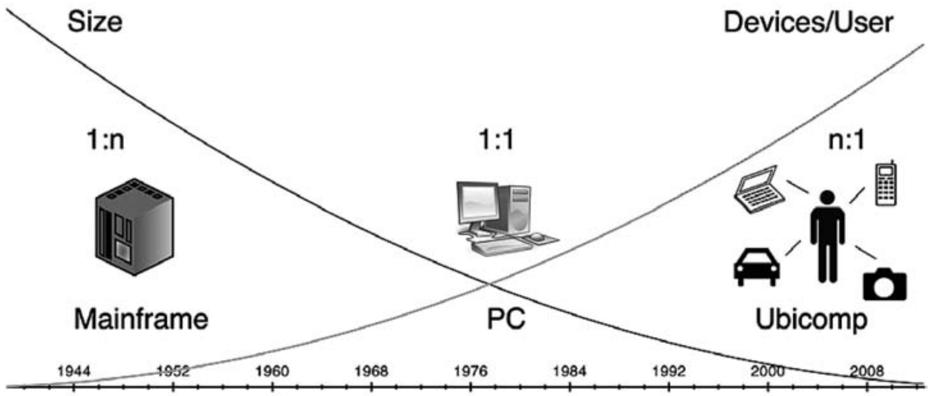


Figure 1. Ubiquitous computing seen as the quantitative relationship between computers and users. First, many users have used a mainframe computer, later every user had access to a personal computer. Today one user may use many devices at the same time[4].

ible and easy-to-use, so that not anymore the handling (*human-to-computer interfaces*), but just the use of the interconnecting technology embedded in the device (*human-to-human interfaces*) becomes the main centre of interest (Weiser 1999):

> "Pervasive [or ubiquitous] computing describes the trend that inter-connected computational devices become interwoven with artefacts in our everyday life. Hence, processing, sensing, activation and communication are embedded into devices and environments, making computing an integral part of our life." (Schmidt 2008)

The larger number of devices used often in ubicomp scenarios has led to the simplification as shown in Figure 1, that associates ubiquitous computing with the 3rd Era of Computing: The 1st Era of Computing was the mainframe computing era (*many users share one computer*), while the 2nd Era has been marked by using desktops and personal computers (PCs) (*one computer per person*). The 3rd era is now the Ubiquitous Computing Era (*one person uses multiple computers/devices with a multitude of functionalities at the same time*), which is marking the cutting edge of the Ubiquitous Information-Centred Service Society (Mannermaa 2007). 20 years after the publication of Weiser's visions in 1991, we see how far Weiser's ideas have influenced computing and human-computer interaction. Although a lot of his visions have found their way into real life, some research challenges still remain and further ubiquitous technologies arise (Schmidt 2011).

Research in ubiquitous computing became more popular by the end of the 1990ies. In 1999 the first international conference on *Handheld and Ubiquitous Computing* was initiated, which later became the UBICOMP conference series.

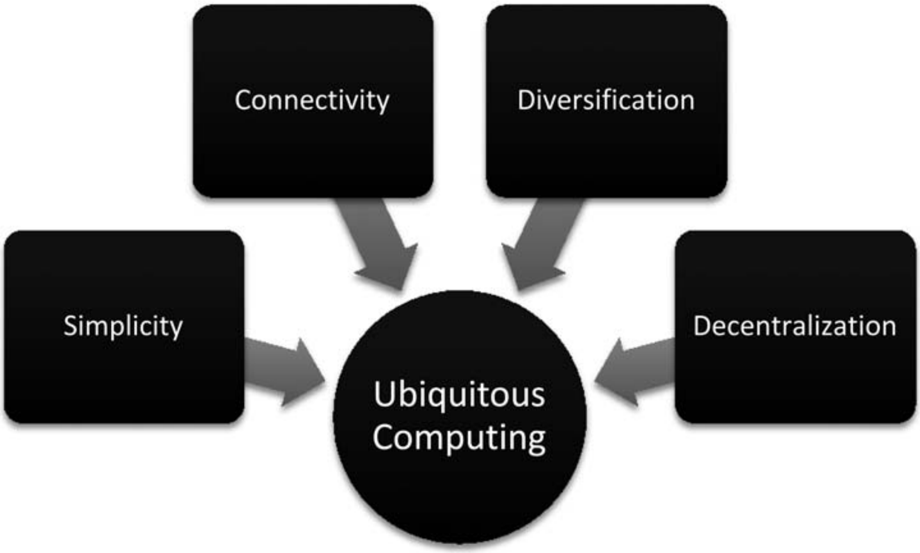


Figure 2. Paradigms in *Ubiquitous Computing* from (Diekmann 2003).

The UBICOMP research community has grown massively since then, and there are now a number of distinguished international conferences focusing on slightly different angles of the subject, most notably UBICOMP, PERVASIVE and PerComp. Since 2003, the IEEE Magazine on Pervasive Computing and since 1997 the Personal and Ubiquitous Computing Journal (Springer) have been published.

The four paradigms that are used in ubiquitous computing are shown in Figure 2.

2.2. *Relationship to neighboring communities*

Ubiquitous computing is a highly interdisciplinary area that spans across several different areas of computer science, material science, and electronics. Even fields like ethnography, sociology, ethics, and design are examples of disciplines that have a tight relationship with ubicomp applications and technologies. During the last decade, a significant upturn of new and more complex ubiquitous computing applications has been noticed, as well as research going far beyond the original *hardware-centred approach* that has initiated the ubiquitous computing area as such. Especially with the rise of social networks and the proliferation of social network software to mobile devices and everyday artefacts, ubiquitous computing is reaching end users in large numbers. Nowadays, ubiquitous computing research and applications are merging with applications from neighbouring communities like, e.g., the *Ambient Intelligence* (AmI[5]) and

Ubiquitous Knowledge Discovery (KDubiq[6]) among others, even if they all build on different technological foundations and research traditions. The European Community supports numerous collaborations, projects and exchanges between these inter-related communities[7]. Even if, from its beginnings, AmI has pursued the more *information processing-oriented approach* (Acampora 2008), whereas ubiquitous computing has been seen as the more *hardware-centred approach*, both communities cannot deny the close relatedness of applied technologies and applications, as well as the blurring distinction between them (Schmidt 2007):

> Ambient Intelligence is considered as the composition of three emergent technologies: Ubiquitous computing, ubiquitous communication and intelligent user interfaces. The aim of integration of aforesaid technologies is to make wider the interaction between human beings and information technology equipment through the usage of an invisible network of ubiquitous computing devices composing dynamic computational-ecosystems capable of satisfying the users' requirements. (Cook 2009)

2.3. Current trends in ubiquitous computing

Technological trends in several areas of advanced computing, such as mobile and distributed computing, wireless sensor networks, wearable computing, smart homes/living, as well as intelligent environmental monitoring become enabling technologies and building blocks for ubicomp. The development and broad use of advanced ubiquitous systems is initiating a lot of radical changes in the interaction and behaviour of users, and consequently goes along with changes in social and communication structures.

In the following section, we illustrate the technological trends that are giving impulses to the rapid advancements of ubiquitous computing, and vice versa the social trends that emerge the more our society will use ubiquitous technologies and devices (further references cf. Schmidt 2009, May 2010 and Schmidt 2011).

2.3.1. Ubiquitous technology trends

Trend (1): Technology becomes widely available – Digital technologies that are very expensive today and used for special purposes, can quickly become part of everyday devices, since technological advancements are faster than ever and their developments and use lifecycles are becoming shorter. Two examples that have gone through this transition are GPS receivers and digital cameras. This trend predicts that many technologies that are now shared and expensive will, due to rapid developments in digital systems, become personal commodities soon. Ubiquitous technologies that were expensive at some point may not even be considered as additional cost in the future. Technologies that are at present still expensive, but have the potential to become ubiquitous, are: physiological sensing, biometric authentication, book printing on demand, eye gaze tracking, 3D interaction, scanning, and printing, and integrated production systems. For 3D interaction, for instance, we see with Microsoft Kinect[8] the first steps in this direction.

Trend (2): Computing, storage and communication are not the limit – Computing power and storage space have massively increased over the last 30 years. Moore's Law – which is still valid – is a driving force behind this. Already in current desktop systems, we see that processing power and storage space do not limit what we can do with these computers. A typical office computer is running – even if heavily used with text processing and spreadsheet calculation – mostly idle. Adding a fast processor will not be likely to increase productivity, since the limitations lie in the capabilities of the software. We expect consequently that massive computing power, large storage capability and huge bandwidth with high-speed communication will become ubiquitously available. It is likely that in future ubiquitous computing systems, limitations will not be due to technological limitations. There may be limitations due to wireless data transmission and due to power consumption; however, we expect that the rapid technical advancements in ubiquitous computing and related disciplines will lead to a different angle on system development where creativity may become the bottleneck and not technology.

Trend (3): Mobile communication is ubiquitous – Terminals for mobile communication have advanced significantly over recent years. In one mobile device we now have several complex functionalities like GPS, camera, video recording, Internet, Bluetooth connection, etc. The whole infrastructure is nowadays ubiquitously deployed. It is a clear trend that mobile devices are networked and in many regions of the world users just assume that there is network coverage wherever they go. Within this trend we see that also very small components become networked, e.g. the Eye-Fi SD-card[9] is a memory card that includes a communication interface and acts as storage device, but transparently transmits the data to the backend. A further development is WhisperNet on the Amazon Kindle[10] device that provides users with transparent connectivity in more than 150 countries. Extrapolating from these two products, one can imagine a world in which all electronic devices are seamlessly and ubiquitously connected.

Trend (4): Mechanical and electro-mechanical systems are going to be computer-controlled – In all branches of industry we observe that production chains for all mechanical and electro-mechanical systems become more and more computer-controlled. User interfaces for mechanical and electro-mechanical systems have a tradition of being tangible. Often this tangibility was a direct result of the mechanical design, e.g. a lever to change a setting in pump. In these times many of the design restrictions were due to the mechanical design. In new computer-controlled actuators, these restrictions are not anymore relevant.

However, the technological trends are going here in the direction of new advanced interface options using complex processing of tasks like e.g. sensing of actions, haptics that draw on the physical interaction used for making systems more understandable (for more information on multimodal communication see Lücking and Pfeiffer 2012 in this volume).

Trend (5): Functionality is non-discriminative – Ubiquitous computing products and systems embedded in mobile devices, phones, mp3 players or TV sets have more or less similar features and functionalities, (e.g. same decoder chip, same display panel),whereas aspects like design, experience, user interface have become the relevant distinctive factors for quality differences between devices and products. This trend is accelerating as certain functionalities are "maxing out". For example, for a device that can record and play back audio with a resolution of 8bit and 12kHz doubling resolution and bandwidth (e.g. 16bit and 24kHz) will improve the device massively, but increasing the capacity of the device to 24bit and 96kHz will not make any difference for the users, because they are not able to perceive the difference. We see this "maxing out" effect currently with audio, visual screens and digital cameras.

Trend (6): Ubiquity lies in the process – Ubiquitous computing technologies allow us to change the way we organize any kind of business, communication and consequently innovation processes – in factory, as well as in social settings. In many cases, optimizing a single phase within the process chain may not make any difference, whereas optimizing the whole process from a ubiquitous perspective may bring new business innovations, totally new communication channels – and like a chain reaction – also potentials for new technological opportunities and developments. We have to look at the process chain as a whole, in order to understand the change in the process itself. For example ubiquitous RFID[11] technology in logistics has revolutionized the way manufacturing and delivery are organized nowadays, and also ubiquitous communication forms over smart phones, and multimedia mobile devices, etc. using social networking platforms like Facebook, Twitter and LinkedIn have changed radically the ways people are interacting today. Here the effect of change is not due to a single technology, it is rather a re-consideration of the entire business and communication process value chain with regard to the new technologies available.

2.3.2. Ubiquitous user trends

Trend (7): Declining willingness for training – Ubiquitous technologies rely on advanced user interfaces that should be intuitive and allow "*natural interaction*"[12] with minimal learning effects. Looking at the way people use smart phones in everyday life, we see that an average person today acts as telephonist, photographer, film-maker, and typesetter without spending much time for training. Many of these functions now included in a mobile phone required, only 50 years ago, long-term training. This is also important from an economic perspective, as a reduced need for training will increase the flexibility of deploying the work force.

Trend (8): Ubiquitous technologies are enabling new ways of education and training through social interaction – Driven by the tremendous technological and social developments, our society has turned into an "*always learning*" or Ubiquitous User-Centred Knowledge Society. Ubiquitous technologies are the driving technologies for the "*everytime-everywhere-learning*" experience, given the learning, collaborative and information-rich applications now available over mobile and wireless devices. Furthermore, the technological trends in ubiquitous learning are going in the directions of learning in virtual labs, as well as learning embedded in everyday life over intuitive tangible interfaces. Ubiquitous networking and information providing technologies may support many aspects of natural learning better than traditional knowledge technologies have done so far. Information can now be acquired in many ways, e.g. learners may piece together and connect various content and conversation elements to create an integrated (even if at times contradictory) network of information. Our learning and information acquisition becomes a *mashup*. We take pieces, add pieces, dialogue, reframe, rethink, connect, and ultimately, we end up with some type of pattern that symbolizes what's happening "out there" and what it means to us. New patterns in communication procedures and communities of knowledge are evolving and changing daily, social networking technologies and new interaction and knowledge sharing habits have evolved so rapidly over the last few years – to the point that the loose collection of many resources is often seen and even more and more used as an alternative to a *Learning Management System* (LMS) that has been the dominant technology for many years in eLearning. Through the use of Google Docs, Skype, blogs, wikis, podcasts, Flickr, YouTube, LinkedIn, del.icio.us and other tools, academics can provide a rich learning experience often exceeding the static experience of an LMS (see also Diewald 2012 as well as Waltinger and Breuing 2012 all in this volume).

Trend (9): Perception of privacy changes – A very important question in the ubiquitous society is how we will deal with personal information in a highly interconnected world. We see a clear trend that the way people protect and share their personal information is changing. In social networks, such as Facebook, we can observe a high degree of sharing in order to become connected within a group of friends. The absolute value of privacy as previously suggested is likely to be replaced by a model where personal information becomes a commodity that people trade. Most people are willing to make their private information available (publicly, to corporations, to authorities, or to service providers) if they see a benefit for themselves. The perception what is private information and what value in return is appropriate is very different between people, for example some people find sharing their location in order to have their friends know where they are convenient, whereas others would never want to share this information.

2.4. Example 1: Smart environments

Smart Rooms belong to the sub field of Intelligent or smart environments within ubiquitous computing. The components of smart rooms applications are interwoven and integrated into an augmented spacious environment (Montoro 2004), which is a distributed computing system able to perceive context information from many perspectives of the room environment through sensors and to execute intelligent logic on computing devices serving its occupants by actuators[13]. An early example of such scenarios for *situated interaction* is the EasyLiving research project (Brumitt 1998), that has been enhanced with a geometric context model by (Brumitt 2001). This area of ubiquitous applications belongs to the group of *location-aware computing services* and includes tasks like driving directions, redirecting phone calls or maintaining interactive sessions from one room into the other depending on user's movements, or simply turning off the lights when the user leaves the room. Looking at this example, we perceive technical communication in a ubiquitous, smart environment in its rather general sense of "*delivering information*" at the right time with the right content to the right device, all this way being patterned primarily by a user's context (including location and time, etc.), experiences and needs with regard to the specific environment. Ubiquitous computing applications and devices are enabling people and groups to exchange instantly information and collaborate or react on the fly to relevant information from their environments. The information is reaching the people independently of time and space, but smart environments should be able to react and send only the information that is relevant to the user for the present moment. There are several relevant and challenging research issues that need to be solved here, such as the interconnection of computing devices on many different scales[14], the handling of various mobility problems caused by user's movement, network protocols, software infrastructure, application substrates, user interfaces issues, etc. From their scenarios descriptions, the projects on Smart Environments seem quite heterogeneous, since they are focusing on different aspects of technological development, for details, see for example (Jiang 2000), and some target at facilitating the collaboration of multi-user multi-device within a technology-rich environment (Interactive Workspace 2000, Fox 2000) (for more information on multimodal computing see Martin and Schultz 2012 as well as Lücking and Pfeiffer 2012 both in this volume). The current Smart Environments application trends go far beyond smart multimodal and smart interfaces, the UBICOMP Smart Environments are now inter-connected over several worlds, like WLAN, LAN, Area network (cf. Vanhoof 2010). For example, the

House_n Project at MIT[15] investigates methods for creating new smart homes technologies and products with person-centred design (Intille 2006). Another example of large ubiquitous applications are the South Korean projects on building Ubiquitous Cities or U-Cities[16] and also the projects from the area called Pervasive Health (Bardram et al. 2006)[17].

2.5. Example 2: People-centric sensing

The second large area of ubiquitous applications is regarding the people-centric sensing and comes from the improved sensor technology, as well as from using every day thousands of sensors invisibly integrated in our environments. For instance sensors for activity recognition and GPS are now being shipped in millions of high-end mobile phones, enabling a myriad of new sensor-based applications for personal, social and public sensing. This complements other sensors already integrated on mobile phones such as high-quality cameras, microphones, and digital compasses, as well as sensors installed in urban environments in support of more classic environmental sensing applications, such as real-time feeds for air quality, pollutants, weather conditions, and traffic-congestion conditions. Collaborative data gathering of sensed data for people by people, facilitated by sensing systems comprised of everyday mobile phones and their interaction with static sensor webs, present a new frontier driven by people-centric sensing. Promising attempts in setting up the right infrastructure for this kind of ubiquitous applications is for example the MetroSense project (Campbell 2008) working with industry and agencies to develop new applications, classification techniques, privacy approaches, and sensing paradigms for mobile phones enabling a *global mobile sensor network* capable of societal-scale sensing. Several other projects, e.g. SoundSense, CenceMe, Sensor Sharing, BikeNet, AnonySense, and Second Life Sensorare part of this large project.

Another interesting example in this area is the project called *Emotional Cartography.*[18] In this project people's physiological and emotional reactions to places in an urban area are mapped. By having people wearing sensing equipment while walking in a town and aggregating this information, a map is created that shows a community's emotional response to its environment. Here the initial vision that ubiquitous technologies should melt the frontiers between virtual and physical worlds, seem to become reality. If we now imagine that in the near future such physiological sensors are integrated in garments and that a sufficient number of users share (potentially anonymously) their data, we would have real-time maps of towns that include the emotions they evoke.

3. A Selection of ubiquitous computing technologies

In this section, we intend to give a rough overview on technologies and methods used in ubiquitous computing. The ubicomp area is much too vast to cover the whole state-of-art in ubicomp in this chapter. For more detailed overviews, please check the following references: Burkhardt (2001), Chen (2008), and Krumm (2009). In the following subsections, we start from most simple devices in order to go to most complex technologies and methods.

