

DE GRUYTER
MOUTON

MULTIMODALITY

DISCIPLINARY THOUGHTS AND THE CHALLENGE
OF DIVERSITY

*Edited by Janina Wildfeuer, Jana Pflaeging,
John Bateman, Ognyan Seizov, Chiao-I Tseng*

Copyright 2019. De Gruyter. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

EBSCO Publishing : eBook Collection (EBSCOhost) - printed on 2/9/2023 10:35 PM

via
AN: 1303 ; Janina Wildfeuer, Jana Pflaeging, John Bateman, Ognyan Seizov,
Chiao-I Tseng.; Multimodality : Disciplinary Thoughts and the Challenge of
Diversity
Accession Number: 335141

DE
G

Multimodality

Multimodality

Disciplinary Thoughts and the Challenge of Diversity

Edited by
Janina Wildfeuer, Jana Pflaeging, John Bateman,
Ognyan Seizov and Chiao-I Tseng

DE GRUYTER
MOUTON

ISBN 978-3-11-060798-7
e-ISBN (PDF) 978-3-11-060869-4
e-ISBN (EPUB) 978-3-11-060805-2

Library of Congress Control Number: 2018946711

Bibliographic information published by the Deutsche Nationalbibliothek

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

© 2019 Walter de Gruyter GmbH, Berlin/Boston
Printing and binding: CPI books GmbH, Leck

www.degruyter.com

Preface and Acknowledgements

The editors of this volume would like to thank, first and foremost, the authors for their insightful contributions and hard work during the reviewing process. We are also grateful for the valuable suggestions made by an anonymous reviewer. Without their dedication and responsiveness, this volume would not have been possible.

We also want to extend our thanks to our authors as well as to the other colleagues for taking part in #BreMM17: *Third Bremen Conference on Multimodality: Towards a New Discipline* on 20–22 September 2017 in Bremen, Germany. Their creativity and active participation in the interdisciplinary dialogue throughout the event sparked our enthusiasm to document some of the conference’s major themes and directions in this volume.

The *Bremen Conferences on Multimodality* have become a firm pillar in the multimodality context. Starting in 2014, *BreMM14* was dedicated to building bridges between various multimodality-ready disciplines, *BreMM15* continued with theoretical and methodological explorations in 2015 and built first foundations for the discipline-building endeavor at *BreMM17* and in this book. *BreMM19*, in September 2019, focuses on empirical inroads in multimodality research and with this further establishes a more robust disciplinary core. We hope to be able to keep this high level of continuity and fruitful exchange in one form or another and push multimodality’s development further.

The BreMM17 conference was a success also thanks to the financial support of the German Research Foundation (Deutsche Forschungsgemeinschaft – DFG) as well as the logistical support of the University of Bremen, and we gratefully acknowledge their contributions to this first step towards the volume’s development.

Janina Wildfeuer, Jana Pflaeging, John A. Bateman,
Ognyan Seizov, and Chiao-I Tseng
Groningen, Salzburg, Bremen, and Berlin, September 2019

Contents

Preface and Acknowledgements — V

Part I: Introduction

Janina Wildfeuer, Jana Pflaeging, John A. Bateman, Ognyan Seizov, and Chiao-I Tseng

**Multimodality: Disciplinary Thoughts and the Challenge of Diversity –
Introduction — 3**

Part II: Disciplinary Thoughts

Hartmut Stöckl

**Linguistic Multimodality – Multimodal Linguistics:
A State-of-the-Art Sketch — 41**

Martin Thomas

**Making a Virtue of Material Values:
Tactical and Strategic Benefits for Scaling Multimodal Analysis — 69**

Kay L. O'Halloran, Sabine Tan, Peter Wignell, Rui Wang, Kevin Chai, and Rebecca Lange

**Towards a Discipline of Multimodality:
Parallels to Mathematics and Linguistics and New Ways Forward — 93**

Part III: Diversity

Axel Schmidt and Konstanze Marx

**Multimodality as Challenge:
YouTube Data in Linguistic Corpora — 115**

Lauren O'Hagan

**Class, Culture, and Conflict in the Edwardian Book Inscription:
A Multimodal Ethnohistorical Approach — 145**

John Harnett

Cognitive Pathfinders:

Highlighting Cross-Modal Interaction and the Orchestration of Memory in Comics — 171

Christopher Taylor

Audio Description:

A Multimodal Practice in Expansion — 195

Edward Larkey

Narrative as a Mode of Communication:

Comparing TV Format Adaptations with Multimodal and Narratological Approaches — 219

Wendy Nielsen, Pauline Jones, Helen Georgiou, Annette Turney, and Mary Macken-Horarik

Learning Science through Generating Multimodal Digital Explanations:

Contributions to Multimodality in Educational Practice — 247

Dušan Stamenković and Milan Jačević

Video Games and Multimodality:

Exploring Interfaces and Analyzing Video Screens Using the *GeM* Model — 277

Part IV: More Disciplinary Thoughts

John A. Bateman

Afterword: Legitimizing Multimodality — 297

List of Contributors — 323

Index — 329



Part I: Introduction

Janina Wildfeuer, Jana Pflaeging, John A. Bateman, Ognyan Seizov, and Chiao-I Tseng

Multimodality: Disciplinary Thoughts and the Challenge of Diversity – Introduction

Abstract: In this introduction, we discuss the idea of establishing a discipline of multimodality, considering both how this might be defined and potential benefits and challenges of attaining such an independent status. This builds on previous rounds of discussion within the Bremen Conferences on Multimodality (BreMM) series concerning this issue, where diverse approaches to the study of multimodality have come together to create a broad-based set of agenda items where a more systematic engagement with the phenomena of multimodality is key.

Keywords: multimodality, discipline, interdisciplinarity, transdisciplinarity, multidisciplinarity, semiotics

1 Conceptualizing Multimodality’s Disciplinary Status and Challenges

This book presents a collection of position articles and reports of empirical and theoretical results in multimodality drawn from the most recent of our series of international conferences on multimodality. The broad range of topics, approaches, and research objects covered by the contributions in this volume attests clearly to the growing *diversity* of what is typically subsumed under the label of *multimodality*. Not only does our volume showcase this remarkable diversity, including many innovative transdisciplinary and cross-national research projects, it also seeks to draw attention to the inclusivity and flexibility of current approaches in our field.

Be it in different ways, both of these characteristics encourage a more explicit and more thorough discussion of multimodality’s current disciplinary status, and any future developments that result from it: On the one hand, the overlap in research interests and similarity of approaches seems to justify thinking about a commonly shared name, concepts, or institutionalization. On the other hand, and perhaps more importantly, the prevalent diversity of approaches and perspectives suggests considering the unifying potential of a disciplinary ‘home’, which promotes an ongoing exchange of ideas and mitigate less productive divergences and fragmentation.

<https://doi.org/10.1515/9783110608694-001>

As documented by this volume's contributions, scholarship within the 'broad context of multimodality' has recently been expanding into a wide range of areas in which an even broader variety of methods is adopted. This developmental trajectory indicates that multimodality may indeed be at risk of fragmentation if (meta-)disciplinary guidelines are not adopted that fully embrace the necessity of associative and hierarchical approaches, and channels to communicate. As a field naturally poised for further growth, multimodality needs to be well-prepared to tackle a host of diverse theoretical and empirical challenges, both now and in the future. We believe that achieving a more mature status of critical self-reflection that will ultimately be able to offer better mutual support across research initiatives and methods, as well as strengthening interaction across such initiatives, will be instrumental in moving forward successfully.

This introduction, in particular, is entirely devoted to the pressing discipline-related questions pointed out, placing a particular focus on the diversity and breadth of multimodal research and critically conceptualizing the question of whether multimodality could, and perhaps even should, be seen as a discipline in its own right. In addition, some of the discussions led in particular chapters of this book also consider, either theoretically or by drawing on the results of empirical research, whether a move towards establishing multimodality as a stand-alone discipline is desirable and, if so, what would be necessary to accomplish it.

Our disciplinary thoughts have been heavily influenced by contributions from the international community of multimodality researchers, particularly those scholars that shared their work at one of the previous Bremen Conferences on Multimodality. These voices strongly support our assumption that the "broad context of multimodality research is now on its way toward forming its own fully acknowledged discipline" (Wildfeuer & Seizov, 2017, 279).

As our conferences as well as the many handbooks and introductions to the field continuously show, activities grouped under the term 'multimodality' constitute one of the most influential contemporary approaches to the study of all kinds of communicative artifacts and performances. Multimodal perspectives on such research objects can now be found in almost every discipline interested in the analysis of communication. This popularity even transcends such broad orientations as the 'humanities' and includes scholars and practitioners from diverse fields wherever the goal is to scrutinize combinations of communicative forms and the meaning-making mechanisms deployed. As a consequence, these efforts readily employ tools and frameworks that have evolved, and which continue to evolve, drawn from diverse research areas as well. In this sense, multimodality can already be seen to be far more multidisciplinary than many other fields of research; van Leeuwen (2005, 1), for example, even sees multidisciplinary as an

“essential feature” at least of social semiotics, upon which many approaches to multimodality build.

This situation brings both benefits and challenges. Among the benefits are precisely that broad range of different voices and empirical applications that become included from various areas concerned with communication. Conversely, however, among the challenges comes the task of dealing productively with often quite disparate conceptual anchorings, frameworks, and questions. Indeed, one of the most frequent and reoccurring doubts raised concerning multimodal research is that of whether such diverse voices can communicate at all. As a consequence, attempts to procure funding still often founder on the shoals of methodological doubt among reviewers relatively unversed in the radical multidisciplinary that has long formed the everyday practice of many of those active in multimodal research.

Moreover, and as we have set out at length previously (see Wildfeuer, 2015; Seizov & Wildfeuer, 2017; Bateman et al., 2017), although the theories and methods of multimodality are indeed naturally heterogeneous, they have also nevertheless often remained confined to national and regional research communities, sometimes cross-cutting theoretical or philosophical lines of demarcation, sometimes aligning with them. This in turn continues to create problems of recognition both within multimodality and without. The current state of ‘multimodality’ is then one of a rather uncoordinated field that does not always grant its own wide-ranging scholarly community and the results they have achieved equal respect. Even appropriate mutual knowledge of related approaches is often wanting. It is particularly in response to issues of these kinds that we consider a more explicit orientation to the status of multimodality as a potential discipline not only timely, but also increasingly urgent.

The recent historical development of the diverse activities potentially contributing to multimodality has been complex. By the 1960s and 1970s, multimodality had already become an object of study in several disciplines and research directions, even if the term was not itself always explicitly used. Since then, and particularly after a first phase of development and theory formation in the 1990s and early 2000s, multimodal questions have increasingly taken on the role of an informing and supplementing research area incorporated by other disciplines or included in interdisciplinary projects, calling for practitioners and theorists alike to combine complementary approaches rather than seeking insufficiently motivated discipline-internal replacements or by beginning anew. Today, further developments in attitudes and awareness have begun to promote an openness that has increased both the motivations for, and the rewards of, broader inter- and transdisciplinary approaches (see Wildfeuer & Seizov, 2017). These not only focus on theory and method but also include a substantial growth in empirical applications and evaluations. Taken together, such developments have already

widened horizons and opened up fundamental academic debates leading to more robust definitions and conceptualizations of the field's main terms, and these, particularly, now invite explicit discipline-building explorations.

Although the current diversity and multidisciplinary of multimodality could reasonably be seen as a significant constraint on sustainable development and advancement of the field, these properties have also created, on the other hand, a multiperspectivity which holds considerable potential for shaping a disciplinary status of a rather particular kind. The contributions to this book mainly reflect this potential of the diversity of disciplines and interests they represent. Their collection here serves to contribute further way stations requiring consideration when envisioning roadmaps for multimodality as the field reaches towards disciplinary status. Multimodality has clearly not achieved this status yet, however, and there are still serious concerns with regard to other disciplines and developing trends that make it difficult to assess whether this will happen in the near future. Nevertheless, below we further outline some relevant developments of the field to date which would serve as prerequisites for such disciplinary status, and we critically discuss how multimodality may benefit from further efforts in this direction.

As preparations for this, we explicitly included in the call for papers for BreMM17 several questions central when considering moves towards a discipline of multimodality. In Section 2 following, therefore, we begin by presenting recent relevant developments drawing on the conference contributions and presenting an overview of the chapters appearing in the volume. These all more or less respond to the questions raised in the call for papers by providing theoretical and methodological discussion or practical reports exemplifying the current status of multimodality. In the subsequent sections, we turn to consider the potential of these contributions as participants in a broader discipline-building endeavor. In Section 3, we collect criteria from more general characterizations of the features of disciplines as well as presenting a more particular assessment of one research field—the Digital Humanities—which is often argued to have gained the status of a fully-developed discipline despite being constituted by a similarly heterogeneous range of approaches and questions. In Section 4, this collection of discipline-specific characteristics is taken as a backdrop against which multimodality's current disciplinary status can also be assessed. At the same time, this offers a frame of reference which makes apparent further prerequisites for the advances necessary for a discipline to emerge.

We then suggest an approach to a discipline of multimodality which, on the one hand, would provide a complex definition for the discipline, while, on the other hand, leaving appropriate room for extension. Finally, in Section 5, we round up the introductory discussion with some further reflections and an outlook on future directions relevant for establishing multimodality as a discipline in its own right.

2 Towards a New Discipline: Developments to Date and This Book's Contributions

In this section, we focus particularly on what steps have already been taken towards multimodality as a 'stand-alone' discipline, regardless of whether they were announced as such at the time or have only become identifiable in retrospect, and then situate the contributions to this book against this background.

2.1 Developments to Date

Efforts to make multimodality a more homogeneous field or provide common foundations for a more unified theory or methodology are not new and can be found in various discussions early on, albeit often not in a particularly differentiated fashion. In the late 2000s, when the first handbooks and comprehensive overviews of the field were published, multimodality was described as “a theory, a perspective, a methodological application or a field of enquiry” (Jewitt, 2009a, 127)—a conceptualization that is not without ambiguity and vagueness. However, with calls for “a general project of multimodality” (Constantinou, 2005, 604) and the implementation of more “systematic rigour” (Forceville, 2010, 2607), the “need for methodological development” (Björkqvall, 2012, 8) became increasingly recognized as a pressing issue.

Nevertheless, while experiencing rapid growth, as noted above, multimodality research remained extremely heterogeneous, exhibiting a broad diversity of theoretical and methodological developments. Indeed,

[t]his diversity is [...] one of the constituting features not only of the field and its contexts of investigation, but also of all the work combining under this keyword, which makes it a particularly challenging and exciting endeavour. (Wildfeuer, 2015, 21)

Moreover, there were (and still are) considerable differences between national and international approaches. Crossing disciplinary boundaries therefore poses a considerable challenge, as demonstrated with the example of the situation in Germany in comparison to the international context discussed in Wildfeuer (2015).

One of our main aims at that time, which we now see as a first initiative towards discipline-building, was to find relations between the various approaches engaging in multimodality research and to construct bridges between them as necessary, identifying similarities and differences within this inter- and multidisciplinary mix. Here it was necessary to emphasize that “multimodality is interdisciplinary in itself and there is absolutely no need for any discipline to claim the leading role for

itself” (Wildfeuer, 2015, 22); it could already be observed that many different disciplines involved in multimodal research were “not compet[ing] against each other but rather gain from their reciprocal influence and mutual exchange” (Wildfeuer, 2015, 22). The contacts established between theories, methods, and empirical applications made visible further potential for developments that increasingly “help strengthen multimodality as a still to be fully established paradigm” (Wildfeuer, 2015, 30).

Parallel to our own initial efforts in the early 2010s, several other colleagues have made similar attempts to revisit multimodality by adopting a broader perspective and framing it as a discipline. Not least, several multimodality-related handbooks and textbooks published in recent years have been crucial for reflecting and negotiating this diversity in multimodality research, albeit in rather different ways. Handbooks, on the one hand, usually highlight the multi- and transdisciplinary perspective of multimodality and conceive and conceptualize the notion as broadly as possible (see, e.g., Norris, 2015; Klug & Stöckl, 2016). This allowed major achievements in multimodal research to be showcased and the similarities (and differences) in the theories, methods, and empirical applications to be identified. However, none of the handbooks available to date has succeeded in critically addressing, let alone synthesizing, these perspectives in a manner that would elevate the discussion to the more abstract level essential for a discipline to develop. Textbooks, on the other hand, naturally aim at laying foundations for more systematic and problem-oriented studies of multimodal phenomena (see, e.g., Jewitt et al., 2016) and thereby, in principle, also pave the way towards more discipline-building elaborations (see, e.g., Bateman et al., 2017). Most textbooks to date have, however, only dealt with multimodality as a broad research field or practice and thus may not explicitly address questions of disciplinary status either. And, moving beyond textbooks and handbooks, most recent work in the field of multimodality has also not focused particularly on defining the criteria that would need to be applied to assess and establish multimodality’s status as a discipline. Works that do include discussions of multimodality’s possible disciplinary status, such as, for example, van Leeuwen (2005), Feng et al. (2013), Forceville (2013), and van Leeuwen (2014), only touch upon the topic briefly and remain rare exceptions.

But regardless of whether due to a perhaps naive optimism or reflecting an actual conviction in multimodality’s status as a discipline, the fact that scholars are increasingly adopting a broader perspective and draw on the discipline-label, reflects in our view a significant sea-change. Calls “to account for the disciplinary status of multimodality” (Wildfeuer & Seizov, 2017, 279) are growing in volume and taken together point to the field’s potential to continuously renew and rethink current ideas of multimodal and media analysis and, in so doing, to move towards

more concerted efforts that are indeed resonant with ideas of disciplinarity (cf. van Leeuwen, 2014, 252).

In Wildfeuer & Seizov (2017), we developed further considerations of this potential in an initial practice-based definition of multimodality as a

modus operandi [...], a habit of research [..., which] starts far-reaching discussions that cover description, terminology, methodology, as well as practical analysis in order to build the basis for a profound disciplinary status. (Wildfeuer & Seizov, 2017, 279–280)

This allowed us to engage in interdisciplinary exchange and to establish a discussion of multimodality's status as a distinct new discipline at the intersection of other disciplines—including, but not limited to, linguistics, communication and media studies, pedagogy, sociology, cultural studies, argumentation theory, musicology, information design, translation studies, experimental psychology and neuroscience, as well as artificial intelligence, human-computer interaction and interface design. But it remained evident that “a discipline of multimodality existing individually and independently alongside linguistics, communication studies, digital humanities, or other disciplines [...] institutionalized via professorial chairs and study courses with the respective denomination” (Wildfeuer & Seizov, 2017, 278) had not yet emerged, leaving the majority of multimodality researchers still more or less uncomfortably placed ‘between’ disciplines. And so, in order to come together in a concerted endeavor of engaging in active exchange and joint research, contributors from different disciplines are commonly faced with the need to transcend traditional borders often maintained by institutional structures at universities, research facilities, and funding agencies.

As a consequence, the Third Bremen Conference on Multimodality (BreMM17) held at the University of Bremen in September 2017 was dedicated entirely to this issue. We took previous efforts as a starting point to call for further and more detailed discussions of the topic of a discipline of multimodality. As set out in the Call for Papers, BreMM17 aimed ‘at repositioning the field as a well-grounded scientific discipline with significant implications for future research in all areas involved in the study of multimodal communication’. For this reason, we sought to start more far-reaching discussions that ‘cover description, terminology, and methodology, [bring] a multitude of approaches to multimodal analysis into the fold and [let] previously disparate directions in theory and practice converge’ (CFP, BreMM17).

The Call for Papers then asked particularly for contributions that would guide discipline-building explorations by addressing the following questions:

- What previously established disciplines should inform multimodality’s disciplinary delineation? What is the place of, for example, semiotics, systemic-functional linguistics, discourse analysis, interaction analysis, and other popular methods if multimodality were to be defined as a stand-alone discipline?
- Where can multimodality find its most inclusive and exhaustive theoretical foundation? Do we need ways of combining the pioneers’ work to produce a new theoretical basis for the discipline? Or do we start a new theory from scratch?
- What belongs in multimodality’s methodological toolbox? What existing empirical approaches define the field, how can we develop them further or combine them, and do we need new methods to capture multimodality’s breadth?
- What are multimodal media and how do their various semiotic affordances shape multimodality within and across media formats? Are all media inherently multimodal?

On this platform we envisaged that the papers to be presented at the conference would help characterize multimodality as a discipline. The resulting submissions and the selection from these presented in this collection certainly show that these aims retain their legitimacy, but also show that the situation continues to be far from straightforward. Although most of the contributions in the current volume take explicit positions on the question of whether and how to move towards multimodality as a discipline in its own right, signs of disciplinary boundaries are still endemic and further detailed and continued discussion is essential. A brief synopsis of the chapters of the book will clarify these concerns; we then return to these positions and their specific contributions to the multimodality question below.