3.1. Ubiquitous technologies

3.1.1. Sensors and perception

Sensors are the most essential devices used in ubicomp applications to "perceive" the physical environment, in other words, these tiny devices are generating the basic data input to the ubicomp applications by converting the physical parameters (e.g. temperature, blood pressure, humidity, speed, etc.) into a signal which can be measured electrically[19]. Consequently the notion of "perception"[20] implies here not only the process of continuous information gathering from sensors, but even more it constructs certain dimensions of the "context" in which the users and the ubiquitous devices or applications are interacting to each other. The main ingredients for the "smartness" of a ubiquitous system are the ways in which we are able to "sense" (or collect) data and "perceive" the context relevant for the user. Sensing data has become commonly integrated in devices (e.g. accelerometer, GPS, compass etc. in mobile phones). We have enough systems available to collect ("sense") this sensor data centrally or locally and interpret it. Nevertheless, major current challenges to be solved are e.g. ubiquitous knowledge discovery technologies embedding more "intelligence" in the device, especially if they have to deal with resource-constrained devices (cf. Vanhoof 2010), as well as transmitting data and certain parameters to different locations[21].

3.1.1.1. Sensing data – types of sensors

A big variety of sensors is available to capture information about light intensity, density, type of light (sunlight, artificial light, etc.), reflection, color temperature, and wavelength. Low cost and energy saving sensors comprise photo diodes, color sensors, infrared (IR)- and UV sensors and many more. They allow extracting information about the current light conditions as well as, e.g., movement within the scene. If more detailed information is required, camera chips (e.g., CCD chips or CMOS chips) are used. Depending on the level of image processing implemented, they provide a lot of different information. Very easy implementations calculate measurements like the prevailing color or illumination level or changes within subsequent images (for motion detection). More elaborate image processing methods might, e.g., be used to detect and distinguish different objects, faces, people or gestures. While the sensor used can be cheap, some situations require elaborate optics to fit to the scenario as well as advanced image processing options can require some computational power to extract the desired information. Camera chips differ mainly with respect to their size (image resolution) and sensitivity (black/white, color, IR, etc.). Infrared sensitive cameras could for example be used to detect different heat distributions within a scene. In order to not only extract information that is visible in traditional pictures, two different technologies can be used to extract even depth information from the scene. The first method to do this is to use a stereo camera setup which comprises two (typically horizontally displaced) calibrated cameras (stereo vision). By projecting one camera image to the position of the other camera, the parallax allows to see the relative shifts between both views. The second way to extract 3D information has become famous with the introduction of the Kinect[22] technology which is mainly based on the PrimeSensor[23] technology. Depth information is extracted by using an IR sensitive camera chip which records the scene that has been illuminated with a IR light source that projects a specific pattern onto the scene[24]. This technology is somehow a variant of using structured light[25] to extract depth maps. A different way is to use time-of-flight (TOF) cameras[26] which use an (IR) light source to illuminate the scene. In contrast, this technology uses the delay between emitting the light and receiving the reflected light to calculate a depth image. Furthermore embedded cameras in mobile devices are able to track the eye gaze of users (Jacob 1990, Bolt 1981). Eye gaze tracking is the sensing technology using the movement of the human eye to interact directly with a computer or another embedded device. Interaction is for example possible by using dwell-based activation (Jacob 1990) or specific eye gestures (Drewes 2007). Similar to camera chips, microphones can provide information at different processing levels. They can, e.g., monitor the acoustic environment and extract basic information about noise level or prevailing frequencies. More elaborate setups allow recognizing commands (speech recognition) or distinguishing speakers. By using an array of microphones, one can even use beam forming to extract the sound at a specific location within a room without having a dedicated microphone at this location or locate the source of a noise within a room (Burkhardt 2001). The advent of modern smart phones brought a lot of different sensors able to provide more information about the current context. Accelerometers provide information on inclination, motion, and acceleration of a device allowing inferring knowledge about orientation and movement of a device. The orientation can also be measured by a gyroscope, or magnetic field sensors can be used to provide the direction/movement of a device. If location information is necessary, different technologies provide the de-

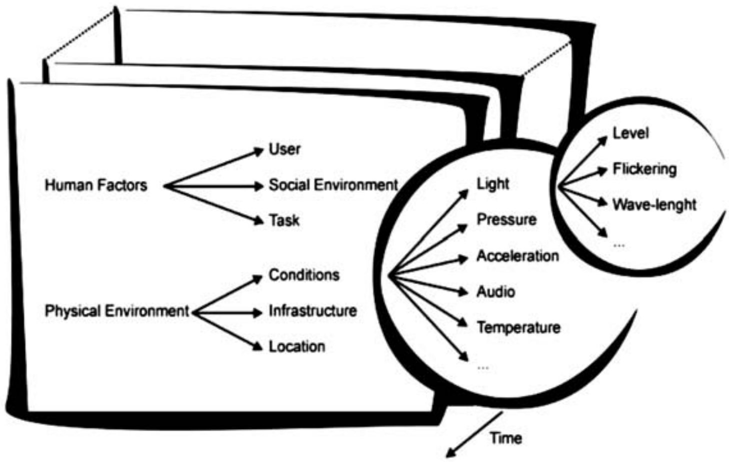


Figure 3. Elements of situational context are splitted in human factors and physical environmental factors (Schmidt 2000).

sired information. For outdoor/global uses cases, GPS, Galileo, or the GSM network can be used to infer the current location or collocation of users and devices. If used indoor/locally, technologies like Ubisense[27]and Optitrack[28] might be useful. A lot of interactive systems which require some user interaction also make use of touch technologies that go beyond traditional haptic sensors like keys. Different technologies exist to enable touch/trackpad and touch screens. Temperature-sensitive sensors might provide additional information about the current situation as well as biosensors (e.g., skin resistance, pulse, blood pressure, brain etc.), integrated into garment, provide detailed information about a user's condition.

3.1.1.2. Sensing context

The definition of context is indispensable for ubicomp, since interaction is embedded and happens only in context[29]. The notion of physical and virtual context is melting here. Content can be reused, syndicated and has different meanings depending under which context and by which user:

> Context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and application themselves … A system is context-aware if it uses context to provide relevant information and/or services to the user, where relevancy depends on the user's task. (Dey 2001)

The thematic shift towards "*implicit human-computer interaction*" (Schmidt 2000) plays an important role in defining ubiquitous computing applications, since it helps to elaborate how the availability of processing power and advanced sensing technology can enable a shift in HCI from *explicit interaction*, such as direct manipulation GUIs, towards a more implicit interaction based on *situational context* (Schmidt 2000).

3.1.2. Displays and actuation

If a system needs to provide feedback to its users, technologies can be used to stimulate all human senses to submit the intended information. If a detailed feature-rich feedback has to be transmitted, a lot of systems use the visual cue to convey this information. The information can be submitted by rather traditional displays (e.g., LCD, LED, or plasma screens), projectors or newer technologies like e-ink, electronic paper, or (AM)OLED displays. Loudspeakers may be used to provide acoustic feedback to a user. Vibro tactile feedback is another technology to transmit information. It is for example used, to notify the owner of a phone about an incoming call (the phone is vibrating) or in the automotive domain assistance systems use a vibrating steering wheel to notify the driver about other cars driving in the blind spot while changing lanes or if the driver tends to depart from the current lane. Olfactory displays[30] could provide another way to transmit information to the user (for more information on multimodal communication see Allwood and Ahlsén 2012 as well as Lücking and Pfeiffer 2012 both in this volume).

3.1.3. Smart objects, ubiquitous services and devices

The amount of smart objects, ubiquitous services and devices is increasing every day. Therefore, the following objects and services just represent a small part of what is nowadays available. Smart phones seem to be the most widespread and well-known smart devices. As it does not only contain loudspeaker and microphone, but also GPS receiver, accelerometer, gyroscope, digital compass, camera(s), NFC modules and light sensors, a lot of different applications and services can be provided with this platform. Gradually, also smart meters find their way into our homes. They do not only observe our power or water usage. Instead, they also predict the near future, anticipate the users' needs, and could even lead to behavior understanding. For example, just-in-time services like a daily or weekly payment interval could be introduced when using power or water. In "*natural reality*", objects are assigned semantic sense by humans interacting about them: person A tells person B that "this is a chair", which for example means that one can sit on it. When such objects have clear affordances, people will intuitively or, often more precisely, based on their previous experiences with other objects, perceive what they can do with the object (such as sit on it because it has a horizontal plane and support for the lower back). Creating such obvious affordances is a large part of design for usability. When objects are more complex and have a wide range of functionalities, the *intuitive approach* alone often is not sufficient – Technical Communication, usually in the form of an associated written manual, will complement what meets the eye and will need to be consulted to take full advantage of the object's potential.

There are several challenges associated with this approach:

- *Challenge 1: Media disruption* – the manual needs to be picked up consciously, it needs to be read (which requires basic literacy), and this is an activity not naturally linked to the actual use of the object.
- *Challenge 2: Static nature* – the manual was created at some point in time and does not change.
- *Challenge 3: No interaction* – the manual cannot respond to how the object is being used, including to features of its user. Since there is no interaction, there can also be no intelligence – a learning from how it is being used.
- *Challenge 4: Unilateral definitional power* – the manual was written by the vendor, thus specifying a limited set of uses and meanings.

An interesting and low-threshold alternative to the permanent association of an object with an ID (like the RFID object) and sensor that delivers information to a computing architecture is the option for an object to be associated permanently or transiently with an ID and to allow everyone to perform this association. This in some way restores the "natural" situation in which it is the people in the situation who define a chair to be a chair (or something else).

The Semapedia project[31] allowed people to "hyperlink their world" – in 3 simple steps. First, anyone who wants to hyperlink a physical object, place, etc., generates a QR code (a two-dimensional barcode) that encodes a URL (more specifically, a Wikipedia page), prints it out, and attaches it to the object. Other people can then use their mobile phone cameras plus free software to decode the image and to automatically request the respective Wikipedia page to their phone browser. This idea addresses Challenge 1 and Challenge 2 in essentially the same way that RFID tags do, and there is scope (though as of today no realization in Semapedia) for an informational combination with mobile-phone GPS and other data and services to "personalize" and "contextualize" what is being delivered, which would result in addressing Challenge 3 in a similar way as RFID does. Most interestingly, though, the Semapedia idea also addresses Challenge 4. Thus, for example, different people could attach different QR codes to the same object, thus assigning different meanings and associating with different information and action recommendations. The use of the collaboratively authored Wikipedia as the information source of choice is in itself a further democratizing facet of Semapedia. In sum, while Semapedia obviously faces a number of physical limitations – no one wants objects and places to disappear under a carpet of QR-code printouts, printouts can be removed, made unusable by weather or other events –, it addresses a number of challenges for (not only) technical communication in inspiring ways and is therefore a potential source of new ideas for designing information-for-use. Semapedia terminated operations, but the idea has been developed further in projects like QRPedia (qrpedia.org).

3.1.4. Local and global communication technologies

The ability of processors and sensors to wirelessly communicate with each other through different networks is a big pathfinder for new interconnected, ubiquitous applications and services. As the requirements concerning speed and data volume vary, different technologies are used to provide the intended connectivity. While some devices set up a local network to connect to other systems, another option is to connect devices through mobile communication networks and the Internet. In general, the amount of connected devices increases continuously which in turn requires a permanent development of fast and reliable communication modes. This section shall give an overview of current and upcoming technologies. One of the first (proprietary) standards that have been adopted widely to connect devices is Bluetooth[32].

Bluetooth was developed to provide a radio-based short-range (1m – 100m) communication network, which has low power consumption and directly connects different devices. The main purpose of Bluetooth is to connect non-resident devices and applications. Depending on the implemented Bluetooth wireless technology system, data rates between 721.2 kbit/s (Basic Rate) and 24 Mbit/s (AMP) are supported[33]. In contrast to Bluetooth, Wi-Fi was intended to connect resident devices and application, replacing wired local area networks. The technology provides wireless local area networks (WLAN) based on the IEEE 802.11 standards that are created and maintained by the IEEE LAN/MAN Standards committee. The latest standard 802.11n[34] allows maximum raw data rates from 54 Mbit/s to 300 Mbit/s and 600 Mbit/s. If a long-range connectivity is required for an application scenario, most devices use the cellular communication systems that are set up by mobile phone providers. The first widely used technologies were the General Radio Packet Service (GPRS) and the Enhanced Data rates for GSM Evolution (EDGE)[35] using the Global System for Mobile Communication (GSM) standard. While GPRS allowed peak data rates of 57.6 kbit/s, EDGE offered faster peak rates at about 473.6 kbit/s. A new EDGE version called EDGE Evolutions even promises peak rates of 1894 kbit/s. As both technologies extend the second-generation wireless telephone technology (2G) using the GSM norm, they are also called 2.5G (GPRS) and 2.75G (EDGE) technologies. In Europe, Universal Mobile Telecommunications System (UMTS) is the most common third-generation (3G) mobile telecommunications technology, which is also being developed into a 4G technology.

While the name UMTS, is usually used in Europe the technology is better known as FOMA (Freedom of Mobile Multimedia Access)[36] or W-CDMA outside of Europe. Depending on the technologies used, UMTS allows peak download rates of 384 kbit/s to 1920 kbit/s. The protocols High-Speed Downlink Packet Access (HSDPA) and High-Speed Uplink Packet Access (HSUPA) mark the next step towards faster mobile data transmission and are known as 3.5G or 3G+ technologies. Current HSPA data rates allow theoretical rates of up to 14,4 Mbit/s for downloading data (HSDPA) and 5.76 Mbit/s for uploads (HSUPA). Future releases of the 3GPP (*3rd Generation Partnership Project*) specification increase the data rates of what is called HSPA+ then to 84 Mbit/s for downloads and to 23 Mbit/s for uploads. In 2008, the ITU-R organisation defined the minimum speed requirements for the fourth generation of cellular network standards, 4G: For high mobility scenarios (communication from trains or cars), a minimum of 100 Mbit/s is required while for low mobility communication (stationary usage or pedestrians) a peak speed of even 1 Gbit/s is required[37]. Although the above mentioned requirements are not completely fulfilled, technologies like the Mobile WiMAX standard[38] are branded as 4G technology. The current Mobile WiMAX standard offers peak data rates of 128 Mbit/s (downlink) and 56 Mbit/s (uplink) while a new standard, *Mobile WiMAX Release 2*, is currently developed which shall fully comply with the 4G requirements postulated by the ITU-R. Concerning LTE, the maximum bandwidth is currently 300 Mbps for download and 75 Mbps for upload. Similar to WiMAX, an extension of LTE, LTE-Advanced, will comply to the requirements of the ITU-R. 4G may allow roaming with wireless local area networks, and may interact with digital video broadcasting systems.

3.2. Ubicomp methodologies

3.2.1. Interaction beyond the desktop

The new way of interaction beyond the desktop has been marked by the development of *tangible user interfaces* (TUIs) (Ishi and Ullmer 1997). This means the user interface in which a person interacts with digital information through the physical environment. The initial name, *Graspable User Interface*, is no longer in use. Hiroshi Ishii, leader of the Tangible Media Group at the MIT Media Laboratory, is one of the pioneers in TUI development. His particular vision for tangible UIs, called Tangible Bits, is to give physical form to digital information, making bits directly physically graspable and perceptible to changes (Ishi and Ullmer 1997). Tangible bits pursues seamless coupling between these two very different worlds of bits and atoms.[39]

Programming of applications, which is currently done on personal computers, can be done for example on mobile devices and first examples are

TouchDevelop (Tillmann 2011) and MobiDev (Seifert 2011). The concepts intend to support end-users to adapt their personal devices to their own needs. By programming on the device (e.g., a smart phone), no special environment like a personal computer is necessary. This might especially be helpful to people living in developing countries as the numbers of people having access to personal computers is rather low compared to people having access to a mobile phone (ITU 2010).

3.2.2. From the lab to the real world

The European Network of Living Labs (ENoLL) is the "International Federation of Benchmarked Living Labs" in Europe and worldwide[40]. This EU-infrastructure project has been initiated in November 2006 under the auspices of the Finnish European Presidency, and has to this date 4 Waves launched resulting in 212 accepted Living Labs. A Living Lab is a real-life test and experimentation environment where users and producers co-create innovations. Living Labs have been characterized by the European Commission as *Public-Private-People Partnerships* (PPP Alliance) for user-driven open innovation. A Living Lab employs four main activities:

1. Co-Creation: co-design by users and producers
2. Exploration: discovering emerging usages, behaviors and market opportunities
3. Experimentation: implementing live scenarios within communities of users
4. Evaluation: assessment of concepts, products and services according to socio-ergonomic, socio-cognitive and socio-economic criteria.

3.3. Impact on society

A ubiquity of devices translates into a ubiquity of collecting, processing and disseminating data. This has a wide range of consequences on society – most notably, on knowledge and education (information *on* a wider range of events, phenomena etc. can be produced and published *by* and become available *to* more people) and on personal data and privacy. The sphere of education has already been discussed in Section 2.3.2; here, we will concentrate on privacy and closely related topics.

3.3.1. *Challenges of and for ubicomp: privacy, human rights, surveillance and freedom*

Privacy is a contested term, which is used in many different meanings; a meta-analysis of different disciplines' literatures has led us to distinguish between (a) privacy as hiding / confidentiality – the "the right to be let alone"; (b) privacy as control / informational self-determination – the ability to control what happens with one's personal information; and (c) privacy as practice / identity construction – the freedom from unreasonable constraints on the construction of one's own identity, be it by strategically being able to reveal or conceal data (Gürses and Berendt 2010). All of these forms of privacy are recognized by many legal systems as rights (even if of different "priorities" within the respective systems, ranging from basic human rights to rights in specific contractual or citizen-state relations), cf. Gürses, Berendt, and Santen (2006). With regard to personal data and privacy, ubicomp inherits many of the opportunities and threats of the Internet – simply put, there is (a lot) more data from and on (a lot) more people and (a lot) more of their concerns and activities, and this situation gives rise to (even more) new forms of combining data and drawing inferences from them. The implications of these changes have been described in a huge literature, ranging from the dystopian of Big-Brother-like societies (see cf. Garfinkel 2000) to the utopian of increased democracy and new forms of transparency (see cf. Ishi and Lutterbeck 2001 or Brin 1998). So is there a categorical (rather than just quantitative) difference between networked desktop/laptop computers and ubiquitous devices? The former are large enough to be visible, they are discernible as computing machines, and they are comparatively stationary; the latter are small, embedded and mobile. This leads to three fundamental differences:

- New types of data can be recorded with high accuracy, data that may be deemed particularly sensitive (location information, health information, etc.).
- The distinctions between situations, activities, contexts and the social roles associated with them are increasingly blurred: whereas formerly, work-related and public information-gathering activities were the mainstay of computational and therefore recordable activity, now essentially *all* situations are potentially being recorded by an unknown array of observers and the recordings made available to an unknown array of viewers.
- The distinction between situations with and without data collection becomes more difficult to make (the small size of sensing devices means that people cannot in general know whether sensing takes place and how).

For technical communication, these specifics of ubicomp harbor certain opportunities: more genuine data on usage activities of technical artifacts can be collected and tied to the technical communication data used in those situations.

This can make it possible to improve, maintain and personalize technical communication to an unprecedented degree.