2.2 Overview of Contributions

As indicated above, the contributions to the current volume address a diverse range of media and genres, each exhibiting broad constellations of semiotic modes at work. The media addressed include TV series, interfaces of computer games and videos of computer game play, advertisements, paintings and their audio descriptions in museums, digital presentations, live performance, graphic novels, and inscriptions of ownership within books. In each case, attention is drawn both to the complexity of intermodal relationships and to the consequences of those relations for the use and interpretation of the artifacts considered.

Following the title of this volume, we organize the contributions into four parts: As suggested above, this introductory chapter, which constitutes **Part I**, is devoted to the question of whether, or to what extent, multimodality can be considered a discipline in its own right. Adopting a decidedly broad meta-perspective, it seeks to set the stage for a systematic consideration of multimodality as a discipline. It is complemented by three further contributions, grouped into **Part II** of the book, that offer more specific perspectives on these issues. In particular, they clearly focus on essential discipline-building pillars such as a strong linguistic foundation, a more empirically- and material-oriented scale of multimodal analyses, as well as the orientation back towards semiotics and mathematics.

Hartmut Stöckl offers a view that brings out some of the ways in which multimodality might be considered a discipline, focusing quite specifically on ‘linguistic multimodality’—that is, occurrences of multimodal communication essentially involving language. Stöckl suggests that a set of core concepts has now emerged that can be seen to support a ‘more or less unified’ base theory, together with an increasing diversity of empirical studies and lively scientific interchange. He suggests that the fact that there are, nevertheless, still many ‘subdisciplines’, or distinct kinds of approaches to multimodality, as well as contrasting and not always compatible definitions of terms, may be more due to the youth of the discipline and that, as the field of activities involved matures, there may well be movement towards a more consolidated disciplinary orientation. As part of the discussion, Stöckl works through some of the terms more frequently used, including *communicative action*, *dialogicity*, *meaning multiplication*, *mode*, *affordance*, *semiotic resource*, and so on, placing them in connection with one another. He then proposes three further areas of disciplinary activity that he considers necessary for achieving a robust foundation both for theorizing and for conducting practical research: multimodal rhetoric, cognitive semiotics, and transtextuality.

Martin Thomas considers some of the issues raised in building corpora of multimodal artifacts and performances and suggests that a further re-orientation to materiality would help in this task. The author particularly emphasizes that multimodality, in order to grow, needs to engage with substantially larger data sets than has typically been the case hitherto. And, for that, it is necessary to find methods that ‘scale’. By drawing on properties of data that can be derived automatically, for example by the application of techniques from the field of computer vision, Thomas proposes that more abstract and general questions can be addressed building on a more robust basis ranging across genres. This is suggested to offer not only an advance of scale but also to avoid premature interpretation: an important consideration for placing multimodality on a more secure empirical footing.

Kay O'Halloran, Sabine Tan, Peter Wignell, Rui Wang, Kevin Chai, and Rebecca Lange put forward the case that multimodality, as a field of endeavor, is poised to take an analogous role as such strongly enabling disciplines as mathematics and the language sciences, and so, by analogy, could well be considered a discipline in its own right. The argument here focuses on the multiple and diverse applications of mathematics and the language sciences, which in many respects seem to have precisely the same kind of 'transdisciplinary' status that has been suggested for multimodality. They also present their view of how multimodality research, in the future, will need to draw far more on larger and broader sources of data, making computational techniques both of analysis and of visualization and presentation crucial. For these authors, such techniques are not an optional extra but relate directly to effective theory and practice.

Part III attests to the considerable diversity and breadth of multimodal research. These different research contexts, orientations, and approaches give rise to different lines of argumentation with regard to the discipline question. For some scholars, the sheer diversity of connections that are necessary with other established disciplines makes the disciplinary status of 'multimodality' questionable. For others, it is precisely that diversity that allows us to argue for an independent status, since across the board there do appear to be a range of concepts and, increasingly, analytic methods that reoccur in different contexts of use and application.

The first contribution in this part by **Axel Schmidt** and **Konstanze Marx** addresses concrete methodological and theoretical issues—in this case those raised by the challenges of establishing multimodal corpora as foundations for empirical work. On this basis, they question whether assigning disciplinary status to multimodality is either necessary or beneficial. The particular type of corpus they consider is one formed of YouTube 'Let's Plays', which are recorded to show a set of players playing a video game and simultaneously commenting on their play and the play of the others depicted. Data of this kind raises many challenging issues for any corpus-based approach and Schmidt and Marx show how an interactionist, pragmatic account drawing on Goffman's notion of participation frameworks has much to offer. This allows the authors to draw out particularly well the complex, diverse, and interacting levels of engagement that need to be teased apart and related in analysis when dealing with media use of this kind. The fact that the design of such corpora itself relies upon a broad range of disciplines, and may be required to support an equally diverse range of research questions, is what then leads Schmidt and Marx to argue that multimodality itself should not be considered a discipline because it lacks a specific research object of its own. Consequently, they suggest multimodality is better seen as contributing methods to other research disciplines and questions.

Lauren O'Hagan shows how multimodal analysis can also benefit from ethnohistorical approaches that appropriately situate the objects of analysis within their particular socio-historical contexts of use. By examining three sample texts taken from a larger corpus of inscriptions found in Edwardian books, O'Hagan draws out variations correlating with social class, showing culture and conflict concerning the practice of making book inscriptions and their precise multimodal composition. Far from being 'insignificant markers of ownership', O'Hagan shows how the detailed form of these inscriptions itself reveals a rich dynamic of rather different goals and aims among the distinct cultural groups whose inscriptions are analyzed, a perspective that may well be missed when considering these artifacts independently of context.

John Harnett discusses the diverse multimodal strategies and intermodal interactions employed in three exceptional graphic novels, David Mazzuchelli's *Asterios Polyp*, Alan Moore and Eddie Campbell's *From Hell*, and Chris Ware's *Building Stories*, in order to document some of the challenges that such works raise for any account of multimodality. Each of the graphic novels addressed is shown to employ its own particular constellation of modal resources and, in so doing, brings about very different narrative effects. The complexity of these interactions is itself seen as deliberately challenging readers' orientations to the visual and the textual as well as their modal competence. By drawing out discourse interpretations of the three works' narratives, Harnett demonstrates how close attention to multimodal design also requires being open to the fact that very different modal configurations can be made to serve narrative purposes.

Chris Taylor then reports on new developments in audiovisual description (AD), i.e., methods that improve access for blind and sight-impaired persons. Considering a significant extension of AD to include applications in museums and galleries, Taylor argues that there is a clear need to go beyond image/word symbiosis and bring in other senses in a truly multimodal approach to AD. For instance, the sense of touch can bring an artifact alive, though necessarily accompanied by an oral input. By describing several case studies showcasing new developments of AD in museums, Taylor addresses the question as to how diverse multimodal properties can benefit the growing community of persons with sight loss. Verbal description for the visually impaired, still the cornerstone of any attempt to recreate images in the mind, may well be usefully supported and enriched by other senses substituting vision.

Broadening the diverse medial contexts where multimodality comes into its own, **Edward Larkey's** chapter addresses a revealing source of information concerning cultural differences and similarities: the varied realizations of a TV show 'localized' to a broad range of target countries. The original TV show, produced in French in Canada, enjoys a wide circulation across the world, and Larkey selects

one scene in which gender, generational and family values take center stage for a detailed multimodal analysis. The producers of the scene in the different cultural contexts are shown to make rather different decisions concerning precisely what is said/depicted and how, clearly revealing specific cultural orientations to the issues enacted.

Wendy Nielsen, Pauline Jones, Helen Georgiou, Annette Turney, and Mary Macken-Horarik discuss applications of multimodality in an educational context, focusing on how the learning of science can be supported by having students produce digital explanations that naturally draw on a range of semiotic resources. In this approach, systemic-functional social semiotics and Legitimation Code Theory (Maton, 2013) are combined to consider science learning in a broader pedagogical context and to show ways in which concepts from multimodality theory can beneficially contribute to a shared language for critical reflection. Several kinds of digital explanations are considered and one artifact is worked through in detail and used to discuss both processes of knowledge construction and how such processes can be characterized drawing on the theories of multimodality adopted.

Finally, **Dušan Stamenković** and **Milan Jačević** consider the extent to which some proposals in the multimodal literature for addressing broadly ‘page-based’ media from a corpus-based perspective may also be applicable to certain aspects of video game interfaces. In their analysis they show that many of the components of a page-based scheme can indeed be applied to such interface screens, with the exception of some extensions that are clearly needed which they identify and discuss. They also by these means make the case that video games are certainly one of the media that multimodality should address and that doing so will in turn be useful for developing multimodality further.

Part IV then refers back to this introduction and the initial argument we develop here: In his Afterword, **John Bateman** takes up once again the issue of disciplinary status, comparing the situation of multimodality across several of the disciplines within which multimodality is now practiced and asking whether multimodality has its own specific contributions to make. In addition, and looking back on the contributions of the volume, he also considers in what respects the availability of a more explicit disciplinary anchoring specific to multimodality might help or hinder further developments. Again drawing on discussions from Legitimation Code Theory, Bateman assesses the relative capabilities of different forms of disciplinary discourse and applies the result to the state of play in multimodality, asking if further guidelines useful for development might be derived.

2.3 Breadth and Diversity as a Challenge

As has become apparent from these summaries, the chapters in this volume present a diverse range of disciplines and approaches to the analysis of multimodal artifacts and performances. As a matter of fact, the chapters do not constitute a representative sample of approaches typical of multimodal research in general.¹ Neither do the contributions document a joint effort in shaping a new discipline by establishing a shared interest or explicit connections to particular theoretical or methodological endeavors. One might even go as far as saying that the chapters only represent a ‘fraction’ of the various contemporary approaches to multimodality. In that sense, our book rather documents how the broad field is expanding into even newer directions and contexts with even more diversity and breadth in the use of theories, methods, and analytical frameworks.

Given this rapid growth of the field, we argue that this is actually one of the biggest challenges for multimodal research. However, in our view, it is a great potential at the same time: Capturing the mechanisms of synergetic knowledge construction and transmission via diverse forms of expression is an essential task in all research projects concerned with communication and it is likewise essential to have methods and frameworks available that follow this task systematically. As we argue in our own introductory textbook, multimodal analyses

[move] beyond bare recognition that there *are* such ensembles [of expression]: questions of *method* are going to become particularly important. Many issues at this time, for example, need to be seen as demanding empirical research—there is much that we simply do not know. What we *can* do at this stage, however, is to bring together robust starting points for analysis so that subsequent research can make progress—either for some practical task or research question that you need to find solutions or methods for, or as part of an educational process in some area of study where such questions have been found relevant. (Bateman et al., 2017, 19)

We therefore suggest that deliberate interaction (not competition!) between the respective disciplines make use of this diversity and breadth, and that such endeavors should ultimately be complemented by further “more inclusive accounts” that “move beyond the confines of individual disciplines” (Bateman et al., 2017, 20). We are convinced that this goal can only be reached through joint efforts in critically assessing and building disciplinary (and meta-disciplinary) foundations for a systematic analysis of communication that incorporates diverse methods and frameworks and builds bridges between them. If such connections are made

¹ Such more comprehensive overviews have been given in recent handbooks or introductory textbooks, some of which are mentioned in Section 4 below.

and channels of communication established, a sufficient degree of coherence in goals, concepts, and approaches and an exchange of ideas can continue to take place in the future—despite, or perhaps because of, further growth and increasing diversity.

In the remainder of this introduction as well as in the Afterword to this volume, we elaborate on how this common goal motivates us to explicitly call for strong discipline-building initiatives.

3 What Makes a Research Field a Discipline?

In comparison to the individual ideas voiced in the book's contributions for, or against, a discipline of multimodality, in this section we take up more explicitly the criteria and characteristics that would hold for a discipline. We begin with a characterization of the nature of disciplines articulated by Armin Krishnan. Then, to illuminate issues of disciplinarity further, we consider some particular reports drawn from the area of Digital Humanities, since this is a field that has itself relatively recently moved to claim the status of a fully-developed discipline. This comparison will allow us to discuss more deeply the potential of the current state of the art of multimodality to act as a discipline and to identify some of the common trajectories that can be observed in such developments.

3.1 Krishnan's Characteristics of a Discipline

As part of our previous initiatives and thinking about multimodality's disciplinary status, especially at the BreMM conferences and in the resulting conference proceedings (Seizov & Wildfeuer, 2017), we drew inspiration from Krishnan's (2009) working paper entitled "What are academic disciplines?", which was originally intended to consider and develop survival strategies for smaller disciplines amidst current trends towards larger interdisciplinary networks. In his discussion, Krishnan generally distinguishes between a technical dimension of the term *discipline* that combines "the organization of learning and the systematic production of new knowledge" and a moral dimension that includes "a specific and rigorous training that will turn out practitioners who have been 'disciplined' by their discipline for their own good" (Krishnan, 2009, 8). Going beyond the simple fact that "something is a subject taught in an academic setting", Krishnan provides a list of characteristics that indicate "whether a subject is indeed a distinct discipline" (Krishnan, 2009, 9). According to him, a discipline is characterized by:

1. a particular object of research (which may be shared with other disciplines),
2. a body of accumulated specialist knowledge (which is specific and not shared with other disciplines),
3. theories and concepts to organize the knowledge systematically,
4. a specific terminology and/or technical language,
5. specific research methods,
6. an institutional manifestation in the form of subjects taught at universities or colleges with academic departments and professional associations (Krishnan, 2009, see also Wildfeuer & Seizov 2017).

Krishnan emphasizes that a research field does not need to satisfy all of these criteria in order to be a stand-alone discipline. English literary studies, Krishnan argues, lacks a unifying theoretical paradigm or method as well as a neatly definable, uniform object of research; its status as a distinct discipline, however, hardly seems disputable. Accordingly, the category *discipline* is prototypically organized, with more central and more peripheral members. However, “the more of these boxes a [potential, J.W. et al.] discipline can tick, the more likely it becomes that a certain field of academic enquiry is a recognized discipline capable of reproducing itself and building upon a growing body of own scholarship” (Krishnan, 2009, 10).

Back in 2017, when compiling BreMM15’s proceedings (Seizov & Wildfeuer, 2017), our endeavors to chronicle recent developments in theory-building and methods, and to explore the synergies between previously disconnected concepts in semiotics and multimodality, ended in a call for the further work needed to make more significant moves towards establishing multimodality as a stand-alone discipline. We also concluded that book by giving Krishnan’s (2009) list (see Wildfeuer & Seizov, 2017), primarily with the intention of outlining possible milestones to guide future discipline-building ventures. Several years, many publications, and another BreMM conference later, we now have the chance to take up this thread again, and to return to Krishnan’s list in order to evaluate multimodality’s current disciplinary status. Before doing so, however, we will complement this list of criteria by looking more closely into another particularly active area of recent discipline-building: that of the Digital Humanities.

3.2 Similar Developments in Other Fields: Perspectives from Digital Humanities

In order to critically engage in a process of “disciplinarianisation and institutionalisation”, as Krishnan (2009, 10) phrases it more generally, and to target more

specifically the particular case of multimodality, we can complement Krishnan's list of features by considering similar discussions from other research fields. Transcending the traditional borders of the humanities and the natural sciences, for example, some fields and areas have only recently taken shape as disciplines in their own right and have, along the way, faced questions similar to those raised in this introduction for multimodality.

As suggested above, among the most recent and prominent examples are the Digital Humanities (DH), whose disciplinary status is a recurring theme in scholarly debate—even, and in fact predominantly, among its own practitioners. Several strategies have been pursued to move towards a precise definition of DH and its consolidation as a discipline. In the following, we turn to two recent accounts of DH's disciplinary status which will further inform our evaluation of the development of multimodality below.

Low-end and high-end disciplines.

Patrick Sahle, a prominent German DH-researcher and coordinator of the Cologne Center for eHumanities, has approached the issues under question here quite directly with a paper entitled “DH? But there is no such thing!” (Sahle, 2015). Sahle starts out by going through a checklist for distinguishing academic fields and disciplines similar to Krishnan's list of criteria in order to ensure an ‘unbiased’ assessment of strategies and characteristics present in the field. In addition to the aspects mentioned by Krishnan, Sahle lists characteristics such as a noticeably broad range of research projects, an organizational consolidation in associations, and a ‘living research community’, as well as the use of, and reference to, the discipline in popular media, i.e., beyond the academic context (Sahle, 2015, 1–4). Based on his assessment, Sahle concludes that DH has clearly reached the status of a “fully developed, independent subject” and an “autonomous discipline” (Sahle, 2015, 5).

However, he also characterizes DH as a ‘bridge subject’, spanning various fields and disciplines, and is thus essentially trans- and interdisciplinary in nature. At the same time, DH adopts a highly specialized position within individual discipline-specific fields. On this basis, Sahle claims that a discipline is not merely defined by its demarcation from other fields and disciplines, but also by its relations to them (Sahle, 2015, 5). Disciplines, he argues, evolve from specific subject-related movements within other disciplines and eventually lead to a further differentiation of the research landscape, a position also pursued by Muller (2011) in the context of Legitimation Code Theory, which will be returned to in the Afterword to this volume.

Sahle provides a definition of DH that reflects these considerations as follows:

The digital humanities deal with problems which go beyond the neighboring individual subjects in two ways: on the one hand, they deal with questions that apply to many fields; on the other hand, they deal with questions that cannot be addressed by other fields, since they are from their perspectives either too specific or demand competencies that are not available in those areas. (Sahle, 2015, 5, our translation)²

Calling DH a ‘bridge subject’, Sahle argues, allows one to locate its various manifestations in the research landscape and to provide a more differentiated description of its respective statuses: either as areas within other fields or disciplines (where it may initiate processes of transformation of traditional subject groupings), or as a discipline in its own right.

In reference to Burghardt & Wolff (2014) and their discussion of the relation between digital humanities and the humanities more broadly, Sahle (2015, 6) suggests assessing DH’s disciplinary status by placing it on a scale from a ‘low-end’ to a ‘high-end’ discipline. A low-end disciplinary status shows in the use of a (potential) discipline’s generic tools in other areas and disciplines, while a high-end disciplinary status involves the development of discipline-specific tools and methods, as well as their ongoing critical assessment and theoretical reflection. These ideas are rendered in graphic form in Figure 1, which also points out that the relations and interactions between fields of study are central to discussions of the extent to which DH is justifiably to be considered a discipline, especially when precise demarcations vary.

Theoretical versus applied dimensions and a discipline’s research object(s).

In a more recent proposal, Michael Piotrowski (2018) identifies the “confounding [of] a number of related, but actually distinct issues” as a major difficulty in defining the field of DH, and delivers an extended “explication” intended to clarify the discussion:

In short, we posit that any field of research (regardless of whether one wants to consider it a discipline or not) is ultimately defined by a unique combination of (1) a *research object* and (2) a *research objective*. Research methods are secondary in that they are contingent on the research object and the research objective, as well as on technical and scientific progress, which both requires them to adapt and permits them to evolve, whereas the research object and the research objective remain stable. Furthermore, disciplines have always used a

² “Die Digital Humanities befassen sich mit Problemen, die über die benachbarten Einzelfächer in einem doppelten Sinne hinausgehen. Zum einen betreffen sie Fragen, die für viele Fächer gleichermaßen gelten; zum anderen betreffen sie Fragen, die von den benachbarten Fächern nicht behandelt werden, weil sie aus ihrer Sicht zu speziell sind oder Kompetenzen erfordern, die in den Fächern nicht enthalten sind.” (Sahle, 2015, 5)

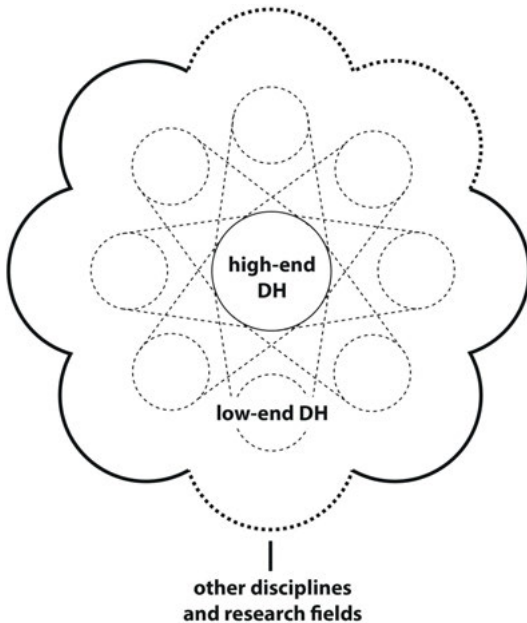


Fig. 1: The (current) relations of DH with other disciplines. DH research activities reach out into several disciplines and research fields as generic digital tools are applied. Each of these areas constitute the low-end poles of scales that all extend into an area of high-end DH, where DH-specific methods are developed and theoretical reflection takes place.

variety of methods; for example, while qualitative methods are certainly ‘typical’ for many humanities disciplines, quantitative methods have always been used as well. This means that it is not useful to attempt to define digital humanities (or any other field or discipline for that matter) by way of the methods it happens to use at some point. (Piotrowski, 2018, 2)

On this basis, Piotrowski provides a two-fold definition of DH that is again meant to clarify DH’s relationship to other disciplines, subdividing the discipline into theoretical DH and applied DH. While theoretical DH refers to “research on and development of means and methods for constructing formal models in the humanities”, applied DH means “the application of these means and methods for the construction of concrete formal models in the humanities disciplines” (Piotrowski, 2018, 2). This distinction is similar to Sahle’s concepts of a high-end discipline, which corresponds to the theoretical dimension, and a low-end discipline, which refers to the applied dimension.