On the other hand, ubicomp – even more so than the Internet, and regardless of whether the situation involves technical communication or not – conjures up a perception of the world as panopticon.[41] Panoptical structures involve or at least provide for ubiquitous *surveillance*: the monitoring of the behavior, activities, or other changing information, usually of people and often in a surreptitious manner. 'Surveillance' most usually refers to observation of individuals or groups by government organizations or corporations. However, the ready availability of – for example – video-recording functionality on standard mobile phones and video-publishing functionality on the Web also means that anybody can collect and disseminate highly personal data without any possibility of (e.g. democratic) control over these processes (Zittrain 2008).

3.3.2. Ethical consideration of ubiquitous computing research and deployment

Privacy research and policy start from the above-mentioned positing of privacy as a right and focus strongly on the individual. One basic fault line in this discussion is the question to what extent privacy and personal data are a fundamental human right or a personal possession. In the first case, privacy may have to be protected by the state potentially even against the explicit wishes of the data subject herself, just like the state has to prevent people from other forms of (even consensual) behavior that "*violates human dignity*". In the second case, personal data can become a tradable commodity – ranging from comparatively innocuous tradings of information for personalization to consenting to being filmed 24–7 in TV Reality Shows and beyond (cf. Berendt 2012). The privacy questions raised by highly networked environments like social network sites increasingly show that the focus on the individual in these discussions ignores essential "external effects" of one person's decisions on other people's privacy (cf. Gürses and Berendt 2010). Surveillance studies (Lyon 2007) takes such considerations one step further by stressing the societal (as opposed to only individual) effects of surveillance and data collection. The close association between surveillance, the internalization of its values by the observed, and resulting power over them, was stressed already by Bentham and Foucault. A clear expression of the need for data protection to safeguard also democracy was given by the German Constitutional Court in its 1983 ruling on the National Census, a ruling which became highly influential in later EU privacy legislation:

> A person who does not know whether deviant behaviors are, at all times, noted and stored, used or disseminated as data, will try not to become conspicuous through such behaviors. […] This would not only have an impact on the chances of individual pursuits of interests, but also on the common welfare, because self-determination

> is an elementary condition of functioning for a free democratic society based on its citizens' ability to act and participate. This implies: The free development of the individual requires, under the modern conditions of data processing, the protection of the individual against the unlimited collection, storage, use and dissemination of his personal data[42].

Privacy and surveillance researchers build on strong mono-disciplinary developments (such as cryptographic solutions to reduce data linkability and traceability, or privacy-preserving data mining / data publishing to reduce inferences from data or data analyses), and they increasingly stress the need for interdisciplinary and policy-related action[43]. Some current examples of EU-funded projects are SAPIENT[44] and for privacy-respecting forms of data mining is MODAP – Mobility, Data Mining and Privacy[45]. In sum, these discussions show that privacy and surveillance legislation and practice are affecting individual and societal values and developments in similarly fundamental ways as, for example, bioethical questions, and therefore merit broad, deep and ongoing societal and political debate. Ubiquitous computing research and deployment do not exist in an ethical vacuum and will have to take these debates into account more strongly in the future.

4. Research challenges in ubiquitous computing

In this subchapter, we will give only a brief selection of current research trends and challenges. We apologize for not being able to cover the full landscape of current research areas in ubiquitous computing. Nevertheless, we focus on two thematic aspects:

- Section 4.1 discusses general research challenges including interaction beyond the desktop (Section 4.1.1), recent developments in environmental and GIS computing (Section 4.1.2), how we can predict future events (Section 4.1.3) and how to handle resource-constrained devices (Section 4.1.4), volatile execution environments (Section 4.1.5), heterogeneous execution environments (Section 4.1.6), fluctuating user environments (Section 4.1.7) and recent developments in invisible computing (Section 4.1.8).
- Section 4.2 on next generation media will deal with ubiquitous applications using Web 2.0 embedded technologies (Section 4.2.1), embedded speech recognition, natural language processing (Section 4.2.2).

4.1. Current research challenges

As we are currently half way in the post-desktop era, there are still some major research challenges as well as requirements for developing ubiquitous computing systems. In the following section, we will shortly introduce the main challenges. The interested reader is directed to (Bardram 2010 and May 2010).

4.1.1. Interaction beyond the desktop and natural user interfaces (NUI)

Starting from "*computing is all about interface*", we look back at the development of user interfaces from "*text*" (command-line interfaces – CLI) to "*graphics*" over GUI (Graphical User Interface) to TUI (Tangible User Interface), now to "*objects*" over NUI (Natural User Interface). NUI is *touch, gesture and manipulation of physical/digital content*. NUI can perform object recognition on a tangible interface, blurring physical with digital media. The interesting idea behind this is called "*smart matter*"[46] and is rooting from micro-electro-mechanical systems. Fundamental properties of materials, that are usually taken for granted (e.g. colour, shape, elasticity, and texture), have embedded very small sensing devices like e.g. micro cells and these can change their physical status on demand.

> Matter itself becomes dynamic, such that computing then becomes something that no longer just resides in distinct computing devices, but becomes a permeating part of almost every manufactured product. Such a vision is, of course, far off in the future, but the research for such possibilities is being done today. (from http://www.cc.gatech.edu/classes/cs6751_97_fall/projects/gacha/daniels_essay.html#Intro – retrieved on April 1, 2010)

Ubiquitous application including components built on microscopic scale are for example "Digital Ink"[47] which implies producing a new kind of "intelligent paper" with embedded microcells capable of changing their color and consequently of changing or deleting contents, downloading new headlines and information on the fly. Other possible complex ubiquitous products and applications[48] with this regard are, for example:

- clothing material with a thin outer layer possible of changing its "color" when requested;
- embedded GPS sensors, that can be embedded in the label on a package, so that the package can always be "aware" of its location;
- small chips that are deliberately tilled into the soil on a farm, so that each chip can "report back" about the temperature and humidity at its location;
- building materials that are capable of strengthening or weakening their flexibility in response to weather conditions;
- insulation materials capable of dynamically adjusting the amount of heat or cold that they allow to pass through[49].

4.1.2. Environmental computing and GIS with ubicomp

Ubiquitous computing is dealing with different kinds of geographic and environmental data that can be instantly linked to the location-relevant information about the user. Current Geographic Information Systems (GIS) are able to lead the user to the desired location and at the same time to track users' trajectories and combine this information, e.g. in telling the user, who – which friend or known person – is next to him. See EU-funded project MODAP[50] and UBIQ-TRACKs. OpenStreetMap[51], also called free WIKI World Map, is an impressive Wiki project that involves people from all over the world in creating and sharing free geographic data such as street maps, tracking data, etc.

4.1.3 Predicting the future

With the advances of connecting smart devices systems can more and more recognize what is happening in the closer environment. If this data is used carefully, our life – and the world in general – might become more predictable. If objects like buses, trains, and cars as well as people are tracked in real-time (e.g., position, activity, consumption/food intake, eye-gaze) and if the environmental conditions are permanently sensed, models of the current world can be generated. Founded on the sensed data, a likely future can be predicted (Schmidt 2011a). Technologies like a physics engine support the prediction of at least the immediate future.

4.1.4 Resource-constrained devices

Although there are already ubiquitous devices like the latest series of Smart Phones (e.g. Motorola Atrix) which have highly powerful CPUs, a lot of ubiquitous devices have significantly less resources and computing power than current PCs. Thus, one goal is to stick to the paradigm of resource-aware computing:

> "Resource-aware computing is an approach to develop technologies where the application is constantly notified about the consumption of vital resources, and can help the application (or the user) to take a decision based on available resources now and in the future." (Bardram 2010: 42)

With the advances of computing power the most constraining resources are energy and connectivity: "One of the main hardware constraints to consider when building ubicomp systems and applications is power consumption and/or opportunities for energy harvesting – including recharging" (Bardram 2010: 42). Solutions to deal with those restrictions are, e.g., adapting the display brightness to the current battery level of a ubiquitous device or harvesting the kinetic energy of a user while he/she is walking (*power foraging*). Depending on the location and the available energy also the network connectivity might vary – applications have to deal with unpredicted network speeds as well as they might have to decide between high-speed, high power consuming connections and connections that might require less energy but which also provide a slower rate at the same time.

4.1.5 Volatile and heterogeneous execution environments

With the proliferation of wireless networks of all kinds, devices are able to discover and connect to each other, allowing using further services which the connected devices provide. Thus, e.g., a smart phone while being in the car might provide applications that can be shown on the central display of the car and that can be controlled by the input devices which the car-integrated infotainment system provide. An example for this scenario is the Pandora webradio application on the iPhone which can be connected to specific Mini and BMW cars. A typical ubiquitous computing scenario might include different hardware and software components. They might differ relating to hardware components, network technology, operating systems, available input and output modalities, resources, and connected sensors. As the ubiquitous application might span across several devices, technologies are required to coordinate the operation and provide the application's features. Assume that a user carrying a smart phone enters a smart room providing means to display contents on a large screen and some input capabilities. In order to use those services in combination of the phone, standard technology stacks are necessary to allow a standardized communication between those devices.

4.1.6. Fluctuating user environments

As mentioned before the ubiquitous computing era led to a one-to-many relationship between end user and devices. This situation gets more complicated as the same device might even be shared between several users extending the aforementioned relation to a many-to-many relation. Thereby, a simple task might not be bound to a specific device of a single user but distributed across numerous possibly heterogeneous devices. One big challenge in many cases is the effort that needs to be spent in order to setup and manage services across all those devices. The fluctuation of the user environment can be related to the changing location of users (e.g., location-based services, navigation tasks), to varying user contexts (e.g., who is using the application?, who is around?), and fluctuating user activity.

4.1.7. Invisible computing

As part of Marc Weiser's visionary paper back in 1991, his vision of ubiquitous computing is that computers should be invisible to the user – in two ways:

1) the computer should be embedded/hidden within other devices to make it disappear visually (physical invisibility).
2) Additionally, the computer should behave impercible in the periphery of the user's attention to make it mentally invisible.

Topics that are dealt with include autonomic computing (systems that mainly manage themselves), proactive computing (anticipate future handling and adapt accordingly), and graceful degradation (no dependence on volatile resources).

4.2. Next generation media

4.2.1. Ubiquitous applications using Web 2.0 embedded technologies

InfoPad is one of the first ubiquitous applications using integrated web 2.0 technologies. The InfoPad platform allows real-time access of multimedia data through the use of inexpensive, portable devices, where the data is accessed over high-speed networks[52]. InfoPad is sending data into the ubiquitous devices and tasks such as playback of video and audio files, collaboration on projects, and sharing of database information and data, scheduling are possible in real-time over hand-held devices. Another ubiquitous application in this area is the Touch MewE@r (="Touch me Wear") (Beach 2009) project that explores possibilities of association between existing social network information like e.g. Facebook real-time information with real-world physical interaction embedded in clothes or other wearable of the users. Information from the real-world, like for example who was hugged by whom and where and when can be sent to Facebook directly from the wearable of the user and appear under his social network profile. A lot of ubiquitous applications using Web 2.0 embedded technologies are available over mobile devices (like PDAs, smart phones, iPhones, etc. The sense-sation platform (Sahami 2010), for example, uses web technologies and protocols to integrate smart devices into the web. The idea is to promote (locally and remotely) the available sensor data of current smart phones via RESTful web services. Thus, (web) developers are able to create applications that can use and process mobile context data without having to deal with mobile phone developing at all.

4.2.2. Embedded speech recognition and natural language processing

Automatic speech recognition (ASR) and natural language processing (NLP) tasks require usually high computation capability and enough memory, and at the beginning, it was difficult to implement those tasks successfully in embedded and mobile devices (see also Martin and Schultz 2012 in this volume). In the meantime, embedded speech recognition especially for handheld mobile devices has improved its performance considerably and tasks like voice dialling, talking dictionaries, e.g. searching for a location in GPS device by acoustic input, contributing to games in internet over micro etc. are no longer a technological challenge, but furthermore they have become interesting future business challenges. The embedded speech and NLP markets are evolving very fast especially in information-rich ubiquitous telematics applications, automotive and call centre applications, and also in ubiquitous health care sector, as e.g. embedded speech devices in health assistance equipment or in barrier-free entertainment products for disabled or elderly people. Large companies like for example Sun, Apple, AT&T, Dragon Systems, IBM, Novell, Philips and Texas Instruments are already collaborating on the development of standards e.g. Java Speech API (see Burkhardt 2001: 223–240) and standards for VoIP (=Voice over IP – Internet Protocols) applications.

5. Conclusions: Ubicomp and technical communication – Potentials and challenges

We have illustrated that especially in terms of hand-held ubiquitous devices and basic sensoric, context awareness and adaptation, the ubiquitous technologies and products have already successfully entered and revolutionized the market and business infrastructures. Nevertheless when it comes to complex ubiquitous applications like smart homes, ubiquitous cities and smart matter, these applications may work in a well-defined laboratory setting, but when it comes to real business settings these applications are still marching on most trivial limitations given by today's hardware and usability settings. For example even if the sensors are getting more powerful in terms of capacity and cheaper in terms of price, this does not make a huge price difference for the complex hand-held devices, like smart phones, PDAs that are still pretty expensive for one user to have too many of them[53]. Large ubiquitous infrastructures are still lacking sufficient bandwidth, and the most urgent limitations that have to be solved are the ones given by the troublesome software infrastructure, that needs to be formally installed, updated, maintained, adjusted, etc. The only exception to this may be the embedded firmware programs, which are burned in as part of the hardware device[54].

If we now try to summarize our journey through the ubiquitous computing field, we are tempted to claim that the most urgent challenges that in ubicomp information-rich applications have to be solved are of communicative nature. The way in which the right information has to be processed and properly sent to the user is still the major issue. The infrastructure for setting up ubiquitous systems that are able to "perceive" information from different environments on the fly is still a huge challenge. Several EU-funded projects as, for example, CONET (= Cooperating Objects) NOE[55] make the attempt to develop such an infrastructure that brings us even beyond the "Internet of things" into complex "cooperating infrastructures" enabling wireless networks systems (sensory systems) to communicate with pervasive systems and embedded devices on the fly like in Figure 4:

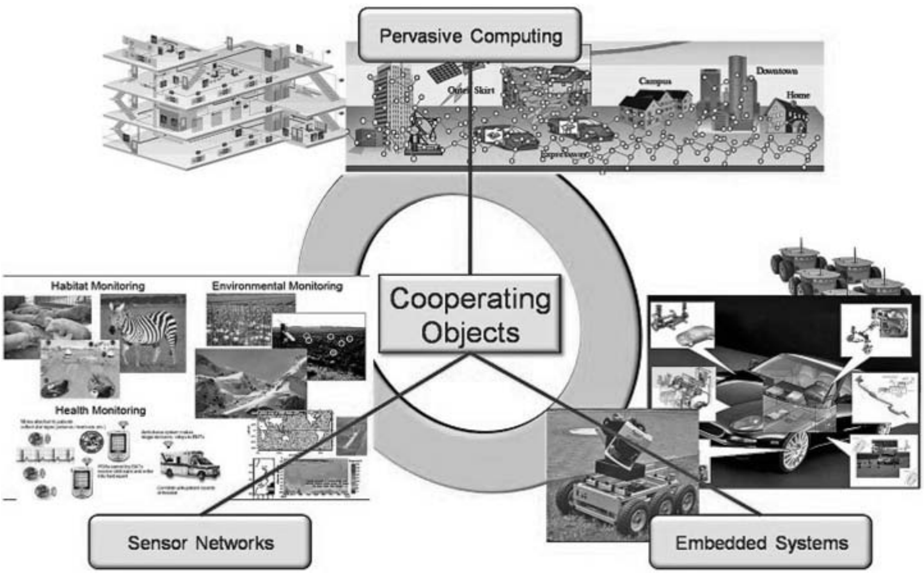


Figure 4. Cooperating Objects enable the communication between sensor networks, embedded systems and pervasive applications[56].

In order to overcome the resource-constraints for data processing and transmission, we should make use of advanced visual functionalities, for example to solve complex search and navigation through content-rich landscapes problems:

> A picture can connect the strategic with the tactical in a way no other communication form possibly can. (http://semanticstudios.com/publications/semantics/000633.php – retrieved on April 1, 2010)

The idea here is of having "experience maps" to help the user to participate to his journey through the contents and help the system to understand on the fly what is the right application or the information he/she needs for his situation.

This statement can be seen as the "ubiquitous service design" perspective and it pleads for improving methodologies for a better "*information architecture*" in terms of a bridge between user experience and service design embedded in the device[57].

In order to realize advanced ubiquitous technologies, not only the technological development but everything that comes with the technology has to be reconsidered and developed appropriately.

> "The goal of Communication artists is certainly not to produce first level meanings, but above all to make us aware as to how, in the end, the generalized practice of Communication interacts on the whole of our sensory system. This evolution is about to put into place the data for a "new awareness" at the edge of our perception, and then, along with new "ways of feeling", it will open up new aesthetic paths. " (from http://www.hcipatterns.org/ – retrieved on April 1, 2010)

Notes

1. On web syndication see http://en.wikipedia.org/wiki/Web_syndication (retrieved on April 1, 2010).
2. An example is www.bibsonomy.org.
3. Information Society and Media – In ICT Security Research in FP7: ftp://ftp.cordis.europa.eu/pub/fp7/ict/docs/security/fp7-brochure-infso-f5-low-resolution_en.pdf (retrieved 01. 04. 2010).
4. See http://www.clker.com/clipart-23386.html, http://en.wikipedia.org/wiki/File:Computer-aj_aj_ashton_01.svg (retrieved on April 1, 2010).
5. http://de.wikipedia.org/wiki/Ambient_Intelligence (retrieved on April 1, 2010).
6. See KDubiq (Knowledge Discovery in Ubiquitous Environments) CA (Coordination Action) under FET open FP6: http://www.kdubiq.org (retrieved on April 1, 2010).
7. See KDubiq (Knowledge Discovery in Ubiquitous Environments) CA (Coordination Action) under FET open FP6: http://www.kdubiq.org (retrieved on April 1, 2010).
8. On Microsoft Kinect see http://www.xbox.com/kinect (retrieved on April 1, 2010).
9. http://de.eye.fi/ (retrieved on April 1, 2010).
10. On Whispernet on Amazon Kindle see http://www.amazon.com/gp/help/customer/display.html?nodeId=200375890 (retrieved on April 1, 2010).
11. On RFID see http://en.wikipedia.org/wiki/Radio-frequency_identification (retrieved on April 18, 2011).
12. "Natural interaction is defined in terms of experience: people naturally communicate through gestures, expressions, movements, and discover the world by looking around and manipulating physical stuff. The key assumption here is that people are meant to interact with technology as they are used to interact with the real world in everyday life, as evolution and education taught them to do." (Valli 2007)
13. Actuators are usually mechanical devices for moving or controlling a mechanism or system.

14. On Smart Grid technology see http://en.wikipedia.org/wiki/Smart_grid (retrieved on April 1, 2010). Further information for example on the currently emergent area called Smart Grid under EU (on Smart Grid EU-funded projects see http://ec.europa.eu/energy/gas_electricity/smartgrids/smartgrids_en.htm – retrieved on April 1, 2010) and KD2U (= Knowledge Discovery in Distributed and Ubiquitous Environments) Smart Grid Symposium (see http://www.kd2u.org/ – retrieved on April 1, 2010).
15. On the House_N project at MIT see http://architecture.mit.edu/house_n/ (retrieved on April 1, 2010).
16. On Ubiquitous Cities see http://en.wikipedia.org/wiki/Ubiquitous_city (retrieved on April 1, 2010).
17. On Pervasive Health see also http://www.pervasivehealth.org/?page_name=call-for_papers (retrieved 01. 04. 2010).
18. On Emotional Cartography see http://emotionalcartography.net (retrieved on April 1, 2010).
19. For a tutorial on sensors see http://www.engineersgarage.com/articles/sensors (retrieved on April 1, 2010).
20. Perception is the process by which an organism attains awareness or understanding of its environment by organizing and interpreting sensory information.
21. See KDubiq (Knowledge Discovery in Ubiquitous Environments) CA (Coordination Action) under FET open FP6: http://www.kdubiq.org (retrieved on April 1, 2010) and KD2U (= Knowledge Discovery in Distributed and Ubiquitous Environments): http://www.kd2u.org/ (retrieved on April 1, 2010).
22. On Microsoft Kinect see http://www.xbox.com/kinect (retrieved on April 1, 2010).
23. See http://www.primesense.com/?p=487 (retrieved on April 1, 2010).
24. See http://www.primesense.com/?p=487 (retrieved on April 1, 2010).
25. See http://community.middlebury.edu/~schar/papers/structlight/structlight.pdf (retrieved on April 1, 2010).
26. See http://en.wikipedia.org/wiki/Time-of-flight_camera (retrieved on April 1, 2010).
27. See www.ubisense.net (retrieved on November 27, 2011).
28. See http://www.naturalpoint.com/optitrack/ (retrieved on November 27, 2011).
29. "In Ubiquitous Computing, interaction with computers is inevitably in context and in most cases context matters for not only the users directly, but it also matters indirectly for the system." (Schmidt 2002: 5).
30. See http://www.irc.atr.jp/~noma/papers/Paper/CHI-2002-OrfactDisplay. pdf (retrieved on April 1, 2010).
31. http://en.semapedia.org/ (retrieved on April 1, 2010).
32. On Bluetooth SIG see https://www.bluetooth.org/Technical/Specifications/adopted.htm (retrieved on April 1, 2010).
33. On the specification of the Bluetooth System (Architecture & Technology Overview, Covered Core Package version 4.0, Vol. 1, p. 17, June 2010, Bluetooth SIG) see https://www.bluetooth.org/docman/handlers/downloaddoc.ashx?doc_i=229737 (retrieved on April 1, 2010).
34. See http://standards.ieee.org/findstds/standard/802.11n-2009.html (retrieved on April 1, 2010).
35. On Enhanced Data rates for GSM Evolution (EDGE); Project scheduling and open issues for EDGE see http://www.3gpp.org/ftp/Specs/html-info/50059.htm (retrieved on April 1, 2010).

36. http://de.wikipedia.org/wiki/Ubiquitous_Computing (retrieved on April 1, 2010).
37. See http://www.itu.int/rec/R-REC-M.1645/en as well as http://www.itu.int/rec/R-REC-M.1645-0-200306-I/en and http://www.itu.int/dms_pubrec/itu-r/rec/m/R-REC-M.1645-0-200306I!!PDF-E.pdf (Recommendation M.1645: Framework and overall objectives of the future development of IMT-2000 and systems beyond IMT-2000, ITU) (retrieved on April 1, 2010).
38. IEEE 802.16e-2005.
39. “Tangible Bits” is our vision of Human Computer Interaction (HCI): we seek a seamless coupling of bits and atoms by giving physical form to digital information and computation.(from http://tangible.media.mit.edu/projects.php – retrieved on April 1, 2010)
40. On ENoLL see http://www.openlivinglabs.eu/ (retrieved on April 1, 2010).
41. A panopticon is a structure that allows an observer to observe all to-be-observed without the latter being able to tell whether they are being watched. Originally an architectural design principle for prisons described by J. Bentham in 1785, the panopticon became a metaphor for institutional/societal structures that function like this (M. Foucault described modern ‘disciplinary’ institutions and societies and their pervasive inclination to observe and normalise in this way, in his book Discipline and Punish, original: *Surveiller et punir, naissance de la prison*, Paris: Gallimard, 1975).
42. The 1983 “Census Judgement” is a fundamental decision of the German Federal Constitutional Court (Az.: 1 BvR 209, 269, 362, 420, 440, 484/83); it is available at http://zensus2011.de/fileadmin/material/pdf/gesetze/volkszaehlungsurteil_1983.pdf (retrieved on April 1, 2010).
43. For resource collections, see for example http://www.epic.org and http://people.cs.kuleuven.be/~bettina.berendt/SPION.
44. On SAPIENT – Supporting Fundamental Rights, Privacy and Ethics in Surveillance Technologies see http://uc.inf.usi.ch/sapinet (retrieved on April 1, 2010).
45. On MODAP – Mobility, Data Mining and Privacy see http://www.modap.org (retrieved on April 1, 2010).
46. See http://www.cc.gatech.edu/classes/cs6751_97_fall/projects/gacha/daniels_essay.html#Intro (retrieved on April 1, 2010).
47. See http://www.cc.gatech.edu/classes/cs6751_97_fall/projects/gacha/daniels_essay.html#Intro (retrieved on April 1, 2010).
48. See http://www.cc.gatech.edu/classes/cs6751_97_fall/projects/gacha/daniels_essay.html#Intro (retrieved on April 1, 2010).
49. See http://www.cc.gatech.edu/classes/cs6751_97_fall/projects/gacha/daniels_essay.html#Intro (retrieved on April 1, 2010).
50. On MODAP EU-funded project under 7FP see http://www.modap.org (retrieved on April 1, 2010).
51. See http://www.openstreetmap.org/.
52. On the InfoPad see http://bwrc.eecs.berkeley.edu/research/old/infopad/index.html (retrieved on April 1, 2010).
53. See http://www.cc.gatech.edu/classes/cs6751_97_fall/projects/gacha/daniels_essay.html#Intro (retrieved on April 1, 2010).
54. See http://www.cc.gatech.edu/classes/cs6751_97_fall/projects/gacha/daniels_essay.html#Intro (retrieved on April 1, 2010).
55. On semantic studios see http://semanticstudios.com/publications/semantics/000633.php (retrieved on April 1, 2010).

56. See http://semanticstudios.com/publications/semantics/000633.php (retrieved on April 1, 2010).
57. See http://semanticstudios.com/publications/semantics/000633.php (retrieved on April 1, 2010).

References

Allwood, Jens and Elisabeth Ahlsén
2012 Multimodal communication. In: Alexander Mehler, Laurent Romary and Dafydd Gibbon (eds.), *Handbook of Technical Communication*. Berlin/ Boston: De Gruyter.

Acampora, Giovanni and Vincenzo Loia
2008 A proposal of ubiquitous fuzzy computing for Ambient Intelligence. *Information Sciences* 178(3): 631–646.

Athanasios, V. Vasilakos
2008 Special Issue: Ambient Intelligence. *Information Sciences* 178 (3): 585–587.

Augusto, Juan Carlos
2007 Ambient Intelligence: The Confluence of Pervasive Computing and Artificial Intelligence. In: Assaf Schuster (ed.): *Intelligent Computing Everywhere*, 213–234. London: Springer.

Bardram, Jakob and Adrian Friday
2010 Ubiquitous Computing Systems. In: John Krumm, *Ubiquitous Computing Fundamentals*, 37–94. Boca Raton, FL: Chapman & Hall / CRC.

Bardram, Jakob E., Alex Mihailidis and Dadong Wang (eds.)
2006 Pervasive Computing in Healthcare. CRC Press.

Beach, Aaron, Raz Baishakhi and Leah Buechley
2009 Touch Me Wear: Getting Physical with Social Networks. In: *IEEE International Conference on Computational Science and Engineering*, vol. 4, 960–965.

Berendt, Bettina
2012 More than modelling and hiding: Towards a comprehensive view of Web mining and privacy. In: *Data Mining and Knowledge Discovery, Special Issue on 10 Years of Mining the Web*. 24(3), 697–737.

Bikakis, Antonis, Theodore Patkos, Gregoris Antoniou and Dimitris Plexousakis
2007 A Survey of Semantics-Based Approaches for Context Reasoning in Ambient Intelligence. In: Max Mühlhäuser, Alois Ferscha, Erwin Aitenbichler (eds.), *Constructing Ambient Intelligence – AmI Workshops, Darmstadt, Germany, Nov. 2007, Revised Papers*, Springer CCIS 11, 14–23.

Bolt, Richard A.
1981 Gaze-orchestrated dynamic windows. In: *Proceedings of SIGGRAPH'81*, ACM Press, 109–119.

Botthof, Alfons et. al.
2010 Internet der Dinge – Leitfaden zu technischen, organisatorischen, rechtlichen und sicherheitsrelevanten Aspekten bei der Realisierung neuer

RFID-gestützter Prozesse in Wirtschaft und Verwaltung, Dokumentation der BMWI, Berlin.

Brin, David
1998 The Transparent Society: Will Technology Force us to Choose Between Privacy and Freedom? Perseus Books.

Brumitt, Barry, Brian Meyers, Daniel Robbins, John Krumm, Mary Czerwinski and Steve Shafer
1998 The New EasyLiving Project at Microsoft Research. In: *Proceedings of the DARPA/NIST Smart Spaces Workshop*, July 1998, 127–130.

Brumitt, Barry and Steven A. Shafer
2001 Better Living Through Geometry. In: *Personal Ubiquitous Computing*. 5, 1, 42–45.

Bullinger, Hans-Joachim and Michael ten Hompel
2007 Internet der Dinge. Springer, Berlin 2007.

Burkhardt, Jochen, Horst Henn, Stefan Hepper, Klaus Rindtorff and Thomas Schäck
2001 Pervasive Computing – Technologie und Architektur mobiler Anwendungen. München: Addison-Wesley.

Cai, Yang
2007 Ambient intelligence: From interaction to insight. *International Journal of Human-Computer Studies* 65(5), 419–420.

Campbell, Andrew T., Nicholas D. Lane, Emiliano Miluzzo, Ronald A. Peterson, Hong Lu, Xiao Zheng, Mirco Musolesi and Kristof Fodor
2008 The Rise of People-Centric Sensing. *IEEE Internet Computing*, 1089–7801/08.

Chen, Tsung Teng and M. Lee
2008 Ubiquitous Computing in Prospect: a Bibliographic Study In: *Proceedings of the 2008 International Symposium on Ubiquitous Multimedia Computing*, IEEE Computer Society.

Cook, Diane J., Juan C. Augusto and Vikramaditya R. Jakkula
2009 Ambient intelligence: Technologies, applications, and opportunities. *Pervasive and Mobile Computing* 5(4), 277–298.

Dey, Anind K.
2001 Understanding and Using Context. In: Personal and Ubiquitous Computing, 5(1) 4–7.

Diekmann, Thomas and Svenja Hagenhoff
2003 Ubiquitous Computing: State of the Art. Georg-August-Universität Göttingen, Institut für Wirtschaftsinformatik: Arbeitsbericht Nr. 24/2003. (http://www2.as.wiwi.uni-goettingen.de/getfile?DateiID=470, retrieved 2012-04-08).

Diewald, Nils
2012 Decentralized Online Social Networks. In: Alexander Mehler, Laurent Romary and Dafydd Gibbon (eds.), *Handbook of Technical Communication*. Berlin/Boston: De Gruyter.

Drewes, Heiko and Albrecht Schmidt
2007 Interacting with the Computer Using Gaz Gestures. In: *Proceedings of INTERACT 2007*, 475–48.

Fleisch, E. and F. Mattern, (ed.)
2005 Das Internet der Dinge – Ubiquitous Computing und RFID in der Praxis. Berlin: Springer.
Garfinkel, Simson
2000 *Database Nation: The Death of Privacy in the 21st Century*. Beijing/Cambridge/Farnham: O'Reilly Media.
Gürses, Seda, Bettina Berendt and Thomas Santen
2006 Multilateral security requirements analysis for preserving privacy in ubiquitous environments. In: Bettina Berendt and Ernestina Menasalvas (eds.), *Proceedings of the Workshop on Ubiquitous Knowledge Discovery for Users at ECML/PKDD'06*, Berlin, Germany, 22 September 2006, 51–64.
Gürses, Seda and Bettina Berendt
2010 The social web and privacy: Practices, reciprocity and conflict detection in social networks. In: Elena Ferrari and Francesco Bonchi (eds.), *Privacy-Aware Knowledge Discovery: Novel Applications and New Techniques*. Boca Raton, FL: Chapman & Hall/CRC Press.
Intille, S. Stephen, Kent Larson, Emmanuel Munguia Tapia, Jennifer S. Beaudin, Pallavi Kaushik, Jason Nawyn, and Randy Rockinson
2006 Using a live-in laboratory for ubiquitous computing research. In: F. K. Fishkin and B. Schiele, Paddy Nixon, and A. Quigley (eds.), *Proceedings of PERVASIVE 2006*, LNCS 3968, Berlin: Springer, 349–365.
Ishii, Kei and Bernd Lutterbeck
2001 Unexploited resources of online education for democracy: Why the future should belong to open courseware. *First Monday* 6(11) (http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/896, retrieved 2010-07-30).
Ishii, Hiroshi and Brygg Ullmer
1997 Tangible bits: towards seamless interfaces between people, bits and atoms. In: *Proceedings of the SIGCHI conference on Human factors in computing systems* (CHI '97). ACM, 234–241.
ITU
2010 *Measuring the Information Society 2010*. Geneve, Switzerland.
Jacob, R. J.
1990 What you look at is what you get: eye movement-based interaction techniques. In: *Proceedings of CHI '90*, ACM Press, 11–18.
Jiang, Changhao, Shi, Yuanchun, Xu, Guangyou and Xie, Weikai
2000 Classroom in the Era of Ubiquitous Computing. Smart Classroom. In: IEEE International Conference on Wireless LANs PANs and Home Networks ICWLPHN (http://www.cs.cmu.edu/~jiangch/pub/smartroom.pdf, retrieved 2012-04-08).
Juels, A.
2006 RFID security and privacy: a research survey. *IEEE Journal on Selected Areas in Communications* 24(2), 381–394.
Kranz, Matthias; Spiessl, Wolfgang Spiessl and Albrecht Schmidt
2007 Designing Ubiquitous Computing Systems for Sports Equipments. In: *Proceedings of the Fifth IEEE international Conference on Pervasive Comput-*

ing and Communications PERCOM 2007, March 19–23, 2007, IEEE Computer Society, 79–86.

Lücking, Andy and Thies Pfeiffer
2012 Framing multi-modal technical communication. In: Alexander Mehler, Laurent Romary and Dafydd Gibbon (eds.), *Handbook of Technical Communication*. Berlin/Boston: De Gruyter.

Lyon, David
2007 *Surveillance Studies: an Overview*. Cambridge: Polity Press.

Mannermaa, Mika
2007 Living in the European Ubiquitous Society In: *Journal of Future Studies* (online available by http://www.jfs.tku.edu.tw/11-4/R01.pdf – retrieved on April 1, 2011).

Martin, Jean-Claude and Tanja Schultz
2012 Multimodal and speech technology. In: Alexander Mehler, Laurent Romary and Dafydd Gibbon (eds.), *Handbook of Technical Communication*. Berlin/Boston: De Gruyter.

May, Michael and Loranza Saitta
2010 Ubiquitous Knowledge Discovery. Berlin: Springer.

Montoro, German, Xavier Alamán and Pablo. A. Haya
2004 A Plug and Play Spoken Dialogues Interface for Smart Environments. In: *Proceedings to the 5th Conference on Computational Linguistics and Intelligent Text Processing*, Seoul, Korea, 360–371.

Sahami, Alireza, Paul Holleis, Albrecht Schmidt and Jonna Häkkilä
2008 Rich Tactile Output on Mobile Devices. In: *European Conference on Ambient Intelligence (Ami'08)*. LNCS. Berlin: Springer, 210–221.

Sahami, Alireza S., Christian Winkler and Albrecht Schmidt
2010 SENSE-SATION: An extensible platform for integration of phones into the Web. In: *Internet of Things (IOT)*, 1–8.

Schmidt, Albrecht, Michael Beigl and Hans-W. Gellersen
1999 There is More to Context than Location In: *Computation & Graphics Journal* 23(6), 893–902.

Schmidt, Albrecht, Kofi Asante Aidoo, Antti Takaluoma, Urpo Tuomela, Kristof Van Laerhoven and Walter Van de Velde
1999 Advanced Interaction in Context In: *1st. International. Symposium on Handheld and Ubiquitous Computing (HUC'99)*, Karlsruhe. Germany, LNCS 1707, Berlin: Springer.

Schmidt, Albrecht, Hans-W. Gellersen and Christian Merz
2000 Enabling Implicit Human Computer Interaction In: *A Wearable RFID-Tag Reader International Symposium on Wearable Computers (ISWC2000)*, October, 16–17, 2000. Atlanta, GA, USA, 193–194.

Schmidt, Albrecht
2000 Implicit Human Computer Interaction Through Context In: *Personal Technologies* 4(2&3), 191–199.

Schmidt, Albrecht and K. Van Laerhoven
2001 How to Build Smart Appliances? In: *IEEE Personal Communications* 8(4), 66–71.

Schmidt, Albrecht
2002 Ubiquitous Computing – Computing in Context. Ph.D. Thesis. Computing Department, Lancaster University, UK (http://comp.eprints.lancs.ac.uk/687/1/Albrecht_Schmidt_PhD-Thesis_Ubiquitous-Computing_print1.pdf, retrieved 2012-04-08).
Schmidt, Albrecht, Lucia Terrenghi and Paul Holleis
2007 Methods and Guidelines for the Design and Development of Domestic Ubiquitous Computing Applications *Special Issue PerCom2007 of Pervasive and Mobile Computing* 6, 721–738.
Schmidt, Albrecht
2008 Ubicomp Applications and Beyond – Research Challenges and Visions. In: *Proceedings to 2nd KDubiq Industry Applications Workshop*, 06/03/2008, University of Porto.
Schmidt, Albrecht, Dagmar Kern, Sara Streng and Paul Holleis
2008 Magic Beyond the Screen. In: *IEEE Multimedia*, 8–13.
Schmidt, Albrecht and Dagmar Kern
2009 Embedding Interaction – Human Computer Interaction beyond the Desktop. Lecture, 16/02/2009, DFKI, Saarbrücken.
Schmidt, Albrecht
2010 Invisible Computing – Are We There Yet. In: *IEEE Computer* 43(2), 95–97.
Schmidt, Albrecht, Bastian Pfleging, Florian Alt, Alireza Sahami Shirazi, and Geraldine Fitzpatrick.
2011 Interacting with 21st-Century Computers. In: *IEEE Pervasive Computing*. To Appear. http://dx.doi.org/10.1109/MPRV.2011.81
Schmidt, Albrecht, Marc Langheinrich, and Kristian Kersting
2011 Perception beyond the Here and Now. *Computer* 44, 2 (February 2011), 86–88. DOI=10.1109/MC.2011.54 http://dx.doi.org/10.1109/MC.2011.54
Seifert, Julian; Bastian Pfleging Elba del Carmen Valderrama Bahamóndez, Martin Hermes, Enrico Rukzio, Albrecht Schmidt
2011 Mobidev: a tool for creating apps on mobile phones. Mobile HCI 2011: 109–112
Siemens, George and Peter Tittenberger
2009 Handbook of Emerging technologies for Learning (http://umanitoba.ca/learning_technologies/cetl/HETL.pdf; retrieved 01. 05. 2011).
Tillmann, Nikolai, Moskal, M., and de Halleux, J.
2011 TouchStudio – Programming Cloud-Connected Mobile Devices via Touchscreen, Microsoft Technical Report MSR-TR-2011–49, Microsoft Research. http://research.microsoft.com/apps/pubs/default.aspx?id=147663
Valli, Alessandro
2007 Natural Interaction White Paper. (http://www.naturalinteraction.org/images/whitepaper.pdf; Retrieved 01. 05. 2011).
Vanhoof, Koen and Ina Lauth
2010 Ubiquitous Knowledge Discovery Applications. In: May Michael and Lorenza Saitta (eds.), *Ubiquitous Knowledge Discovery*, Berlin: Springer.
Vasilakos, Athanasios and Witold Pedrycz
2006 *Ambient Intelligence, Wireless Networking, Ubiquitous Computing*. MA, USA: Artech House Press.