Although perhaps falling back a little too much on the traditional, and not always helpful, distinction between theory and application, Piotrowski’s main interest is in articulating further DH’s relationship to other disciplines, and es-

pecially to computer science. As described above, Krishnan also considers this feature to be relevant for the description of a discipline but emphasizes that research objects may be shared by several disciplines. For Piotrowski, in contrast, it is more important to define a discipline-specific research object in order to be able to distinguish the field from other disciplines and to describe it on a theoretical level:

Research in theoretical digital humanities (which is as yet rare) is obviously motivated by the requirements of research in the humanities, but has its own research object: the means and methods for constructing formal models in the humanities. (Piotrowski, 2018, 3)

Such a clarification is also crucial for the context of multimodality, particularly when it comes to the description of the discipline's own research object. We take these ideas further in the next section.

4 Where are we? Assessing Multimodality's Disciplinary Status

Drawing on the criteria established above, we can now assess to what extent multimodality's current status can be seen as that of a discipline in its own right. As a first step, we draw on Krishnan's (2009) list of discipline-specific characteristics as outlined in Section 3.1 and discuss these further with regard to multimodality's compliance with these criteria. As a second step, we turn to the criteria elicited from the discussion within and around Digital Humanities in order to provide a frame of reference and a stronger basis for a discipline of multimodality to emerge. With this, we explicitly do not aim at giving a full overview of multimodality's current status, but rather to demonstrate its immense diversity and breadth together with the challenges these bring with them for the discipline to be built.

4.1 Multimodality's Status in Light of Krishnan's Features of a Discipline

With regard to both long-term and recent developments within the field of multimodality, we can critically apply Krishnan's list of characteristics as follows:

(1) Object of research. Today, multimodality's objects of research encompass a broad variety of artifacts, performances, or, more generally, communicative situations in which diverse forms of expression are deployed (cf. Bateman et al., 2017,

7). Examples are provided by all of the contributions to this book. This diversity of research objects, and the even broader range of objects found in multimodality in general, attests to and results from multimodality's equally strong concern with investigating "the combination of different semiotic resources, or modes, in texts and communicative events" (Adami, 2016, 451) wherever they might occur. It is this orienting focus that has driven multimodalists to explore various media and genres of increasing modal complexity with regard to relations between and across modes against the background of an ongoing diversification of the contemporary communicative landscape. As scholars face these analytical challenges and uncover the particularities of multimodal artifacts and their meaning making, new theoretical proposals emerge and empirical questions raise ever more pressing issues. Such developments, in turn, have informed theory-building and, not least, also allow for revisiting analytical approaches to traditional media and genres. We can therefore with justification point to the meta-reflection of just what multimodality is as a phenomenon in its own right as an equally prominent and definitory 'object' of research, in addition to any individual artifacts, genres, or media that may be addressed.

(2) Body of accumulated specialist knowledge. So far, the knowledge gathered and generated in response to the specific research interests and questions posed by multimodalists comprises both theorization and empirical exploration and primarily concerns two main areas: first, the very nature of a semiotic mode, its constitutive elements, and its mode-specific mechanisms of meaning-making, and, second, the structural and functional linkage of modes as part of actual meaning-making processes accomplished by means of artifacts and performances and in specific media environments and communicative situations. However, "with the current state of the art in multimodality research", Bateman (2014, 239) argues, "there are far more questions than answers and so there is much to be explored". As indicated above, theoretical concepts and frameworks have tended to lack consistency and precision, which, once such concepts are drawn on in empirical analyses, often hampers the comparability of results. Also, numerous small-scale case studies, but also larger-scale projects, have generated results that remain tentative and preliminary; more generalized conclusions made on their basis then necessarily also remain hypothetical and demand validation before being accepted. For significant progress to be made, these empirical issues need to be pursued further by investigating larger collections of data from a multitude of perspectives (cf. Bateman, 2014)—ultimately revisiting, qualifying, and even revising earlier theoretical assumptions.

(3) Broader theories and concepts. As suggested above, due to its truly inter- and transdisciplinary nature, there are many fields and disciplines that have informed—and continue to inform—work in multimodality. Linguistics is a par-

ticularly prominent case, as discussed, for example, in Stöckl's contribution to the present volume. Bezemer & Jewitt (2018, 283) consider in detail some of the repercussions that such interdisciplinary ties and exchanges of theories, concepts, and methods may have, noting: "Within linguistics, as indeed in the other disciplines that contribute to the field of multimodality, there's considerable variation in terms of theoretical and methodological outlook." In other words, multimodalists adopt and integrate into their own approaches and frameworks a multitude of theories and concepts as widespread and diverse as the disciplines and fields they interact with. Particularly in the early years of multimodality research, scholars drew on the explanatory power of these adopted concepts and terms and used them productively as a starting point for analyses. The benefits of adopting other disciplines' theories and concepts were thought to outweigh the drawbacks. Such integrative work can, however, lead to conceptual vagueness as well as importing contentious or problematic frameworks and concepts that stumble on the pitfalls of quite specific discipline-internal debates. As work has grown in depth and scope, it is natural that conceptual shortcomings have been discovered, to an increasing extent prompting important processes of terminological renegotiation, refinement, complementation, and replacement. As awareness of the specific research interests of multimodality spreads, more particular and, at the same time, more broadly conceived theories of multimodality and their concepts are being proposed which need then to be related both back to existing terms and to data and empirical studies.

(4) Specific terminology. As suggested above, the increase in multimodality's visibility as a field has been paralleled by the introduction of specific terms and frameworks. As one might expect, much debate has revolved around multimodality's eponymous concept, *mode*, but has also involved (and been influenced by) several immediately related terms that were very prominent from the outset in the multimodality literature, such as *semiotic resources*, *affordances*, and the notion of *materiality* (again see Stöckl, this volume, for an overview of these and many other terms). *Mode*, in particular, has been defined in various, diverging ways, ranging from reductionist conceptualizations on the basis of the five (external) perception channels (see, e.g., Granström et al., 2002, 1) to highly inclusive modelings responsive to "the effect of the work of culture in shaping material into resources for representation" (Jewitt & Kress, 2003, 1–2). Entities as diverse as "written text, spoken language, gestures, facial expressions, pictures, drawings, diagrams, music, moving images, comics, dance, typography, page layout, intonation, [or] voice quality" (Bateman et al., 2017, 16) have each been labeled *modes* in one approach or another. Consequently, over ten years ago, Forceville (2006, 382) had already concluded that "it is at this stage impossible to give either a satisfactory definition of 'mode', or compile an exhaustive list of modes" and today we still have "very

fragmented knowledge of just what semiotic modes exist” (Bateman et al., 2017, 136). This is leading scholars to increasingly stress the importance of empirical analysis to see just which modes participate in a given multimodal artifact (e.g., Bateman, 2008, Thomas, this volume), while more elaborate and robust models of mode are promoted precisely with the aim of supporting analytical work of this kind (e.g., Bateman, 2011, 2016; Bateman et al., 2017; Stöckl, 2016).

Furthermore, terms have been suggested to capture more adequately the ‘interplay’ or ‘interaction’ of modes in processes of meaning making. There has always been much variation in the use of the two roots {*semio*-}, as in “intersemiotic complementarity” (Royce, 1999, 2007), “resemiotization” (Iedema, 2001), “intersemiosis” (Unsworth, 2007), or “intersemiotic texture” (Liu & O’Halloran, 2009), and {*mode*}, as in “in between modes” (Stöckl, 2004), “intermodal relations” (Caple, 2008), or “modal density” (Norris, 2004). There is also talk of “cross-modal” (El Refaie, 2013) relations and meaning-making processes and “transmodal moment[s]” (Newfield, 2014), further focusing attention. Multimodalists consequently show increasing agreement in their use of terms, such as *mode*, in order to orient their discussions, even if the definitions they offer of those terms may remain less than optimally explicit.

(5) Specific research methods. The move towards establishing theoretical concepts and frameworks specifically designed for analyzing multimodal artifacts and performances has been paralleled by adopting, advancing, and developing a rich inventory of methodological approaches. For the purposes of a structured overview, they are grouped here under the headings of *product*, *reception*, and *production*.

The bulk of studies carried out under the flag of multimodality has focused on multimodal artifacts and performances in the broad sense of textual **products** and dynamic discourses. Here, as suggested above, multimodality’s methodological toolkit has been informed by disciplines such as linguistics, discourse analysis, or semiotics. Accordingly, prominent approaches, e.g., social semiotics (Kress & van Leeuwen, 2006 [1996]; van Leeuwen, 2005) or systemic-functional multimodal discourse analysis (O’Halloran, 2008), go back to extensions and advances of systemic-functional linguistics, importing much of the technical machinery of those frameworks.

In a similar fashion, but as an independent branch, the language-based area of conversation analysis has also turned to essentially multimodal phenomena of interaction, spurring the development of a variety of approaches foregrounding the multimodality of interaction or interactional achievements (e.g., Norris, 2004; Deppermann, 2013; Streeck, 2013; Mondada, 2016). At the same time, approaches from other disciplines such as ethnography now inform methods in multimodality and have fueled work aiming at even more integrative methodological frameworks

(see, e.g., Dicks et al., 2011, O'Hagan, this volume). In response to the demand for a more significant move towards an empirical grounding of concepts such as mode, moves towards investigations of larger multimodal corpora are widespread (e.g., O'Halloran et al., 2016; Bateman et al., 2016; Tan et al., 2018). Such a move has led scholars to adapt and synthesize existing approaches (e.g., corpus-assisted multimodal discourse analysis, see Bednarek & Caple, 2014, 151) and also to employ methods from computer science to develop tools, e.g., for (semi-)automated analyses of images or page layouts (e.g., Hiippala, 2015; O'Halloran et al., 2016; Rigaud & Burie, 2018; Tan et al., 2018). Many empirical challenges consequently remain.