Waltinger, Ulli and Alexa Breuing
2012 Internet-based communication. In: Alexander Mehler, Laurent Romary and Dafydd Gibbon (eds.), *Handbook of Technical Communication*. Berlin/ Boston: De Gruyter.

Wegner, P.
2010 What is Computation? – The Evolution of Computation. In: *Ubiquity Symposium*. (http://ubiquity.acm.org; retrieved on April 1, 2011).

Weiser, Marc
1991 The Computer for the Twenty-First Century In: *Scientific American*, 94–100.

Weiser, Marc, R. Gold and J.S. Brown
1999 The Origins of Ubiquitous Computing at PARC in the late 1980s In: *IBM Systems Journal* 38(4).

Weiss, Malte, Florian Schwarz, Simon Jakubowski and Jan Borchers
2010 Madgets: Actuating Widgets on Interactive Tabletops In: *Proceedings of UIST '10*. New York: ACM Press, 293–302.

Wimmer, Raphael, Michael Kranz, Albrecht Schmidt and Sebastian Boring
2007 A Capacitive Sensing Toolkit for Pervasive Activity Detection and Recognition In: *Proceedings of the Fifth IEEE international Conference on Pervasive Computing and Communications (March 19–23, 2007). PERCOM*. IEEE Computer Society, 171–180.

Zittrain, Jonathan
2008 The Future of the Internet – And How to Stop It, Yale University Press, Penguin UK/Allen Lane; Oxford Legal Studies Research Paper No. 36/2008. Available at SSRN: http://ssrn.com/abstract=1125949. Retrieved on April 1, 2010.

22. P2P-based Communication

Gerhard Heyer, Florian Holz, and Sven Teresniak

1. Introduction

1.1. Real-world social networks and technology-based networks

Human communication is essentially based on reaching each other. From a structural point of view, this can be achieved by one of the following three main paradigms. First, there is one-to-one communication where one human being directly communicates with another one like in a face-to-face talk. Then, there is one-to-many communication where one human being addresses a group of other human beings. This is e.g. the case when a teacher talks to his pupils in a class. And, finally, there is many-to-many communication where almost everyone taking part in the communication also sends and receives messages.

All these forms of communication naturally take place directly by talking, just using the air to transmit the messages from the speaker to the hearer(s). But they can also involve technical devices. Telephoning usually is a one-to-one communication, whereas print magazines are instances of a one-to-many communication, and the use of a classic bulletin board constitutes a many-to-many communication.

Using networked computers as a means of communication, we again meet the main paradigms of communication. An ordinary website can be considered a one-to-many communication. Downloading a web page or software is a one-to-one act of communication. And a discussion in a web forum usually resembles the many-to-many paradigm. Web-based social networks like Facebook and Google+ resemble a many-to-many communication on the message level (for these notions and applications see Diewald (2012) in this volume). On the technical level, however, they are server-based services where the users connect to the servers via the web interface.

Peer-to-peer (P2P) networks are computer-based networks and follow the many-to-many paradigm in contrast to many other applications of the WWW (cf. Section 2). These networks are very useful for tasks such as content delivery, communication networks, and tracking and verifying transactions (cf. Section 1.3). In this chapter, we will present the technical aspects of P2P communication and P2P networks. Whenever we speak about nodes/vertices and edges, we take a (graph) theoretic perspective and request the reader to think about the P2P network as a graph, where nodes are peers (clients, users) and the edges are acquaintances, that is, two peers are connected when they know each other.[1] On the other hand, when we speak about peers and neighbors, we take a more technical view in order to discuss practical issues like bandwidth consumption, traffic generation, churn etc. that any P2P software has to deal with. Both views are highly connected, though, as a vertex in our P2P graph represents a peer in a real world peer-to-peer information retrieval (P2PIR) network.

Networks in which peers choose neighbors according to semantic criteria are called *semantic overlay networks* (amongst others Holz et al. 2007 and Crespo and Garcia-Molina 2005).

1.2. Peer-to-peer information retrieval

A main task in every P2P setting is to find the right peer(s) for the specific tasks e. g. to find people from the same school who share the same interests, or to find peers that contribute files in a content-delivery network. In general, this means to find one or more peers who offer a certain content to the network, or have certain properties. The search interest can be compared to the well-known web search interest for web sites. However, for P2P systems efficiency and effectiveness of the search algorithm (and therefor the coverage of the results) are crucial. As search is an issue in information retrieval, any search using P2P systems is called P2P information retrieval (P2PIR).

P2PIR systems are an alternative approach to the centralized services nowadays found in the world wide web. Most of the information sources of the Internet are based on the client server model where a server provides information and many clients just consume this information (this even holds for user-generated content). Nevertheless, there is no need to organize the search client-server-based as well. Even for very "classical" settings with relatively few content servers, the search can be done using a P2PIR system where the search index is distributed over the P2P network and is not stored at a centralized web search server (cluster). In the more generalized setting where almost every user also provides content the same P2PIR approaches can be applied as well.

Most common applications of P2P systems for content-delivery (cf. Section 1.3) operate in a setting where there exists a high replication of almost every part of content which makes it easy to find the requested part of content among one of the many peers. In contrast to this, P2PIR deals with the more general assumption of low replication which means that almost every part of content only exists once in the network, and hence for each content there is only one peer providing it.

In what follows, we shall focus on the information retrieval aspects to deal with the crucial issues of network building and structuring, peer profiling, query routing et al. (cf. Section 3 and 4).

1.3. Common applications of the peer-to-peer technology

There are many applications of P2P networks. The best known ones are in the area of content-delivery, but P2P systems also are well established in the field of communication networks and web transactions.

Applications in content-delivery account for the largest portion of internet network traffic. Different approaches highlight various aspects of the many possible implementations. Freenet is a P2P network which focusses on the freedom of communication and thus aims to prevent censorship (cf. Section 3.2). Bittorrent is a protocol designed for the efficient distribution of arbitrary content at large scale, regardless of whether the content has one original source or many. Edutella is a P2P network with a fixed ontology-based structure and is used for searching semantic web metadata. It is not aimed to directly share the content, but the content description (metadata), and thus makes every categorized content available via semantic web standards using RDF[2].

Communication networks like chats (IRC, Jabber, Skype – see also Waltinger and Breuing (2012) in this volume) mainly use manually created connections (contacts). The contact data often have been exchanged previously via email or other communication channels. These networks rely on direct communication between peers and do not make use of automatic network building and structuring, or routing. Usually users connect to the network via login and a relay server.

Bitcoin is a P2P-based network for tracking and verifying transactions. In the current implementation, these transactions are used to relay payments, and Bitcoin thus is perceived as a digital and decentralized currency.[3]

1.4. Summary of the chapter/Chapter overview

P2P networks resemble a many-to-many communication at the technical level in contrast to the client-server-paradigm. This will be explained in more detail in Section 2. As P2PIR is the most difficult, yet interesting kind of P2P environments, that also well illustrates basic assumptions and consequences of P2P systems in general, the main focus in this chapter will be on P2PIR (cf. Section 3). In particular, we shall also discuss the possible contribution of semantic technologies to P2PIR and, by way of example, present the Sem-PIR projects as an instance of P2PIR systems based on semantics when we consider several technical aspects of P2P communication in Section 4.

2. Servers and clients vs. P2P

There are two main approaches to organize distributed communication systems: the client-server model and P2P-based approaches.

The client-server model contains two kinds of computers, *clients* and *servers*. Clients are consumers of services that are provided by servers. These services can be computing services as well as content resources as, for example, print server, file server, web services and archive storages. Usually the servers are of high-performance while the clients can be just low-performance end-user terminals. To use the services provided, every client only needs a direct connection to the respective server(s).

In a P2P-based organizational scheme there is only one type of computer, for simplicity called a *peer* (sometimes also *servent*), which fulfills both functions: providing and requesting services, for example, communication (chat), filesharing, distributed computing. All peers can have comparable computing power and can access many other peers directly.

The notion "P2P systems" is a collective term for a decentralized paradigm of information exchange. "Peer-to-Peer computing is a network-based computing model for applications where computers share resources via direct exchange between the participating computers" (Barkai, 2002). Currently P2P systems are mainly used for file sharing. In the last years, well known exponents have been, amongst others, Gnutella, FreeNet, and NeuroNet. They mainly differ in the way the queries are processed and routed.

There are different mixed forms of these models. Many servers of similar computing power can work together in a P2P way to better serve a large number of clients. The servers can also be organized hierarchically instead of a flat network. With respect to differences in computing power there also are P2P networks that make use of server-like so-called super-peers that have more functionality than a normal peer as, for example, controlling the network structure of a part of the P2P network.

The main advantages of P2P-based systems over centralized systems are:

Scalability The more clients join a network, the less computing power of the servers can serve every specific client. In a P2P network, every new client also brings additional resources into the network. So, P2P networks are scalable. How the network structure and routing can be tailored to specific circumstances is explained in Section 3.1 and Section 4.2.

Robustness As every peer contributes to the available computing resources, and not only a few servers, the malfunction of some peers and the index of available resources do not affect the network and its functionality as a whole. There is no single point of failure in a P2P network.

Availability Based on the dynamics of the decentralized indexing of the available resources, the retrieved information concerning available resources is very up-to-date. A server-based index usually is updated by re-crawling already known resources, what can only be done from time to time for economic reasons.

Data Safety Every peer provides his own resources, and the index is stored in a distributed way, so, malicious attacks and network-wide censorship either on the resources directly, or on the index, have no single point of access. Every access to the network – either by being a peer, or manipulation of network traffic – is strictly limited to the locally stored and transferred data.

Easy Publishing As every peer can provide resources in a P2P network there is no need to set up an extra server, or to register at, or join, a publishing authority.

3. P2P-based Information Retrieval (P2PIR)

3.1. Background and approaches

The file sharing scenario and the information retrieval scenario as the main applications of P2P systems differ in a substantial way. While in the file sharing scenario many copies of every file distributed throughout the whole P2P network are kept, in the information retrieval scenario (P2PIR) there is usually only one instance of each resource, for example, documents like websites or PDFs. So, to find a specific file in the file sharing scenario, in most cases it is necessary to only search a small part of the network, whereas in the information retrieval scenario, to find a specific document matching a query it must be processed and routed through the network in a way that potentially every peer must be visited in order to retrieve relevant documents from this peer.

In general, these are the main aspects of a P2P system: How are the peers connected, and how are queries routed through the network? Both aspects determine the *costs* of a P2P system in general, and a P2PIR system in particular. If every peer would store the complete information about every other peer in the network, queries could be routed directly to those peers that provide the documents best matching the query. However, it is expensive to store many neighbors (directly reachable peers, edges of the network) and information about their content (profiles, respectively part of the index). Further, flooding, that is, passing a query to all neighbors of a peer, could be used as a simple routing strategy to search the whole network in order to retrieve an answer to a query. But flooding also is expensive. Thus, the challenge is to construct a network with a minimum of edges while providing enough structure to route queries directly into every specific part of the network.

Comparable to the large amounts of protocols used by services based on the client server model, many different approaches exists in the field of P2PIR systems in general. In order to best serve the purpose of a P2PIR system, numerous methods and algorithm exist in the literature to deal with the characteristics of a P2PIR network structure.

A common feature of most P2PIR systems is the fact that each peer in a network only has local knowledge about the available documents and other peers. In contrast to centralized approaches, in P2PIR a peer neither knows which, or how many, documents are contained in the entire network, nor does a peer know how many other peers exist.

In general, every peer that processes a certain query can only answer to this query based on his own documents and his local knowledge about the network structure involving other peers he may know.

When the relevant documents a user may be searching for are distributed over many peers in a network, there is a high possibility that these peers will not be directly connected, and thus that they don't know anything about each other. A clever routing algorithm basically has to direct queries to as many peers as possible that possess relevant documents, even if one would have to also visit peers without matching items. Furthermore, on the search path to peers with relevant documents as few as possible other peers should be bothered. Obviously, without global knowledge about peers and their connectivity, efficient routing is not a trivial task.

There are two approaches to organize peers in a P2P network: *structured* approaches and *unstructured* approaches. In an unstructured P2P system, every peer provides and takes care of his own resources, and can choose his neighbors without constraints. To answer a query, the requesting peer asks all the peers he knows (his neighbors). They reply if they have relevant resources, and they forward the query to all their neighbors, too, until the *time-to-live* (TTL) of the query has expired. This kind of broadcast is also called *flooding*. File sharing P2P systems in general are unstructured. By default, the peers in these networks are randomly connected, although sometimes *superpeers* (see Section 3.2) are used to optimize routing. Flooding is employed with a very short TTL to limit the exponential growth of network load induced by flooding. Since for every file there are many copies, and in most cases at least some of them are located at a nearby peer, this approach reaches a high recall although the matching files have only been searched for in a very small part of the network. In a structured P2P system, every peer stores a clearly defined part of the index of all available resources, and every peer knows exactly how to reach any other peer, directly, or indirectly via intermediate other peers. So, queries can be routed directly to the peer who keeps the relevant resource information in his part of the index. The most common type of structured P2P systems are *distributed hash table*s (DHTs, see Section 3.2).

3.2. Superpeers and structured networks

Concerning the P2P approach for information retrieval purposes, the central issue is the high costs for search in the network and the high network load, both of which are caused by the use of broadcasting for routing queries through the network (flooding). Flooding causes serious problems concerning the scalability of the P2P approach of Ritter (2002). To solve this problem, so-called ultra- or superpeers have been suggested by Singla and Rohrs (2002). Peers with a higher capacity than others – either regarding network bandwidth and/or computing/content resources – can be treated as ultrapeers. Queries are first routed to ultrapeers who either reply to a query more likely, or forward the query to another more appropriate peer. This reduces the amount of messages needed per query (forwarded queries) as well as the average response time. By introducing the concept of ultrapeers, however, the principle of equality of all participants in the network (peers) is violated.

In addition to using manually created metatags, Schlosser et al. (2003), Decker et al. (2002) and Nejdl et al. (2002) present another approach that reduces the costs of broadcasting to its theoretic minimum: the hypercube. A hypercube is built by adding new peers that join the network on the free corners of the overlay hypercube structure. If a hypercube is complete and a new peer joins the network, then a new dimension is added to the overlay structure that provides new free corners for new peers. Broadcasting can be done along the dimensions of the hypercube so that every peer receives every query only once. This structure is the opposite of an ultrapeer network, or a P2P system with centralized or hierarchical elements. Here, every node is completely equal, provides the same functionality, and carries exactly the same load. In reality, however, this is often not the case. Many computers in a network differ in resources like bandwidth or documents. Some peers store and provide a large number of documents while many others provide only very few.

The common feature of all DHT approaches is their deterministic data placement strategy: in systems like Chord (Stoica et al., 2001), Tapestry (Zhao et al., 2004) or CAN (Ratnasamy et al., 2001), the location of files is not arbitrary, but is usually determined by applying hash functions to file keys, that is, peers are responsible for certain ranges of hash values and have to store all keys (and associated files) that yield these values when being fed into a hash function. Of course, the same function can be used to find data, which makes routing very easy.

In effect, data can be located quickly and efficiently. However, some serious drawbacks result from these rigid constructions:

– Since each peer is responsible for providing files with certain hash values, peers joining or leaving the network call for additional efforts, involving costly partitioning or replication of data on other nodes.

– Load balancing: A lot of traffic will be directed towards nodes that are responsible for hash values of very popular data items. Because of the rigid data placement strategy, it is very difficult to introduce any form of load balancing.
– DHTs were designed to provide search functionality for single keys (e.g., file names). Finding files associated with $n > 1$ keys requires n lookup operations (one for each key) and subsequent intersection of the result sets. Searching for more than one key is, however, a typical use case in information retrieval and should therefore be handled more efficiently.

Obviously, DHTs are not really suitable for information retrieval purposes (especially because of the last of the drawbacks just mentioned). Therefore, we will not present their functionality in detail, but rather consider P2P architectures that allow for arbitrary data placement, transient peer populations and a search for more than one keyword.

The FreeNet system uses a unique key for every document in the network (Joseph, 2002a). Every peer stores a database containing a part of the relationship of the keys and his neighbors. Underlying is a similarity measure that ensures that similar documents are likely to be stored on the same computer. This is comparable to the approach of NeuroNet (Joseph, 2002b). These system clearly depend on the quality of the similarity measure.

3.3. Hybrid networks

Recent development in the field of P2P systems have headed towards new usage scenarios (Steinmetz and Wehrle, 2005) other than the "classic" applications like file sharing and computer supported cooperative work. The construction of networks based on semantic neighborhoods not only is suitable for the information retrieval task itself, but can also serve to discover and structure content-oriented communities. Another important development is the combination of the P2P approach with principles of the Semantic Web (Staab and Stuckenschmidt, 2009 see also Waltinger and Breuing (2012) in this volume).[4]

Both approaches try to encode the content and semantics of textual resources, and to generate appropriate representations that are maximally independent from lexical phenomena in order to better serve content-oriented information needs. In most cases, however, the ontologies employed by the Semantic Web still have been constructed manually in order to achieve a sufficient quality.

4. The SemPIR projects

Structured networks provide the ability of optimal recall, that is, a query can reach every peer in the network with the theoretically minimal amount of messages. But this advantage comes at a price: the maintenance of the network is expensive (in terms of messages needed). Every joining or leaving of peers generates traffic to keep the network topology intact.

It suggests itself that a combination between structured networks and unstructured networks may lead to an approach that combines their advantages with just a slight impact of their disadvantages. The question is how to provide cheap and easy network maintenance, but at the same time create enough structure in the network topology to allow for sophisticated routing algorithms that exploit this structure in order to avoid unnecessary network traffic. Every P2PIR system has to provide a solution to this challenge in one way or another.

In what follows we shall sketch some of the specific problems that arise in the process of designing such a P2PIR system by way of reference to the P2PIR system *SemPIR* (semantic P2PIR) as an example.

The goal of the research projects called SemPIR I and SemPIR II (Sem-PIR in the following) was to build a content-based P2P overlay network that allows for fast document retrieval in an efficient manner. The main characteristics of SemPIR are

Content-based network structuring The connectivity between peers is not fixed and not predetermined, but also the linkage is not random. Instead of a geometric network topology, the structuring process is based on the documents owned by the peers, more precisely, it relies on the semantic similarity of the document collections possessed by each peer. The network is thus a *(self-organizing) semantic overlay network* (see Garcia-Molina and Crespo 2003 as well as Crespo and Garcia-Molina 2005 for a formal definition). Joining and leaving of peers is cheap, and so also is network maintenance.

Use of social networks The *small-world phenomenon* is being used to reduce network load by establishing networks with small diameter, that is, every peer is connected to every other peer by just a few hops. Thus, peers with relevant documents tend to be near to each other, or know each other.

Intelligent query routing Query routing and peer selection is based on semantic similarity as well. No flooding is used. Given a query, only peers that promise good results are selected for routing.

Robust and scalable The network scales well in size of peers and is robust against partitioning.

Each point will be described in detail in the following sections. In what follows we shall first deal with the issue representing peers and their "content profiles" and then will explain how to build a self-organizing content-based P2PIR network structure in Section 4.2. In Section 4.4 the routing of messages and the different kinds of messages that can be used will be presented. We conclude by discussing evaluation issues in Section 4.6. However, while SemPIR is always used to exemplify the development of this kind of P2P information retrieval system, the problems discussed in this section are generic to P2PIR in general, and SemPIR is our prime example.

4.1. Profiles

For representing the content of documents the well-known *Vector Space Model* (VSM) is used that became popular through the work of Gerard Salton (see, e.g., Salton et al. 1975) and later was generalized by Wong et al. (1985). In the VSM each document is represented in an n-dimensional space where each dimension represents a concept (typically a term) weighted by its importance for the particular document. Thus, each document is represented as a point in Euclidian space and hence can be compared by Euclidian metrics like cosine similarity or Euclidian distance. The cosine similarity calculates the cosine of the angle between two (document) vectors. Thus, very similar documents get a cosine similarity value near 1.0 while very dissimilar documents get a value near or equal to 0.

While the documents a peer owns are presented by (term-)vectors, peer profiles, and the queries a peer sends to other peers in the network are made up in exactly the same way: Peer profiles are calculated by aggregation (e.g., summarization) of document profiles, queries are simple document vectors with weights equal to zero in almost every dimension except those dimensions representing the terms searched for (as specified by the user). As is usual in systems using the VSM, the similarity between a query and a document, or between a query and a peer profile, can be calculated by using the cosine similarity as described above, or by any other metric. As each dimension in such a term vector generally represents a term, and usually only a fraction of all terms in the document collection (or in the P2P network) that actually occur in a particular document (query, peer profile), dimensions with a weight of zero are suppressed, that is, a vector is represented as a mapping from terms into weights greater than zero. From a technical perspective this helps to reduce the memory footprint of the data structures used and to save bandwidth while transferring these vectors. For example: even if a peer owns many documents with millions of unique terms, the query to submit to search for the terms “P2P” and “information retrieval” is represented by a map like {P2P → 1.0, information retrieval → 1.0} that can be encoded in just a few bytes in length.

With the ability to efficiently calculate the content-based similarity between queries and peers as well as between queries and documents, it is possible to develop routing strategies to calculate the way a query has to pass through a network in order to find those peers with the most suitable document collection to satisfy the query (and thus the user's need).

4.2. Structure building

The aim of the SemPIR project[5] was to combine the benefits of structured networks with the benefits of unstructured networks. In order to achieve this, a P2PIR system was developed by analogy to (human) social networks (but unlike to Facebook) where agents that share interests find together and keep in contact. The algorithms for network structuring, how to arrange the peers, and how to route queries in the graph will be described in detail later. First we briefly give some background on the approach.

An advantage of structured P2P networks (by means of hypercube-based approaches in the following, cf. Section 3.2) is the very efficient search by broadcasting messages. Every peer receives every message by a minimum (optimal) amount of messages generated within the network. In contrast to this approach, the strength of unstructured networks lies in the easy and cheap join and departure of peers into and from the network. There are many approaches to find a compromise between these extremes (see Section 3.3). To bring together the different (sometimes conflicting) demands on a modern scalable P2P system, SemPIR was developed as a *semi-structured* P2P network, that is, the topology of the network is neither predetermined and fixed, nor is the connectivity between peers generated at random.

One general goal of the system was to minimize the network load for basic operations like the joining and departing of peers as well as searching for items (e.g., files) within the network. Further, as the network must not fall apart or collapse over time, it has to be robust and scalable.

4.2.1. Hubs & authorities

To efficiently achieve this goal, SemPIR tried to imitate the way humans tend to find experts (of the desired thematic focus) in their social environment. If someone needs help to solve a problem by means of detailed information, she can ask the person that most probably provides this information directly (i.e., the *expert*) or she can ask a person in the role of a *broker* who most probably knows an expert. If neither an expert nor a broker is available, one can, thirdly, ask a person who has many social contacts in the underlying social network (i.e., a highly connected *hub*). By doing so the chance of finding an expert is increased because of the extended number of persons that can be contacted. Note that a hub in a social network is a person who *knows* many other persons (Kleinberg, 1999a,b). This does not necessarily mean that the hub is also known by many other people: the hub is not necessarily a popular person. Usually, social connections like friendship are mutual and thus a friendship graph (where nodes are persons) is undirected. But many other social relations can be directed. If a person *is known* by many other people, we call this person an *authority*. Popular actors are known, for example, by many more people than they know themselves.

The notion of *hubs* and *authorities* was formalized by Jon Kleinberg in the context of hypertext systems (Kleinberg, 1999a). At the same time, Sergey Brin and Lawrence Page developed the famous PageRank algorithm that strongly relies on the connectivity of websites and the distribution of incoming and outgoing links in the web graph (Brin and Page, 1998).

Like many approaches to P2PIR, SemPIR adopted the latter search for experts based on the notion of hubs and authorities. Furthermore, it also builds on small-world characteristics of P2P networks. This is described in the next subsection.

4.2.2. Small worlds

According to Watts and Strogatz (1998) small-world networks are graphs with certain graph characteristics that have

- a relatively short average path length, that is, the distance (hops) between two random nodes is – on average – short;
- a relatively high cluster coefficient, that is, there is a high probability that for a certain node the neighbors connected to this node are connected, too. Thus, the network tends to build clusters – highly connected regions with only few links between these clusters;[6]
- a relatively low density, that is, there are considerably less edges in the graph than defined by the theoretically maximum $\frac{1}{2}|V|(|V|-1)$ where $|V|$ is the number of vertices in the graph.

There are many graph-like patterns in nature or sociology that have the small-world property (Watts and Strogatz, 1998; Newman, 2003). The most famous finding in this area is an outcome of the letter experiment by Stanley Milgram in 1967 (Milgram, 1967; Travers and Milgram, 1969). In this experiment (followed by comparable experiments by other researchers), Milgram studied the connectivity of social networks. He found that on average every person in the USA is connected to any other person by only 5.5 hops. By definition, a connection was established when two persons know each other on a first name basis. The so-called small-world phenomenon[7] – every person is connected with any other person in the network by on average six hops – is somehow surprising, because six hops establish a rather short path between any two persons.[8]

This small-world characteristic of a short average geodesic distance can be explored when implementing an efficient routing algorithm for queries in a P2P information retrieval network. Every node is just a few hops away from every other node. Duncan Watts and Steven Strogatz studied the creation of graphs with small-world characteristics as a transition state between *regular graphs*[9] and *random graphs*[10] of the same order (i.e., number of vertices) by randomly 'rewiring' edges (Watts and Strogatz, 1998). They found that if the edges of a regular graph are rewired with a certain probability p, the average geodesic path length (of the graphs resulting from this rewiring) decreases much faster than the cluster coefficient the higher the value of p. Thus, in the Watts-Strogatz model, graphs that are located in the range of regular and random graphs, exhibit the small-world property by showing small average geodesic distances in conjunction with high cluster values. However, knowing that a graph almost always has a short path between any two of its vertices is not very helpful, if this path cannot be found efficiently. A peer in a small-world P2P network, for example, does not necessarily know to which neighbor to send a query in order to follow the shortest path to some target peer. Jon Kleinberg investigated this problem and coined the notion of *navigational clues* that have to be present at every node in the graph (Kleinberg, 2000). These clues can represent topological knowledge about distances between nodes in the graph.

In SemPIR, the navigational clues are based on the content of the node's documents as subsumed in their content profiles (this is explained in Section 4.1).

4.3. Building the network

The concept of small-world networks containing hubs and authorities can now be used to loosely organize a P2P network. From a graph theoretic point of view, the vertices (peers) in the graph are arranged in clusters, that is, the nodes are assigned to separate subsets. Each cluster is formed by nodes containing semantically related documents. To build the network, different strategies with respect to exploiting semantic similarity between peers can be followed. The connections within semantically similar clusters are called *cluster links*.[11] The nodes within these clusters are highly connected, similar to social networks in which acquaintances of the same person probably know each other. As a consequence, the network exhibits a high cluster value together with short path lengths. Further, there exist so-called *inter-group links* between semantically coherent clusters. The latter links are to some degree random as they reflect content-related dissimilarities among peers. Their purpose is to keep the graph connected such that its giant component contains all vertices; they establish shortcuts for retrieving documents about certain topics even if the focal peer and his neighborhood do not provide any document about these topics. Technically speaking, these two kinds of links are separated in order to prevent for example that they overwrite each other. However, by viewing the network as a graph, there is no difference between the links. As links between peers are established by the peers themselves and not by any global instance (as, e.g., in Napster), SemPIR can be considered as a *self-organizing* network.

A third kind of neighborhood relation is called *egoistic links*. It establishes smaller neighborhoods than in the case of the neighborhoods described above. In this list, peers are stored that frequently answered a query. In other words, if a peer was able to answer a lot of queries in the past, it seems natural to contact this peer in cases were no other peer can help, that is, if neither the cluster links nor the peers connected via intergroup links help to match the focal query. Hence, the egoistic neighbors provide a fallback solution if a query is not assignable to any of the peers' profiles stored in the peers' neighborhood lists. The list of egoistic peers is continuously updated over time by analyzing the result sets of queries passing a peer. Following Jon Kleinberg, these *egoistic neighbors* can be called *authorities* (see Section 4.2.2).

Based on the considerations outlined so far, a peer can reach other peers within his neighborhood as based on their semantic profile. Moreover, each peer can use long range links to quickly reach other clusters in other regions of the network. In both of these neighborhoods, a peer not only knows the addresses (IPs) of other peers, but also their semantic profile.

If a peer joins the network by simply connecting to a random peer already connected, the joining peer sends a *gossiping query* into the network to achieve the following:

- The gossiping query contains a query consisting of the peer's own profile. That means the peer searches peers that are semantically similar to his own content. Other peers answer with their own peer profile. By analyzing and ranking the query results, a peer selects the most similar peers and puts them in his neighborhood list.
- While the gossiping query works in both ways, other peers can analyze the peer profile the query contains and adjust their own content-based list of neighbors. In this way, a gossiping query *pushes* profiles into the network.
- This process of gossiping is repeated periodically in order to deal with churn, to respond to changed peer profiles, and to successively increase the quality (similarity) of the content-based neighborhood.

Networks in which peers choose neighbors according to semantic criteria are called *semantic overlay networks* (amongst others Holz et al. 2007 and Crespo and Garcia-Molina 2005).

4.4. Routing

After having established a semantic overlay network with small-world characteristics, this network can be exploited for efficient query routing. As mentioned above, the aim is to reduce the costs of searching while not making use of broadcast messages (flooding). Instead an *informative search* is used that is based on semantic routing.

Queries are compact vectors representing the information need of a user that are intelligently routed through the P2P network. On every peer where a query is routed through, the most similar document vectors of documents that the peer keeps are extracted at first, if there exist any such documents. With the document descriptions and the location of the documents appended the query will now be forwarded to the most appropriate peer in the neighborhood list of the peer. That is, the query will be compared to each peer profile and the peer with the highest semantic similarity is selected for the next hop. As long as the time-to-life (TTL) parameter of the query is greater than zero, the query will be forwarded to the next peer. Every peer reduces the TTL by one in the process of forwarding a query. The TTL prevents queries from circling through the network endlessly. A TTL of zero stops the routing process and causes the last peer to send the query (containing all aggregated results) back to the peer the query originated from.

A query log consisting of all visited peers helps to prevent circles in routing. Thus, no peer will be visited twice. Since peers are organized into clusters of high similarity, a query gathers good results once it reaches a matching cluster. The (random) intercluster-links of each peer help to find those clusters fast by reducing the graph diameter.[12] In this way, the routing uses content-based hill climbing: once a cluster is found, each peer in this cluster knows other peers similar to the query.

If a query is returned back to the sender, all documents in the result log are ranked according to semantic similarity and presented to the user. The user now can choose one or more documents to download from one or more peers.

4.5. Caching

To further decrease the network load, peers can use caching algorithms. Results that have been generated for certain queries can be stored for later use. The cached elements (queries and result documents from other peers) help to improve the routing algorithm by sending queries not only to peers with high similarity in the peer profile, but also to those peers who own matching documents. By not answering a query from cached data, no documents are displayed to the user that don’t exist in the network anymore.

4.6. Evaluation issues

The evaluation of P2PIR systems is difficult. Most of the current approaches are evaluated not by using real life systems but in laboratories using simulations. In order to get realistic results, every simulation approach tries to set up simulation parameters with the maximum amount of data gathered in real world systems. Unfortunately, however, until today no real P2PIR system exists that is used by a user base large enough to gather this kind of data needed. Thus, approximations and heuristics must be used where real data are missing.

The following list exemplifies network parameters that could influence the outcome of a simulation:

Latency & Bandwidth The amount of bandwidth a peer can use to send data is limited and often varies over time. Therefore, the amount of data (e.g., queries) a peer can process in time is limited. Furthermore, network connections between peers to some degree are not reliable. The upstream and the downstream peers are not equal in many cases. The simulation of these or other low-level network parameters is difficult because of their complexity.

Churn & time-dependent parameters A user can establish a connection to the network or leave the network. This behavior over time is distinct for every user and thus the simulation of this behavior is complex. It is also difficult to simulate the (offline) arrival of new documents on any peer and furthermore to simulate the time, content and amount of queries a simulated peer has to start to act realistically.

Collection Without any real world data about document collections of users it is difficult to sample document profile over peers in order to provide a realistic initial state of the network. Analyses of P2P file sharing systems cannot deliver this kind of data because file sharing systems are not content-based.

4.6.1. Test data

Depending on the particular aspects of the system that has to be evaluated, many different approaches exist in the literature on how to design simulation scenarios as realistic as possible. Gold standards for centralized information retrieval (like TREC) cannot be used for P2PIR, because of the distributed nature of documents and queries in P2PIR. While centralized IR deals with the problem to choose the documents best matching a given query from a given document collection, in standard P2PIR systems each peer only knows a fraction of the document collection distributed over hundreds or thousands of peers. Further, queries are distributed over peers. Gold standards for centralized IR do not provide this information either.

The task of distributing documents in a network of peers can be performed in many ways. In the following subsection, we sketch some characteristic examples (for more examples see Holz et al. 2007).

In Lu and Callan (2003), web pages are used as a test collection and the prefixes of their URLs define the peers. Lu and Callan use the TREC WT10g web test collection[13], together with queries generated from the documents. Evaluation is done by comparing results to that of a centralized system using plain precision and recall. In another approach is provided by Herschel (2007) who uses Wikipedia in order to distribute documents in conjunction with a term-based retrieval algorithm.

An approach based on statistics is proposed by Cooper (2004, 2005). Histograms over the counts of relevant documents per query, the degree of document replication, and the mean count of relevant documents for a query on one peer are used to generate an artificial testbed that refactors real statistical relations.

Table 1. Small world characteristics for a simulated SemPIR network of 7,500 peers in size. During the structure building process the average path length increases slightly while the cluster coefficient decreases one order of magnitude.

parameter	simulated	random network	quotient
Cluster coefficient	0.181	0.0082	22.1
Average path length	3.443	2.898	1.19

Heinrich et al. (2006) propose latent concept analysis to estimate relevant parameters for a distribution of topics over peers and documents, and use these parameters to distribute another document collection over peers in the same (statistical) manner. In effect, the statistics of a P2P network can be scaled up to a much bigger set of documents and peers.

4.6.2. Simulation

Given a set of documents, the building of a network structure and the routing of queries can be simulated. For the SemPIR projects, a basic simulation with synthetic data was carried out using the OMNeT++ network simulator (see Witschel (2005) for details). Real world data from the OHSUMED corpus[14] (among others) were used in a second simulation run. OHSUMED consists of nearly 350,000 medical abstracts, annotated with on average 10.6 so-called *Medical Subject Headings* (MeSH) terms each. Each of the 14,596 MeSH terms in the collection was treated as a peer and every abstract was assigned to all peers corresponding to his MeSH terms.

Now, with the definition of peers and a distribution of documents over peers, the simulation of *structure building* demonstrated the emergence of small-world characteristics out of a sparse random graph where every peer initially only knows three other peers. The simulation of searching in this network was carried out separately and will be explained later. The results of the structure building simulation are shown in Table 1. Note that the cluster coefficient cannot reach high absolute values, because the amount of neighbors is limited to 30 in this simulation, inducing a limitation of links within semantic similar clusters.

Given the generation of a small-world P2P network based on content similarities, the next simulation showed that this network structure actually helps in terms of query routing and decentralized document retrieval. As the aim of the project neither was to optimize document retrieval nor to build ranking algorithms, the evaluation focussed on the effects of local and global knowledge. Every document got its relevance value (regarding each query/MeSH term) from OHSUMED, while the peer selection algorithm used the Okapi BM25 retrieval function (Robertson et al., 1992; Robertson and Zaragoza, 2009).

When comparing the P2PIR with the centralized IR, the evaluation of precision is not an issue, because precision mainly depends on the similarity measure (between a query and a searched item), and not on the network structure. For the evaluation of recall, however, we need to consider whether or not a peer has limited knowledge about the network and the peers in that network. Clearly, with a limited TTL – which is much below the number of peers in the network – it is impossible for a query to reach every other peer in the network to retrieve relevant documents. If, for example, 21 peers hold relevant documents but the TTL of a query is 20, the recall of that query can never reach 100 %. Thus, the interesting task was to analyze the difference in recall between a P2PIR system where one peer owns every document (the "all-knowing" peer), and a system in which each peer only knows his own documents and the profiles of peers in his neighborhood.

In SemPIR, three ways of searching in the P2PIR system were set up:

Random walker A query does not make use of peer profiles to determine the next hop in the semantic overlay network. This is the baseline for recall, and it was easily reached by the P2PIR approach (Witschel, 2008).

SemPIR search The network topology and peer profiles are used in order route a query through the network. This is document retrieval with only local knowledge.

Informed search The network consists of one single peer and this peer owns every document. Thus, every query on this peer leads to an exhaustive search result. This is the upper bound for recall and the result set of this search is *per definitionem* optimal.