In comparison with product- and discourse-related methods, **multimodal reception analysis** can still be considered an emerging strand of research. While methods such as “thinking aloud during and after the reception process, re-narrations, knowledge tests, interview, or questionnaires” (Bucher, 2017, 92) have long been part and parcel of media reception studies, more recent technological advances have added eye-tracking methods to the multimodal reception analysis toolbox. Early empirical work in this area of multimodality dates back to the mid- and late 2000s (see, e.g., Holsanova, 2008), building on experimental traditions in reading research in the 1980s and even earlier work in visual perception, such as the pioneering studies of Yarbus (1967). Ever since, eye-tracking methods have been employed to study recipients' interaction with multimodal artifacts as diverse as print newspaper spreads (Holsanova et al., 2006), online newspapers (Bucher & Schumacher, 2006), powerpoint presentations (Bucher & Niemann, 2012), comics and graphic novels (Foulsham et al., 2016; Bateman et al., 2018; Tseng et al., 2018), as well as film (see Smith 2012 and Reinhard & Olson 2016 for an overview). Eye-tracking methods allow for critically assessing claims often made previously on the basis of textual products alone and have the potential to “[take] multimodality into a new direction” (Jewitt et al., 2016, 12). Moreover, in addition to eye-tracking, which is becoming relatively widespread as the necessary equipment becomes more accessible, other methods, such as brain imaging of various kinds, are also now beginning to find application in the multimodal context. Examples here include neurocinematics, which uses techniques such as EEG (electroencephalogram) and fMRI (functional magnetic resonance imaging) to find neural correlates of film reception (see, e.g., Hasson et al., 2008; Özerdem & Polat, 2017; Zhao et al., 2018), and Cohn's extensive event-related potential, or ERP-based studies of the neural processing of static visual narratives (see, e.g., Cohn et al., 2012).

Deeper investigations into professional **production** contexts and the professional practice of multimodal artifact design are still relatively rare, even though practitioners' perspectives (e.g., Waller, 2012) have already shown considerable

potential for enriching our understanding of multimodal meaning making. At the same time, media studies scholars have also pointed to the benefits of considering the production-side: micro-level analyses could illuminate how media actors systematically combine forms of expression such as language, image, and sound, and what knowledge, debates, and decisions are involved in the production of such texts. On a meso-level, production processes may show patterns and styles typical of particular editorial departments, for instance. These, in turn, may evolve to form more general discursive patterns tied to broader journalistic-cultural and national areas on a macro-level (cf. Klemm et al., 2016, 297). However, both in media studies and in multimodality, production studies have largely remained on the sidelines.

(6) Institutional manifestation. Over the past decades, an increased interest in multimodal research has also become manifest, though to varying extents, through institutionalization and publications of several kinds that go beyond simply engaging with communicative phenomena that are ‘multimodal’ in some respect. We can characterize these more specifically as follows:

(a) So far, multimodality has perhaps shown its strongest disciplinary consolidation through publications that deal explicitly and predominantly with multimodality and its core research interests as introduced above, e.g., in handbooks and collections (e.g., O’Halloran, 2004; Jewitt, 2009b, 2014; Norris, 2015; Klug & Stöckl, 2016), in dedicated book series (e.g., the Routledge Studies in Multimodality ed. by O’Halloran, since 2011); in dedicated textbooks (e.g., Machin, 2007; Jewitt et al., 2016; Bateman et al., 2017), as well as in peer-reviewed journals with a clear focus on multimodal issues, e.g., *Multimodal Communication* (since 2012), the *Journal of Multimodal Communication Studies* (2014–2017), the *Journal of Multimodal Rhetorics* (since 2017), or *Visual Communication* (since 2002).

(b) A further move towards institutionalization and community-building has been achieved through the establishment of international conferences, e.g., the International Conference on Multimodality (ICOM), a biennial conference, celebrating its 10th event in 2020; the more interaction-focused International Conference on Multimodal Communication (ICMC), held in Osnabrück in 2017 and in Hunan in 2018 with a new event scheduled for 2020 in Osnabrück; and—closely tied to this present collection—the Bremen Conferences on Multimodality, a series of conferences that began in 2014 and which will have its 4th outing in 2019. Nevertheless, a larger number of regular events (even if smaller in scope) would allow for more focused discussions of core-disciplinary themes, recent advances, and current challenges, as well as strengthening the community and increasing multimodality’s visibility as a discipline-to-be through further publications.

(c) The establishment of competence centers and labs solely devoted to the study of multimodality has also contributed to the instantiation of the field as a discipline. Institutions such as the Centre for Multimodal Research at the Institute

of Education, University College London, UK, the Multimodal Analysis Lab at the National University of Singapore, the AUT Multimodal Research Centre in Auckland, New Zealand, the multimodality research group at the University of Bremen, the Natural Media Lab in Aachen, or the Distributed Little Red Hen Lab (a global consortium for research in multimodal communication with many international collaborators)—to name just a few—are encouraging researchers to work in close-knit teams, pool their strengths, share resources, and represent their institutions (and the research field) jointly, not least to acquire funding.

(d) Much slower has been institutionalization through the establishment of professorships with an explicit mention of ‘multimodal’ as part of their denomination—indeed, in most parts of the world, multimodality has in this respect been close to invisible for a long time. However, in recent years, this situation has started to improve and in many countries there are now full professorships with explicit foci on multimodality. To take Germany as just one example, we now find three chairs with an explicit (i.e., named) multimodal specialization: the Chair for German Linguistics, Semiotics and Multimodal Communication (Prof. Dr. Ellen Fricke) at TU Chemnitz, the Chair for Language Use and Multimodal Communication (Prof. Dr. Cornelia Müller) at the Viadrina University in Frankfurt/Oder, and the Chair for the Theory and History of Multimodal Communication (Prof. Dr. Silvia Kutscher) at the Department of Archeology at Humboldt-University in Berlin. However, by far the majority of full professors working in the field of multimodality hold chairs associated (and explicitly labeled) with areas such as (applied) linguistics, media and communication (studies), semiotics, or education, and so clearly reflect a variety of other disciplinary anchor points. Thus, while some first steps towards institutionalization through professorships have been taken, we are still far from having the same degree of institutional recognition as, e.g., historical linguistics, sociolinguistics, media and communication studies, or the digital humanities, as we will see further below.

(e) Perhaps not surprisingly, the situation is similar when we turn to multimodality-related study programs and courses. The MA-course in Semiotics and Multimodal Communication at TU Chemnitz and the MA-course in English-Speaking Cultures at the University of Bremen are rare examples of study programs with explicit specializations in multimodality, although the Universities of Amsterdam, Salzburg, Groningen, and Sydney, among others, also offer BA/MA-programs featuring classes in multimodality. Even though multimodality’s visibility may still be limited here, scholarly works are read and discussed, interests are piqued, and future scholarship in the form of MA theses or PhD projects is inspired.

In sum, we can conclude that many of the criteria/boxes in Krishnan’s list can indeed be ticked for multimodality, although partly with limitations and reservations and with considerable potential for future expansion.

4.2 Multimodality's Status in the Light of DH as a Recently Emerged Discipline

Building on this first step, we now extend our assessment of multimodality's current status by focusing more on the 'disciplinary' discussions pursued with respect to the field of Digital Humanities.

In terms of institutionalization and self-identification, DH has clearly proceeded much further in its disciplinary development than multimodality. This becomes most apparent in the significant growth in the establishment of academic positions, departments, and study programs all over the world in recent years. Considerations in various departments and institutions concerning whether they should integrate perspectives and approaches from DH are relatively common today—despite uncertainties with regard to its precise definition. In contrast and as indicated above, only few institutions have announced positions with an explicit multimodality-related denomination, or at least focus, in the position's profile. Accordingly, only few institutions seem to consider it relevant to incorporate multimodality into their research and application portfolios, despite the fact that an increasing proportion of the objects of study are inherently and complexly multimodal.

With regard to Sahle's (2015) continuum between the conceptual poles of a low-end and high-end discipline, we find indications for multimodality occupying several locations simultaneously, depending on the aspects one investigates. Similarly to DH, multimodality is essentially transdisciplinary in nature, that is, multimodal outlines can be traced within many different fields or disciplines; accordingly, certain parts of multimodal research activities are located towards the *low end* of the scale. However, multimodal research within other disciplines not only encompasses the use of multimodal terminology and frameworks when analyzing another discipline's research objects, but has also initiated processes of transformation within respective disciplines. This can for instance be seen in the development of the field of interaction analysis in linguistics, which, as noted above, no longer only focuses on verbal language in interaction. In addition, even if this has not been echoed with the same level of engagement in text linguistics, studies of written language now commonly draw on accompanying visual features as well. Similarly, the visual sciences have also begun to take into account a variety of modes with respect to their interplay in visual artifacts. Furthermore, as noted above, steps have been taken towards a more theoretically and methodologically grounded field of multimodality by means of introductions to the field for practitioners and students through handbooks and textbooks; these aspects show that multimodality has already extended to cover further parts of the range, much closer to the *high end* of Sahle's suggested scale.

Such an assessment becomes even more plausible if one takes into account the fact that multimodality is practiced by researchers who increasingly view themselves as experts and specialists in *multimodality itself*, and who consequently continue to expand and consolidate the field—or discipline—further. Even if multimodality is still not cited as an independent, fully-fledged discipline as often is the case with DH (see, e.g., the contributions in Gold, 2012; Terras et al., 2013; Berry & Fagerjord, 2017), multimodality’s ‘lived community’ frequently gathers at conferences and interacts through various multimodality-specific mailing lists. Last but not least, as attestable in a small number of sources, including our own work, scholars in the field have begun to refer to themselves as ‘multimodalists’ as well (see, e.g., Priem & Thyssen, 2013; Wildfeuer, 2015; Bezemer & Jewitt, 2018).

Piotrowski’s view of DH as both a theoretical and applied discipline, providing its own means and methods to define its research object, also supports a more detailed characterization of these various manifestations of multimodality. Whereas many approaches and analyses within multimodality indeed orientate more towards an application of concepts and models in analytical work within their own disciplines and fields, there are also more concrete and elaborate efforts being undertaken towards theory-building, aiming at defining and describing multimodality’s core research object: semiotic and communicative modes, material, semiotic resources, and so on.

This has been highlighted and repeated by many researchers in the field over the years. Already in 2000, Theo van Leeuwen and Carey Jewitt were underlining in the introduction to their *Handbook of Visual Analysis* the importance of the “nature of the project [...], the visual material that is being investigated” (van Leeuwen & Jewitt, 2000, 5). This also holds for any multimodal object of investigation: it is the very nature of the multimodal artifact or performance, i.e., the ‘multimodal material’, which needs to stand in the foreground of analysis, from both a theoretical, a methodological, and an empirical perspective. It is this particular focus on the material units of analysis that then often actually makes the various approaches and studies involved ‘multimodal’; Jewitt et al. (2016, 5–6) similarly emphasize this when “doing multimodality” and “adopting multimodal concepts”. And it is this focus “on the role of modalities and their interaction in their own right [that] provides a more systematic path from research questions to the multimodal phenomena lying at the heart of the enterprise” (Bateman et al., 2017, 382). On this basis, specific endeavors and projects of this kind are most plausibly located at the theoretical level of the discipline, suggesting once more that the development towards a ‘high-end discipline’ is already well underway.