Of course, the absolute value of recall depends on a query's time-to-live that has turned out to be an important parameter. But also *relative* recall is important for the evaluation, because the focus for P2PIR is on improving the network structure in order to optimize query routing and not on improving the ranking algorithm as such.

4.6.3. Relative recall

Probably most users of search systems (especially web search engines) only put attention to the first few results in a ranked list (amongst others Guan and Cutrell 2007 and Granka et al. 2004). Therefore it is very important to rank the most relevant documents to the first or upper few ranks in the result list. In consequence, a retrieval system does not have to deliver 100 % recall, because a user is reluctant to inspecting hundreds or more result items. For the evaluation of P2PIR systems this means that we can restrict ourselves to relevant documents presented in the first *k* ranks in the result lists. Thus, the retrieval task is not to find all relevant documents (and to rank them), but to find only the most relevant documents.

To deal with this requirement when comparing recall values, a new evaluation measure based on the well-known *mean average precision* (MAP) measure was developed in SemPIR that judges rankings and makes them comparable. Using this new measure, a system that retrieves the top 20 documents ranked highest by the centralized system receives a better evaluation score than a system that retrieves the 200 lowest ranked documents. The measure proposed is called *average relative ranked recall* with the free parameter *k* (ARRR@*k* for short, with *k* the lowest rank still considered, that is, items below rank *k* are ignored). ARRR@*k* penalizes missing relevant documents in a result list as well as a bad ranking (order) of documents (even if all documents within the first *k* hits are relevant). It plays a similar role as the *k* in the *precision at k* measure (*P*@*k*) commonly used in the IR literature[15].

An important design goal and advantage of ARRR@*k* against other ranking measures is that no human relevance judgements are needed. This makes the evaluation of ranking obtained from decentralized IR systems cheap and fast, because no time-consuming and error-prone human judgement is needed. The basic assumption of ARRR@*k* is that the result set of a centralized system (the results of informed search) is by definition optimal. The performance of a P2PIR system is thus dependent on two aspects:

1) The performance of a similarity measure that selects or ranks the most relevant document regarding a given query (local retrieval aspect), and
2) the ability of the P2PIR network to cope with the lack of global knowledge (decentralized network aspect) by efficiently building and using the network's structure and route queries.

With the help of ARRR@*k* it is possible to compare two rankings with no need for human relevance judgements. In fact, ARRR@*k* highly correlates to ranking measures as the *mean average precision* (MAP, correlation 0.92) or precision at *k* (correlation of precision P@10 is 0.91). Although ARRR@*k* cannot determine the *absolute* retrieval performance of a decentralized system, it can help to optimize a P2P network structure and query routing algorithms.

Let $C = (c_1, \cdots, c_m) \in T^m$ be the ranking of the centralized system, where T is the set of all documents. We assume that the user has specified how many of the top-ranked documents should be retrieved; we call this value k. Further, let $D = (d_1, \ldots, d_n) \in T^n$ be the ranking returned by the distributed system with $n \leq k$ ($n < k$ means that the distributed system retrieves less than k documents in total).

Next, we introduce a function m_D that, for a pair of documents, returns 1 if the first document is ranked ahead of the second within the n-tuple D, else 0:

$$\begin{aligned} \mathrm{m}_D &: T^2 \to \{0,1\} \\ \mathrm{m}_D(c_j, c_i) &= \begin{cases} 1 & \text{if } \exists d_q \in D : d_q = c_j \wedge \exists d_p \in D : d_p = c_i \wedge q < p \\ 0 & \text{else} \end{cases} \end{aligned}$$

With this new function, we define

$$\mathrm{ARRR@}k(D,C) = \frac{1}{\min(k,m)} \sum_{i=1}^{m} \mathrm{m}_D(c_i, d_n) \frac{\sum_{j=1}^{i} \mathrm{m}_D(c_j, c_i)}{i}$$

Figure 1. The ARRR@k-algorithm of Witschel et al. (2008).

Using the ARRR@k measure and a fixed k, parameters could be optimized and promising results have been achieved compared to the centralized system, even if only just a fraction of peers with relevant documents can be visited for a particular query. A detailed explanation and examples are given in Witschel et al. (2008) and in Holz et al. (2007). The algorithm is shown in Figure 1.

Simulation experiments indicate that the amount of peers that have to be visited in order to retrieve the most relevant documents out of a highly distributed collection can substantially be reduced. In comparison to search algorithms using flooding, the network load can also be reduced, and thus the capacity and scalability of the network can be improved. Figure 2 shows some results for two peer selection algorithms: CORI and *by size*. CORI is a resource selection algorithm for distributed information retrieval (Callan et al., 1995), whereas *by size* ranks peers by the size of their document collection.[16]

Note that Figure 2 primarily shows a comparison of measures and thus the absolute values are not quite comparable. The important point is the correlation of ARRR@k with other measures relying on human relevance judgements.

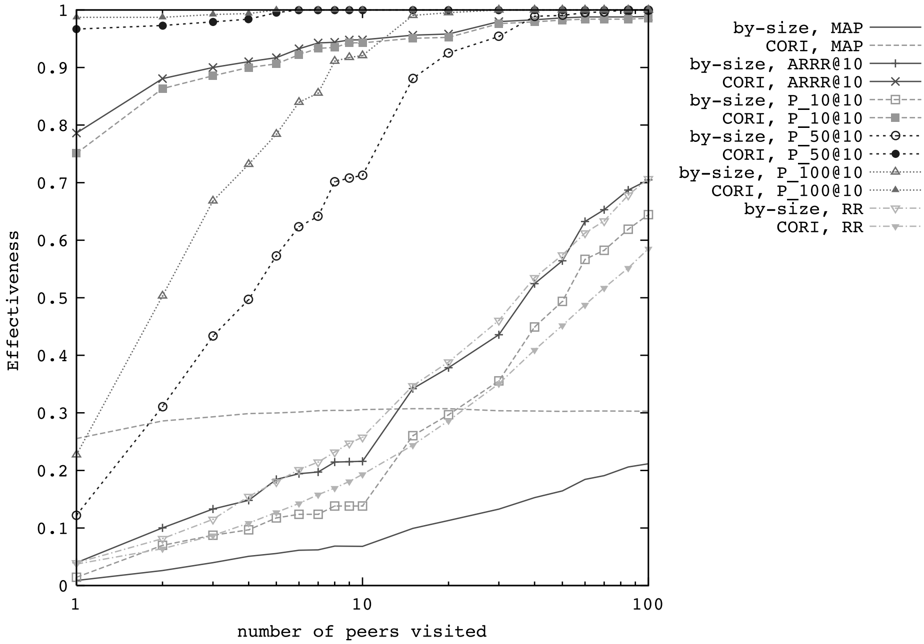


Figure 2. The ranking measure ARRR (*average relative ranked recall*) in comparison to their ranking measures: *mean average precision* (MAP), precision at *k* (P_10, …, P_100), and *relative recall* (RR). See Witschel et al. (2008) for details. Using Spearman's rank correlation coefficient, ARRR strongly correlates (between 0.80 and 0.99) with these other measures, but in contrast to them does not rely on human relevance judgements.

5. Summary and conclusion

We have presented P2P-based communication as a new internet environment for exchanging large amounts of data between users who share similar or related interests. In contrast to the well-known client-server-architecture, the decentralized peer-to-peer-approach (P2P) offers considerable advantages like

- scalability,
- robustness,
- availability,
- data safety,
- easy publishing.

By building decentralized overlay networks on top of existing internet infrastructures, users (called peers from a P2P perspective) are connected in many different ways. Depending on the algorithms used and the purpose of the P2P network, peers can form a network by just using local knowledge about the network itself, that is, without knowing the structure or all the members of that network.

While P2P filesharing networks are most popular for fast and easy exchange of files, the field of decentralized information retrieval aims to apply modern information retrieval techniques in a decentralized manner (peer-to-peer information retrieval, P2PIR). This led to systems where information (mostly files) cannot only be found by using simple metadata like filenames, but also by content – like a semantic footprint of the information itself. Thus, a user can search and retrieve files in a network where he/she does not know the filenames (or part of the name). While P2P filesharing can be regarded as a decentralized file server, P2PIR can be understood as a decentralized search engine.

In the literature there are lots of very different approaches for realizing P2PIR. The so-called SemPIR (for semantic P2PIR) projects were discussed in detail to exemplify some pitfalls, benefits, and constraints a scalable P2PIR design has to deal with.

SemPIR organizes peers in a semi-structured overlay network with small world characteristics. That is, peers with similar content form highly connected clusters within the network and the average path length (hops) between two peers in the network is relatively short. The so-called small world phenomenon can also be found in many natural and social (human-built) networks. We also presented technical details relating to the SemPIR system as well as simulation results.

Unfortunately, P2PIR is not a common, or popular part of today's internet, although the potential benefits of this approach would justify substantial more efforts in this field of research. Moving towards P2P based communication systems, however, would probably require a paradigm shift in the present day information processing with severe implications on the economics of content and the way users share data.

Notes

1. Note that the network topology is spanned by a directed graph: if one peer knows another one, the latter does not necessarily know the first one.
2. http://www.w3.org/RDF/. See also Waltinger and Breuing (2012) and Stührenberg (2012) both in this volume.
3. http://bitcoin.org/
4. See also Waltinger and Breuing (2012) in this volume.
5. http://asv.informatik.uni-leipzig.de/projects/6
6. In social networks, this consideration can also be applied: if a person knows two other persons within a certain community (cluster), the probability is high that these two persons also know each other.
7. Also called "six degrees of separation" by John Guare in 1990.

8. In order to get rid of the mystery in this "experiment", we give a simple gedanken-experiment: Imagine that every person in the world knows 100 people. Let us call these persons *friends*. Then one can potentially reach 10,000 friends through his or her friends, because every one of the 1st-order friends has 100 friends on her part. In a distance of six hops potentially $100 \cdot 100 \cdot 100 \cdot 100 \cdot 100 \cdot 100 = 100^6 =$ 1,000,000,000,000 people can be reached, more humans than living on Earth today. Of course, this is just an upper bound. Because of clustering in social networks, friends tend to know each other so that social networks are not trees but tend to be circular.
9. A regular graph is a graph where each vertex has the same number of neighbors.
10. A random graph is a graph that is generated by some random process. A good overview on random graphs is given by Bollobás (2001).
11. The idea to organize document collections into similarity clusters in order to improve IR results is not new: it was adopted by many other scientists. See, for example, Xu and Croft (1999) and Herschel (2007).
12. In graph theory the diameter of a graph is the maximum length of all its shortest paths, that is, the longest distance of all shortest distances between each pair of nodes.
13. http://ir.dcs.gla.ac.uk/test_collections/wt10g.html
14. http://ir.ohsu.edu/ohsumed/ohsumed.html
15. Cormack et al. (1999) explain and discuss the precision *P@k*.
16. All peers use the BM25 retrieval function to rank documents locally on each peer. CORI and *by size* are used just for peer selection.

References

Barkai, David

2002 *Peer-to-Peer Computing. Technologies for Sharing and Collaborating on the Net*. Intel Press.

Bollobás, Béla

2001 *Random Graphs*. Cambridge University Press, 2 edition.

Brin, Sergey and Lawrence Page

1998 The anatomy of a large-scale hypertextual web search engine. In *Computer Networks and ISDN Systems*, pp. 107–117. Elsevier Science Publishers.

Callan, James P., Zhihong Lu and W. Bruce Croft

1995 Searching distributed collections with inference networks. In *Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '95, pp. 21–28. ACM, New York, NY, USA.

Cooper, Brian F.

2004 A content model for evaluating peer-to-peer searching techniques. In Hans-Arno Jacobsen, (ed.), *Middleware 2004*, volume 3231 of *Lecture Notes in Computer Science*, pp. 18–37. Springer Berlin / Heidelberg.

2005 An optimal overlay topology for routing peer-to-peer searches. In *Proceedings of the ACM/IFIP/USENIX 2005 International Conference on Middleware*, Middleware '05, pp. 82–101. Springer, New York, NY, USA.

Cormack, Gordon V., Ondrej Lhotak and Christopher R. Palmer
1999 Estimating precision by random sampling (poster abstract). In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR ’99, pp. 273–274. ACM, New York, NY, USA.
Crespo, Arturo and Hector Garcia-Molina
2005 *Semantic Overlay Networks for P2P Systems*.
Decker, S., M. Schlosser, M. Sintek, and W. Nejdl
2002 HyperCuP – hypercubes, ontologies and efficient search on p2p networks. In *Proceedings of the International Workshop on Agents and Peer-to-Peer Computing*.
Diewald, Nils
2012 Decentralized social network sites. In Alexander Mehler, Laurent Romary, and Dafydd Gibbon, (eds.), *Handbook of Technical Communication.* De Gruyter, Berlin/Boston.
Garcia-Molina, Hector and Arturo Crespo
2003 Semantic overlay networks for p2p systems. Technical Report 2003–75, Stanford InfoLab.
Granka, Laura A., Thorsten Joachims and Geri Gay
2004 Eye-tracking analysis of user behavior in WWW search. In *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR ’04, pp. 478–479. ACM, New York, NY, USA.
Guan, Zhiwei and Edward Cutrell
2007 An eye tracking study of the effect of target rank on web search. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI ’07, pp. 417–420. ACM, New York, NY, USA.
Heinrich, Gregor, Hans Friedrich Witschel and Sven Teresniak
2006 Entwicklung von Testkollektionen für P2P Information Retrieval. In C. Hochberger, (ed.), *Informatik 2006. Informatik für Menschen. Bd. 1: Beiträge der 36. Jahrestagung der Gesellschaft für Informatik e.V. (GI)*, pp. 20–27. Gesellschaft für Informatik (GI).
Herschel, Sven
2007 Similarity-based document distribution for efficient distributed information retrieval. In *Proceedings of the 8th international conference on Web information systems engineering*, WISE’07, pp. 99–110. Springer, Berlin, Heidelberg.
Holz, Florian, Hans Friedrich Witschel, Gregor Heinrich, Gerhard Heyer and Sven Teresniak
2007 An evaluation framework for semantic search in p2p networks. In *Proceedings of the I2CS 2007.*
Joseph, Sam
2002a *NeuroGrid – Freenet Simulation Results*. Www.neurogrid.net/ng-simulation.html.
2002b NeuroGrid: Semantically routing queries in peer-to-peer networks. In Enrico Gregori, Ludmila Cherkasova, Gianpaolo Cugola, Fabio Panzieri, and Gian Picco, (eds.), *Web Engineering and Peer-to-Peer Computing*, volume

2376 of *Lecture Notes in Computer Science*, pp. 202–214. Springer Berlin / Heidelberg.

Kleinberg, Jon
1999a Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46(5): 604–632.
1999b Hubs, authorities, and communities. *ACM Computing Surveys*, 31(4es): 5+.
2000 The small-world phenomenon: An algorithmic perspective. In *Proceedings of the 32nd ACM Symposium on Theory of Computing*, pp. 163–170.

Lu, Jie and Jamie Callan
2003 Content-based retrieval in hybrid peer-to-peer networks. In *Proceedings of the twelfth international conference on Information and knowledge management*, CIKM '03, pp. 199–206. ACM, New York, NY, USA.

Milgram, Stanley
1967 The small world problem. *Psychology Today*, 2: 60–67.

Nejdl, Wolfgang, Boris Wolf, Steffen Staab and Julien Tane
2002 Edutella: Searching and annotating resources within an rdf-based p2p network. In Martin Frank, Natasha F. Noy, and Steffen Staab, (eds.), *Proceedings of the Semantic Web Workshop*, volume 55 of *CEUR Workshop Proceedings*. CEUR-WS.org.

Newman, Mark E. J.
2003 The structure and function of complex networks. *SIAM REVIEW*, 45: 167–256.

Ratnasamy, Sylvia, Paul Francis, Mark Handley, Richard Karp and Scott Shenker
2001 A scalable content-addressable network. In *Proceedings of the 2001 conference on Applications, technologies, architectures, and protocols for computer communications*, SIGCOMM '01, pp. 161–172. ACM, New York, NY, USA.

Ritter, Jordan
2002 *Why Gnutella can't scale. No, really.* http://www.darkridge.com/jpr5/doc/gnutella.html.

Robertson, Stephen and Hugo Zaragoza
2009 The probabilistic relevance framework: BM25 and beyond. *Foundation and Trends in Information Retrieval*, 3(4): 333–389.

Robertson, Stephen E., Steve Walker, Micheline Hancock-Beaulieu, Aarron Gull and Marianna Lau
1992 Okapi at TREC. In *Text Retrieval Conference*, pp. 21–30.

Salton, G., A. Wong and C. S. Yang
1975 A vector space model for automatic indexing. *Communications of the ACM*, 18: 613–620.

Schlosser, Mario, Michael Sintek, Stefan Decker and Wolfgang Nejdl
2003 HyperCuP: hypercubes, ontologies, and efficient search on peer-to-peer networks. In *Proceedings of the 1st international conference on Agents and peer-to-peer computing*, AP2PC'02, pp. 112–124. Springer, Berlin, Heidelberg.

Singla, Anurag and Christopher Rohrs
2002 *Ultrapeers: Another Step Towards Gnutella Scalability.* Lime Wire LLC, Working Draft, www.limewire.com/developrer/Ultrapeers.html.

Staab, Steffen and Heiner Stuckenschmidt
2009 *Semantic Web and Peer-to-Peer: Decentralized Management and Exchange of Knowledge and Information*. Springer, Berlin.
Steinmetz, Ralf and Klaus Wehrle (eds.)
2005 *Peer-to-Peer Systems and Applications*. Springer, Berlin.
Stoica, I., R. Morris, D. Karger, F. Kaashoek and H. Balakrishnan
2001 Chord: a scalable peer-to-peer lookup service for internet applications. In *Proceedings of the 2001 ACM SIGCOMM Conference*, pp. 149–160.
Stührenberg, Maik
2012 Foundations of markup languages. In Alexander Mehler, Laurent Romary, and Dafydd Gibbon, (eds.), *Handbook of Technical Communication.* De Gruyter, Berlin/Boston.
Travers, Jeffrey and Stanley Milgram
1969 An experimental study of the small world problem. *Sociometry*, 32(4): 425–443.
Waltinger, Ulli and Alexa Breuing
2012 Internet-based communication. In Alexander Mehler, Laurent Romary, and Dafydd Gibbon, (eds.), *Handbook of Technical Communication*. De Gruyter, Berlin/Boston.
Watts, Duncan J. and Steven H. Strogatz
1998 Collective dynamics of ‘small-world’ networks. *Nature*, 393: 440–442.
Witschel, Hans Friedrich
2005 Content-oriented topology restructuring for search in p2p networks. Technical report, University of Leipzig.
2008 *Global and Local Resources for Peer-to-Peer Text Retrieval*. Ph.D. thesis, University of Leipzig.
Witschel, Hans Friedrich, Florian Holz, Gregor Heinrich and Sven Teresniak
2008 An evaluation measure for distributed information retrieval systems. In *Proceedings of the IR research, 30th European conference on Advances in information retrieval*, ECIR’08, pp. 607–611. Springer, Berlin, Heidelberg.
Wong, S. K. M., Wojciech Ziarko and Patrick C. N. Wong
1985 Generalized vector spaces model in information retrieval. In *Proceedings of the 8th annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR ’85, pp. 18–25. ACM, New York, NY, USA.
Xu, Jinxi and W. Bruce Croft
1999 Cluster-based language models for distributed retrieval. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR ’99, pp. 254–261. ACM, New York, NY, USA.
Zhao, B. Y., Ling Huang, J. Stribling, S. C. Rhea, A. D. Joseph and J. D. Kubiatowicz
2004 Tapestry: a resilient global-scale overlay for service deployment. *IEEE Journal on Selected Areas in Communications*, 22(1): 41–53.

Biographical notes

Elisabeth Ahlsén has been professor of neurolinguistics at the University of Gothenburg since 1997, currently in the SCCIIL Interdisciplinary Center and the Division of Communication and Cognition of the Department of Applied IT. Her main research areas are neurolinguistics, pragmatics, multimodal and embodied communication, spoken face-to-face interaction, corpus linguistics, communication disorders, information and communication technology and social signal processing.

Jens Allwood is professor of Linguistics at the University of Gothenburg. He is also director of the interdisciplinary center SCCIIL and directs graduate programs in cognitive science and communication. His research primarily includes work in semantics and pragmatics. He has investigated face-to-face interaction from several perspectives, e.g. corpus linguistics, computer modeling of dialog as well as multimodal and intercultural communication. Presently he heads projects concerned with multimodal communication, cultural variation in communication and the influence of social activity on spoken language.

Bettina Berendt is a professor in the Artificial Intelligence and Declarative Languages Group at the Department of Computer Science of KU Leuven, Belgium. She obtained her PhD in Computer Science/Cognitive Science from the University of Hamburg, Germany, and her Habilitation postdoctoral degree in Information Systems from Humboldt University Berlin, Germany. Her research interests include Web and text mining, semantic technologies and information visualization and their applications, especially for information literacy and privacy. More information under http://people.cs.kuleuven.be/~bettina.berendt

Alexa Breuing is a Ph.D. student in the Artificial Intelligence Group at Bielefeld University and a research assistant at the Center of Excellence *Cognitive Interaction Technology* (CITEC) in the KnowCIT project. Her main research interest is how to enable human-agent interaction and dialog topic detection on the basis of knowledge drawn from the online encyclopedia Wikipedia.

Nils Diewald is a research assistant at the faculty of Linguistics and Literary Studies at Bielefeld University (Germany). He is currently working on his doctoral thesis on communications in the social web. His main research interests are technologies of the social web and applications of finite state technology.

Dafydd Gibbon is emeritus professor of English and General Linguistics at Bielefeld University, with research interests in relations between general linguistics, computational linguistics, language documentation and speech technology, particularly in the fields of lexicography, prosody and multimodality. He has led numerous research projects in these and neighboring areas, with special emphasis on African languages. In the technical communication field, Dafydd Gibbon coordinated the lexicographic group and an experimental speech recognition group in the Verbmobil speech-to-speech translation consortium. He has edited two handbooks and been involved in several research projects concerned with language and resources, data and tools, in particular the European consortia Speech Assessment Methodology (SAM) and Expert Advisory Groups on Language Engineering Systems (EAGLES). Since 2005 he has been Convenor of the International Advisory Group on Speech Databases and Assessment Techniques (COCOSDA).

Anthony Hartley is Emeritus Professor of Translation Studies at the University of Leeds. His interest in terminology stems from a long record of research in the deployment, evaluation and tuning of MT systems, in controlled authoring and in natural language generation. Much of this was conducted in UK and European projects with academic and industrial partners. He is currently based at Tokyo University of Foreign Studies, lecturing on translation technologies and researching into collaborative translation platforms.

Gerhard Heyer holds the chair on Natural Language Processing at the computer science department of the University of Leipzig. His field of research is focused on automatic semantic processing of natural language text with applications in the area of information retrieval and search as well as knowledge.

Florian Holz studied Linguistics and Computer Science at the University of Potsdam and the University of Leipzig. He is currently working as a researcher and lecturer at the University of Leipzig. He is especially interested in text and data mining, statistical inference and time series analysis. His main research foci are knowledge-free and unsupervised algorithms in a decentralized and online manner. He contributed to many projects on the empirical analysis of large corpora and consecutive applications in the field of communications and the humanities. His PhD thesis deals with large scale parallel time series analysis and novelty detection based on diachronic corpora.

Stefan Kopp is head of the research group “Sociable Agents” at the Center of Excellence “Cognitive Interaction Technology” (CITEC), adjunct professor of computer science at the faculty of technology at Bielefeld University and deputy speaker of the SFB 673 “Alignment in Communication”. His research interests include the empirical study and computational modeling of verbal and nonverbal communicative behavior. Combining cognitive modeling principles with probabilistic machine learning techniques, he aims to build artificial conversation partners (with virtual humans or robots) that can flexibly produce and robustly understand multimodal socio-communicative behavior. His current research focuses on natural multimodality with speech and gesture, the process of jointly and incrementally constructing a dialogue including interpersonal coordination and adaptation, and the grounding of these abilities in embodied cognitive architectures combining sensorimotor and social cognitive processing. Stefan is heading a number of projects on the analysis of human behavior, A.I. modeling, and empirical evaluation of human-agent interaction. He is president-elect of the German Cognitive Science Society.

Petra Kubina is a student of text technology, linguistics and computer science at Bielefeld University. One of her main areas of interest is barrier-free communication. She applies her expertise in this field in practical work on the design of websites. Petra Kubina contributed, for instance, to a more barrier-free relaunch of the website of her faculty. Personal acquaintance and communication with people of the target group of assistive and augmentative interaction devices fosters the user-centered perspective of her work. Petra Kubina currently prepares her Master's thesis on barrier-free research and teaching.

Peter Bernard Ladkin is professor of Computer Networks and Distributed Systems at the University of Bielefeld and specializes nowadays in complex-system safety. He invented Why-Because Analysis for determining the causes of accidents and incidents, and Ontological Hazard Analysis for determining hazards arising during deployment of safety-critical systems. He serves on various committees of the German standardization agency for electrical and electronic engineering and information technology, DKE, that deal with issues of safety. He is a director of the tech-transfer company Causalis Limited. Peter holds degrees in Mathematics, Philosophy and Logic from the University of Oxford and the University of California, Berkeley.

Codrina Lauth is affiliate researcher in the Knowledge Discovery (KD) Dept. at Fraunhofer Institute for Intelligent Analysis and Information Systems (IAIS), in Sankt Augustin, Germany. She has obtained her Magister Artium (M.A.) in contrastive computational linguistics from Technical University Braunschweig and Johannes Gutenberg University Mainz, Germany. Between 2002 and 2008 she was in charge of the project coordination and administration of the largest European KD Community - KDNet (NOE, www.kdnet.org, 5th FP) and KDubiq (Knowledge Discovery in Ubiquitous Environments, CA-6FP). She is one of the initiators and public relations chairs of the US-founded KD2U Society for Knowledge Discovery in Distributed and Ubiquitous Systems (www.kd2u.org).

She has been working as external lecturer and senior consulter for IT-Knowledge Management, Ubiquitous Text Mining and Analytics (see UTMA LinkedIn Group), machine translations and ISO/EU/DIN-normed terminology management and localization. She is the founder and CEO of LAUTH transmedia GmbH, an IT company affiliate to the C.R.E.A.M. Europe Private Public Partnership Alliance with the EU-Investment Bank (EIB), responsible for developing large international IT-infrastructure projects funded jointly by EIB, EU-regional and EU-research funding, as well as private investors.

Christopher (Cal) Lee is associate professor at the School of Information and Library Science at the University of North Carolina, Chapel Hill. He teaches graduate and continuing education courses in archival administration, records management, digital curation, and information technology for managing digital collections. His research focuses on curation of digital collections and stewardship of personal digital archives. Cal is PI for the BitCurator project and editor of *I, Digital: Personal Collections in the Digital Era.*

Henning Lobin is professor of applied and computational linguistics and head of the Centre for Media and Interactivity at Justus Liebig University in Giessen, Germany. His research interests cover the field of text technology, multimodal discourse analysis, and the cultural implications of digitalized reading and writing. He is currently heading, among others, a research project on the integration of text annotation, content analysis, and geographical data processing to provide new ways of accessing early Holocaust literature.

Arle Lommel is a senior consultant at the Berlin Language Technology Lab of the German Research Center for Artificial Intelligence (DFKI) and Standards Coordinator for the Globalization and Localization Association (GALA), which represents some 300 language services and technology companies. Formerly director of standards and publications at the Localization Industry Standards Association (LISA), he has more than 15 years of experience in working with language technology standards. He holds an M.A. and PhD in ethnology from Indiana University Bloomington and a BA in linguistics from Brigham Young University in Provo and is also a freelance Hungarian-to-English translator specializing in the domain of music.

Andy Lücking is research assistant at the Text Technology Lab at the Department of Computer Science and Mathematics at the Goethe University Frankfurt am Main, Germany. He received his PhD in linguistics for a unification-based account of iconic gestures and their grammatical integration with verbal meaning. His research interests include the interplay of verbal and non-verbal signs within philosophical and descriptive frameworks of communication. In collaborative work, Andy Lücking currently also approaches the application of multimodal interfaces within educational usage scenarios.

Jean-Claude Martin received the PhD degree from Télécom Paris in 1995. He is a professor of computer science at the University of Paris South XI. He is the head of the group "Cognition Perception and Use" at the Computer Sciences Laboratory for Mechanics and Engineering Sciences (LIMSI-CNRS). He is the Editor-in-Chief of the Springer *Journal on Multimodal User Interfaces* (JMUI). He is studying multimodal expression and perception of affective behaviors in humans and virtual agents.

Aurélien Max is assistant professor in Computer Science at Paris Sud University and LIMSI-CNRS in Orsay, France. He holds a PhD in Computer Science from Grenoble University, and a MPhil in Computer Speech and Language Processing from the University of Cambridge. His work concentrates on issues relating to using context and linguistic information for Natural Language Processing applications such as Machine Translation and Automatic Paraphrasing.

Alexander Mehler is professor of text technology at Goethe University Frankfurt am Main, Germany, were he heads the Text technology Lab as part of the Department of Computer Science and Mathematics. Alexander Mehler is member of the executive committee of the LOEWE Priority Program "Digital Humanities". His research interests include the empirical analysis and simulative synthesis of discourse units in spoken and written communication. He aims at a quantitative theory of networking in linguistic systems to enable multi-agent simulations of their life cycle. Alexander Mehler integrates models of semantic spaces with simulation models of language evolution and topological models of network theory to map the complexity of linguistic information systems. Currently, he is heading several research projects on the analysis of linguistic networks and on historical semantics.

Peter Menke studied computer science, linguistics and literary studies at Bielefeld University. He received his Master's degree in linguistics in 2009. His research focuses on representations of communicative modalities and their structures. Peter Menke aims at an exhaustive ontology of modalities and their related coding systems and data units used in multimodal communication. Currently, he is developing a suite of applications for the administration and processing of complex multimodal dialog corpora. In his doctoral thesis he formalizes its basis, a data model and library for advanced operations on linguistic and nonlinguistic dialog data.

Gerhard Paaß is senior researcher in the Fraunhofer Institute for Intelligent Analysis and Information System IAIS in Sankt Augustin, Germany. In the department of Knowledge Discovery he founded the Text Mining Group and is its leading researcher. Recently he led the IAIS project on semantic extraction in the framework of the German Federal THESEUS program. His research interests cover the area of statistical methods for machine learning, especially applied to structured data and text. Instead of manually prescribing the features for information extraction tasks he aims at automatically extracting relevant features by feature selection methods as well as by generating latent variables to predict the desired outcomes in an optimal way.

Thies Pfeiffer is postdoc at the Artificial Intelligence Group of Prof. Ipke Wachsmuth at the Faculty of Technology of Bielefeld University, Germany. He is interested in multimodal human-computer interaction with a strong focus on communication via gaze and gestures in 3D environments. Since 1999, Thies Pfeiffer is working with the embodied conversational agent Max. His current research focuses on mobile pervasive interaction using speech, gaze and gesture.

Bastian Pfleging is a research assistant at the Human-Computer Interaction Group of the Institute for Visualization and Interactive Systems (VIS) at the University of Stuttgart, Germany. His general research interests are multimodal and natural user interfaces. In particular, he is interested in human-computer interaction in the automotive context. He received his MSc in Computer Science (Diploma) from TU Dortmund, Germany. From 2009 to 2011 he was working as a research assistant at the Pervasive Computing and User Interface Engineering Group (http://www.pervasive.wiwi.uni-due.de/en/) at the University of Duisburg-Essen, Germany. From 2010 to 2011 he was visiting the BMW Technology Office in Palo Alto / Mountain View, CA, USA (http://www.bmw group.com). He is a member of ACM, GI and VDI.

Sebastian Rahtz is head of the Information and Support Group at Oxford University Computing Services, where he oversees the teams responsible for web, mobile apps, help desk, IT staff liaison, and development projects. He has been closely associated with the Text Encoding Initiative for the last decade as a member of its Technical Council, architect of its revised metaschema system, and author of a library of XSL transforms for TEI documents (including the Guidelines documentation and its schemas). Since 2008 he has been part of the team developing CLAROS ("the world of ancient art on the semantic web") at Oxford, for which he leads the Metamorphoses sub-project to manage its place and name linking. He has a degree in Classics and Modern Greek from Oxford, and an MA Archaeology from London.

Laurent Romary is Directeur de Recherche INRIA, France, and guest scientist at Humboldt University in Berlin, Germany. He carries out research on the modeling of semi-structured documents, with a specific emphasis on texts and linguistic resources. He is the chairman of ISO committee TC 37/SC 4 on Language Resource Management, and has been active as member (2001-2007) then chair (2008-2011) of the TEI (Text Encoding Initiative) council. He currently contributes to the establishment and coordination of the European DARIAH infrastructure (www.dariah.eu).

Dietmar Rösler is professor of German as a Foreign Language at Giessen University. He worked as a member of staff in the German departments of University College Dublin, Freie Universität Berlin and King's College, University of London. His main research areas are intercultural learning, learning material design, German grammar and the role of the digital media in foreign language learning. He has published extensively on these and related topics.

Felix Sasaki studied Japanese and Linguistics from 1993 until 1999 in Berlin, Nagoya (Japan) and Tokyo. Since 1999 he worked in the Department of Computational Linguistics and Text-technology, at the University of Bielefeld (Germany), where he finished his PhD in 2004. Felix Sasaki joined the *World Wide Web Consortium* (W3C) in 2005 to work in the Internationalization Activity until 2009. In 2009 he became a lecturer at the University of Applied Sciences Potsdam; in 2010 he joined the Language Technology lab of DFKI. In 2012 he rejoined the W3C team as a fellow on behalf of DFKI. His main field of interest is the application of Web technologies for the representation and processing of multilingual information.

Albrecht Schmidt is a professor for Human Computer Interaction at the University of Stuttgart. He was previously at the University of Duisburg-Essen, the University of Bonn and Fraunhofer IAIS. He studied computer science in Ulm and Manchester and received in 2003 a PhD from the Lancaster University. His research interest is in human computer interaction beyond the desktop, including user interfaces for mobile devices and cars. Albrecht published well over 100 refereed archival publications and his work is widely cited. He is co-founder of the ACM conference on Tangible and Embedded Interaction (TEI). He is an area editor of the IEEE Pervasive Computing Magazine and edits a column on invisible Computing in the IEEE Computer Magazine.

Tanja Schultz received her Ph.D. and Masters in Computer Science from University Karlsruhe, Germany in 2000 and 1995 respectively and got a German Staatsexamen in Mathematics, Sports, and Educational Science from the University of Heidelberg, in 1990. She joined Carnegie Mellon University in 2000 and became a Research Scientist at the Language Technologies Institute. Since 2007 she has also been a Full Professor at the Karlsruhe Institute of Technology (KIT) in Germany. She is the director of the Cognitive Systems Lab, where her research activities focus on human-machine interfaces with a particular area of expertise in rapid adaptation of speech processing systems to new domains and languages. Her recent research focuses on human-centered technologies and intuitive human-machine interfaces based on biosignals, by capturing, processing, and interpreting signals such as muscle and brain activities. Tanja Schultz is the author of more than 200 articles published in books, journals, and proceedings. Currently, she is a member of the IEEE Computer Society, the International Speech Communication Association ISCA, the European Language Resource Association, the Society of Computer Science (GI), and serves as elected ISCA Board member, Associate Editor of ACM Transactions on Asian Language Processing, on several program committees, and review panels.

Serge Sharoff is Senior Lecturer at the Centre for Translation Studies, University of Leeds. He holds a PhD in Computational Linguistics and worked as a freelance translator during his studies. His research interests combine three domains: linguistics, computer science and communication studies. In particular, he investigates the automatic acquisition of representative corpora from the Web and their classification in terms of domains and genres. He also researches methods of finding translation equivalents for both terms and general lexical items from such corpora.

Maik Stührenberg received his Ph.D. in Computational Linguistics and Text Technology from Bielefeld University in 2012. After graduating in 2001 he worked four years as research assistant at Giessen University in different text-technological projects. From 2005 to 2008 he was a member of the Sekimo project of the Research Group 437 Text-technological Modeling of Information funded by the German Research Foundation. After being employed at the Institute for the German Language (IDS) as a member of the CLARIN-D project group, he is currently employed as research assistant at Bielefeld University.

Sven Teresniak studied Computer Science and Linguistics at the University of Leipzig. His scientific interests are in Natural Language Processing, Text Mining, and content-based peer-to-peer networks. For his PhD thesis he's currently working on unsupervised trend mining approaches. As a researcher for the ExB Group his work encompasses the development of advanced language technologies with focus while considering the restrictions of mobile devices.

Thorsten Trippel is a researcher in the area of computational linguistics. As a graduate of Bielefeld University, Germany, he worked at the University of Tübingen since 2010, where he addresses issues related to the archiving and reusing of language resources. Previous work involves lexical resources for technical domains and graph structures in lexical resources. He is an expert in standardization committees for language resources and systems to manage terminology, knowledge and content. For his Ph.D. he worked on data structures of lexical resources.

Ipke Wachsmuth is a professor of Artificial Intelligence at Bielefeld University, a principal investigator in the Center of Excellence "Cognitive Interaction Technology" (CITEC) and an adjoint member of Bielefeld University's Research Institute for Cognition and Robotics (CoR-Lab). He has also been director of the Center for Interdisciplinary Research (ZiF) and coordinator of the Collaborative Research Center "Alignment in Communication". His research focuses on human-machine interaction and virtual reality, including gestural and multimodal interaction, embodied artificial agents and the modeling of communication with robots and virtual humans. Practical aspects of his research aim at advanced human-machine interfaces which realize embodied communication in fully immersive virtual reality.

Ulli Waltinger is a research engineer at Siemens Corporate Technology Munich. He was a post-doctoral researcher at the Center of Excellence Cognitive Interaction Technology (CITEC) at Bielefeld University where he also completed his Ph.D. studies in the field of social semantics in information retrieval.

Index

B

D

E

F

G

H

I

L

M

N

O

P

Q

R

S

T

U

V

W

X

Y