5 Conclusions and Outlook

With this introduction, we have sought to continue the discipline-building explorations that have naturally evolved in multimodality's development over the past decades and, as have been pursued in discussions led at previous BreMM conferences and in subsequent publications. In particular, BreMM15's conference proceedings closed with a list of criteria for elevating a research field to a stand-alone discipline (Krishnan, 2009; Wildfeuer & Seizov, 2017), which we have expanded on here and applied in order to assess multimodality's current disciplinary status.

Krishnan's features of a discipline have allowed us to take stock of multimodality's achievements with respect to objects of research, its accumulated specialist knowledge, broader theories and concepts, specific terminology and research methods, and institutional manifestations. Our assessment has made it evident that multimodality's research objects may be diverse, but its research interests can be spelled out rather clearly: *multimodality investigates meaning made with various semiotic modes, while specifically focusing on the nature of individual modes as well as on the mechanisms of their structural and functional combination*. Our brief look at multimodality's informing theories and frameworks has reaffirmed the field's strong inter- and transdisciplinary past and present—relations that need to be retained for disciplinary advancements through future processes of conceptual integration.

As suggested by Halliday (2003), transdisciplinary research is to be distinguished from inter- or multidisciplinary research. The latter still pursues research focused within the disciplines involved, while the real alternative according to Halliday is to transcend disciplinary boundaries in order to achieve the kind of integrated focus necessary to address, for example, issues at the core of multimodality. In recent years, this kind of transdisciplinary approach to addressing multimodal issues has been productive and quite crucial, particularly for the understanding of emerging communicative formats integrating multiple meaning-making systems, multiple text production devices, and multiple media channels.

At the same time, it has become apparent that it seems desirable to take further steps towards establishing conceptual ties and, ultimately, more robust frameworks in order to strengthen multimodality's theoretical core and move further towards a stand-alone discipline. Developing an understanding of the social and cultural impact and the potential of rapidly evolving communication formats necessitates a broad conceptual framework that brings together the multiple dimensions on offer in order to function as a unified resource for theory-building and practice. In this respect, the multimodality-specific terms that have already been introduced, applied, and renegotiated constitute a strong foundation.

Nevertheless, while the remarks above already suggest some of multimodality's methodological strengths in investigating multimodal artifacts in the sense of textual products, it is only relatively recently that a more deliberate move towards larger-scale empirical investigations has occurred. Several authors, often quite independently of one another and in relation to different subject matters, have discussed this requirement for moving the field on (cf., e.g., Stöckl, 1997; Bateman et al., 2004; Gu, 2006; Carter & Adolphs, 2008; Nakano & Rehm, 2009; Bednarek, 2015; Hiippala, 2015; Pederson & Cohn, 2016), and so this should now certainly be considered an inescapable and necessary step for the field as a whole; several of the contributions to the current volume address this concern. The development and application of methods for multimodal production and reception analysis also require more attention in order to resolve continuing methodological imbalances.

By drawing on Sahle's and Piotrowski's discussions, therefore, we have brought some further precision to the assessment of multimodality's current disciplinary status, not least because both authors considered precisely analogous issues with respect to DH as a recent case of an emergent discipline with which multimodality can readily be compared. The application of their concepts of a continuum ranging from 'low-end' to 'high-end' disciplines suggests that multimodality is evolving into a field (or discipline) with manifestations of several kinds, similarly to the Digital Humanities, and which cover a broad range of the scale from theoretical to application concerns. In a manner comparable to DH, then, multimodality not only exhibits features of an applied discipline but now shows clear tendencies to develop further towards being a theoretical discipline as well—in some respects, arguably, already going beyond DH in this regard (cf. Bateman, 2017).

On this basis, the illustration in Figure 2 attempts to set out graphically more clearly where we are and where we ought to be going if multimodality is to be established as a discipline in its own right. Local movements towards multimodal work within other fields and disciplines mark an early stage of multimodality's disciplinary development (Figure 2, Stage 1). Over time, multimodal applications became more frequent and elaborate, resulting in a considerably diverse and broad research area. At the same time, both theory- and methods-related discourses about multimodality began to surface. By this stage (Stage 2), the field began to gain more visibility with respect to its own unique research objects and interests, its theoretical terms, as well as its methods. Most recently (Stage 3), advancements in theory and methods have evolved into a more integrated body of knowledge that is available for informing, as well as being informed, by applied research.

The contributions to this book reflect all these aspects and interactions as facets of an ongoing enterprise among networked researchers from different research realms who engage in transdisciplinary studies involving their own specializations

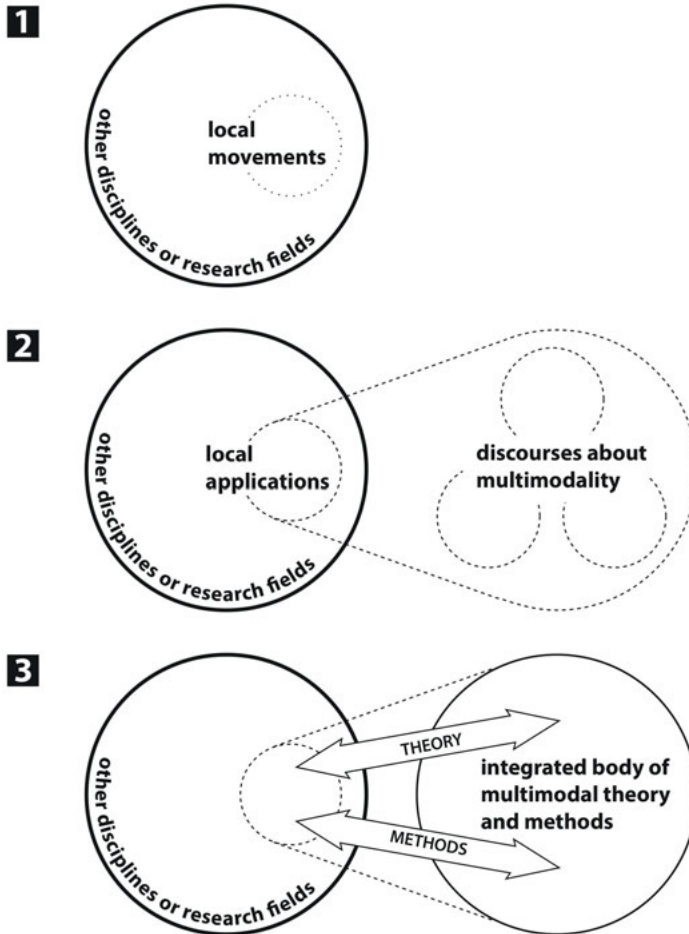


Fig. 2: Multimodality's development across three stages. In stage 1, multimodality-related local movements within other fields/disciplines are marked out only roughly. In stage 2, the visibility of multimodal research activities increases so that a diverse range of local applications are complemented and informed by a collection of discourses about multimodality more generally. By stage 3, these discourses coalesce into a body of multimodal theory and methods; the area of discipline-embedded applications and the area of developing multimodal theory and methods then mutually inform each other and are thus simultaneously and reciprocally modified.

while also developing work in multimodality. The chapters consequently revisit and redefine theoretical or empirical analyses, which are crucial to the study of multimodality from various perspectives, with a view towards evolving issues of

multimodal analysis. Due to the broad range of aspects dealt with and diversity of approaches covered, they attest to the continuing expansion of the field and are relevant for helping to shape a roadmap for further discipline-building explorations to come. As we have seen with regard to DH, preserving such ties fosters fruitful exchange and innovation. Establishing a more robust disciplinary core, however, seems desirable for a discipline to gain visibility, to generate awareness of its explanatory potential, and to ensure further growth. Therefore, even though a long road still lies ahead, the ground has certainly been broken for more decided moves towards what might then emerge as a genuinely new discipline of multimodality.

Acknowledgements

We are grateful to Emilia Djonov and Hartmut Stöckl for making valuable comments on an earlier version of this manuscript, to Theo van Leeuwen and the members of the Centre for Multimodal Communication at the University of Southern Denmark in Odense for feedback on early thoughts on the topic as well as to the anonymous reviewers that have critically evaluated our thoughts of the book and helped us sharpen its main outline and objectives.

Bibliography

- Adami, Elisabetta. 2016. Multimodality. In Ofelia Garcia, Nelson Flores & Massimiliano Spotti (eds.), *Oxford Handbook of Language and Society*, 451–472. Oxford: Oxford University Press.
- Bateman, John A. 2008. *Multimodality and Genre: A Foundation for the Systematic Analysis of Multimodal Documents*. Basingstoke: Palgrave Macmillan.
- Bateman, John A. 2011. The Decomposability of Semiotic Modes. In Kay L. O'Halloran & Bradley A. Smith (eds.), *Multimodal Studies: Multiple Approaches and Domains*, 17–38. London: Routledge.
- Bateman, John A. 2014. Using Multimodal Corpora for Empirical Research. In Carey Jewitt (ed.), *The Routledge Handbook of Multimodal Analysis. Second Edition*, 238–252. London: Routledge.
- Bateman, John A. 2016. Methodological and Theoretical Issues for the Empirical Investigation of Multimodality. In Nina-Maria Klug & Hartmut Stöckl (eds.), *Sprache im multimodalen Kontext / Language and Multimodality* (Handbooks of Linguistics and Communication Science (HSK) 7), 36–74. Berlin: de Gruyter Mouton.
- Bateman, John A. 2017. Multimodale Semiotik und die theoretischen Grundlagen der Digital Humanities. *Zeitschrift für Semiotik* 39(1–2). 11–50.
- Bateman, John A., Annika Beckmann & Rocío Varela. 2018. From Empirical Studies to Visual Narrative Organization: Exploring Page Composition. In Alexander Dunst, Jochen Laubrock & Janina Wildfeuer (eds.), *Empirical Comics Research. Digital, Multimodal, and Cognitive Methods*, 127–153. London and New York: Routledge.

- Bateman, John A., Judy L. Delin & Renate Henschel. 2004. Multimodality and Empiricism: Preparing for a Corpus-based Approach to the Study of Multimodal Meaning-Making. In Eija Ventola, Cassily Charles & Martin Kaltenbacher (eds.), *Perspectives on Multimodality*, 65–87. Amsterdam: John Benjamins.
- Bateman, John A., Chiao-I Tseng, Ognyan Seizov, Arne Jacobs, Andree Lüdtke, Marion G. Müller & Otthein Herzog. 2016. Towards Next-Generation Visual Archives: Image, Film and Discourse. *Visual Studies* 31(2). 131–154. <https://doi.org/10.1080/1472586X.2016.1173892>.
- Bateman, John A., Janina Wildfeuer & Tuomo Hiippala. 2017. *Multimodality – Foundations, Research and Analysis. A Problem-Oriented Introduction*. Berlin: Mouton de Gruyter.
- Bednarek, Monika. 2015. Corpus-Assisted Multimodal Discourse Analysis of Television and Film Narratives. In Paul Baker & Tony McEnery (eds.), *Corpora and Discourse Studies: Integrating Discourse and Corpora*, 63–87. Basingstoke: Palgrave Macmillan. https://doi.org/10.1057/9781137431738_4.
- Bednarek, Monika & Helen Caple. 2014. Why Do News Values Matter? Towards a New Methodological Framework for Analyzing News Discourse in Critical Discourse Analysis and Beyond. *Discourse & Society* 25(2). 135–158.
- Berry, David M. & Anders Fagerjord (eds.). 2017. *Digital Humanities: Knowledge and Critique in the Digital Age*. Cambridge and Malden: Polity Press.
- Bezemer, Jeff & Carey Jewitt. 2018. Multimodality: A Guide for Linguists. In Lia Litosseliti (ed.), *Research Methods in Linguistics*, 281–304. New York: Bloomsbury.
- Björkqvall, Anders. 2012. Multimodality. In Jan-Ola Östmann & Jeff Verschueren (eds.), *Handbook of Pragmatics*, 1–20. Amsterdam: John Benjamins.
- Bucher, Hans-Jürgen. 2017. Understanding Multimodal Meaning-Making. Theories of Multimodality in the Light of Reception Studies. In Ognyan Seizov & Janina Wildfeuer (eds.), *New Studies in Multimodality. Conceptual and Methodological Elaborations*, 91–124. London: Bloomsbury.
- Bucher, Hans-Jürgen & Philipp Niemann. 2012. Visualizing Science: The Reception of PowerPoint Presentations. *Visual Communication* 11(3). 283–306.
- Bucher, Hans-Jürgen & Peter Schumacher. 2006. The Relevance of Attention for Selecting New Content. An Eye-Tracking Study on Attention Patterns in the Reception of Print- and Online Media. *Communications. The European Journal of Communications Research* 31(3). 347–368.
- Burghardt, Manuel & Christian Wolff. 2014. Digital Humanities : Buzzword oder Strukturwandel der Geisteswissenschaften? Stand und Perspektiven anhand Regensburger Beispiele. *Blick in die Wissenschaft* 25. 39–46.
- Caple, Helen. 2008. Intermodal Relations in Image Nuclear News Stories. In Len Unsworth (ed.), *Multimodal Semiotics: Functional Analysis in Contexts of Education*, 125–138. London: Continuum.
- Carter, Ronald & Svenja Adolphs. 2008. Linking the Verbal and the Visual: New Directions for Corpus Linguistics. *Language and Computers* 64. 275–291.
- Cohn, Neil, Martin Paczynski, Ray Jackendoff, Phillip J. Holcomb & Gina R. Kuperberg. 2012. (Pea)nuts and Bolts of Visual Narrative: Structure and Meaning in Sequential Image Comprehension. *Cognitive Psychology* 65(1). 1–38.
- Constantinou, Odysseas. 2005. Review Article: Multimodal Discourse Analysis: Media, Modes and Technologies. *Journal of Sociolinguistics* 9(4). 602–618.
- Deppermann, Arnulf. 2013. Multimodal Interaction from a Conversation Analytic Perspective. *Journal of Pragmatics* 46(1). 1–7.

- Dicks, Bella, Rosie Flewitt, Lesley Lancaster & Kate Pahl. 2011. Special Issue: Multimodality and Ethnography: Working at the Intersection. *Qualitative Research* 11(3).
- El Refaie, Elisabeth. 2013. Cross-Modal Resonances in Creative Multimodal Metaphors. *Review of Cognitive Linguistics* 11(2). 236–249.
- Feng, William D., Deng Zhang & Kay L. O'Halloran. 2013. Advances and Frontiers in Multimodal Discourse Analysis. *Contemporary Linguistics* 1. 88–99.
- Forceville, Charles J. 2006. Non-Verbal and Multimodal Metaphor as a Cognitivist Framework: Agendas for Research. In Gitte Kristiansen, Michel Achard, Rene Dirven & Francisco Ruiz de Mendoza Ibanez (eds.), *Cognitive Linguistics: Current Applications and Future Perspectives*, 379–402. Berlin/New York: Mouton de Gruyter.
- Forceville, Charles J. 2010. Book Review: The Routledge Handbook of Multimodal Analysis. *Journal of Pragmatics* 42. 2604–2608.
- Forceville, Charles J. 2013. Metaphor and Symbol: Searching for One's Identity Is Looking For A Home in Animation Film. *Review of Cognitive Linguistics* 19(2). 250–268.
- Foulsham, Tom, Dean Wybrow & Neil Cohn. 2016. Reading Without Words: Eye Movements in the Comprehension of Comic Strips. *Applied Cognitive Psychology* 30(4). 566–579. <https://dx.doi.org/10.1002/acp.3229>.
- Gold, Matthew K. (ed.). 2012. *Debates in the Digital Humanities*. Minneapolis and London: University of Minnesota Press.
- Granström, Björn, David House & Inger Karlsson (eds.). 2002. *Multimodality in Language and Speech Systems*. Dordrecht: Kluwer Academic Publishers.
- Gu, Yueguo. 2006. Multimodal Text Analysis: A Corpus Linguistic Approach to Situated Discourse. *Text & Talk* 26(2). 127–167.
- Halliday, Michael A. K. 2003. *Collected Works of M. A. K. Halliday. Volume 3. On Language and Linguistics*. London and New York: Continuum.
- Hasson, Uri, Ohad Landesman, Barbara Knappmeyer, Ignacio Vallines, Nava Rubin & David J. Heeger. 2008. Neurocinematics: The Neuroscience of Film. *Projections* 2(1). 1–26.
- Hiippala, Tuomo. 2015. *The Structure of Multimodal Documents: An Empirical Approach*. London: Routledge.
- Holsanova, Jana. 2008. *Discourse, Vision, and Cognition* (Human Cognitive Processes 23). Amsterdam: John Benjamins.
- Holsanova, Jana, Henrik Rahm & Kenneth Holmqvist. 2006. Entry Points and Reading Paths on Newspaper Spreads: Comparing a Semiotic Analysis with Eye-Tracking Measurements. *Visual Communication* 5(1). 65–93.
- Iedema, Rick. 2001. Resemiotization. *Semiotica* 37(1/4). 23–40.
- Jewitt, Carey. 2009a. An Introduction to Multimodality. In Carey Jewitt (ed.), *The Routledge Handbook of Multimodality*, 14–29. London, New York: Routledge.
- Jewitt, Carey (ed.). 2009b. *The Routledge Handbook of Multimodal Analysis*. London: Routledge.
- Jewitt, Carey (ed.). 2014. *The Routledge Handbook of Multimodal Analysis. Second Edition*. London and New York: Routledge.
- Jewitt, Carey, Jeff Bezemer & Kay O'Halloran. 2016. *Introducing Multimodality*. London: Routledge.
- Jewitt, Carey & Gunther Kress. 2003. *Multimodal Literacy* (New Literacies and Digital Epistemologies 4). Frankfurt a.M. and New York: Peter Lang.
- Klemm, Michael, Daniel Perrin & Sascha Michel. 2016. Produktionsforschung. In Nina-Maria Klug & Hartmut Stöckl (eds.), *Handbuch Sprache im multimodalen Kontext*, 277–299. Berlin, New York: de Gruyter.

- Klug, Nina-Maria & Hartmut Stöckl (eds.). 2016. *Handbuch Sprache im multimodalen Kontext*. Berlin: de Gruyter.
- Kress, Gunther & Theo van Leeuwen. 2006 [1996]. *Reading Images: The Grammar of Visual Design*. London and New York: Routledge.
- Krishnan, Armin. 2009. *What Are Academic Disciplines? Some Observations on the Disciplinarity vs. Interdisciplinarity Debate*. University of Southampton National Centre for Research Methods. Working Paper.
- van Leeuwen, T. 2014. Towards a Semiotics of Listening. In E. Djonov & S. Zhao (eds.), *Multimodal Critical Studies of Popular Discourse*, 251–263. London and New York: Routledge.
- van Leeuwen, Theo. 2005. *Introducing Social Semiotics*. London: Routledge.
- van Leeuwen, Theo & Carey Jewitt (eds.). 2000. *Handbook of Visual Analysis*. London: SAGE Publications.
- Liu, Yu & Kay L. O'Halloran. 2009. Intersemiotic Texture: Analyzing Cohesive Devices Between Language and Images. *Social Semiotics* 19(4). 367–388.
- Machin, David. 2007. *Introduction to Multimodal Analysis*. London: Hodder Arnold.
- Maton, Karl. 2013. *Knowledge and Knowers: Towards a Realist Sociology of Education*. London and New York: Routledge.
- Mondada, Lorenza. 2016. Challenges of Multimodality: Language and the Body in Social Interaction. *Journal of Sociolinguistics* 20(2). 2–32.
- Muller, Johan. 2011. Through Others' Eyes: The Fate of Disciplines. In Frances Christie & Karl Maton (eds.), *Disciplinarity: Functional Linguistic and Disciplinary Perspectives*, 13–34. London: Continuum.
- Nakano, Yukiko & Matthias Rehm. 2009. Multimodal Corpus Analysis as a Method for Ensuring Cultural Usability of Embodied Conversational Agents. In Masaaki Kurosu (ed.), *Human Centered Design. International Conference on Human Centred Design*, 521–530. Berlin, Heidelberg: Springer.
- Newfield, Denise. 2014. Transformation, Transduction and the Transmodal Moment. In Carey Jewitt (ed.), *The Routledge Handbook of Multimodal Analysis. Second Edition*, 100–113. London: Routledge.
- Norris, Sigrid. 2004. *Analyzing Multimodal Interaction: A Methodological Framework*. London and New York: Routledge.
- Norris, Sigrid (ed.). 2015. *Multimodality. Critical Concepts in Linguistics*. London, New York: Routledge.
- O'Halloran, Kay L. (ed.). 2004. *Multimodal Discourse Analysis: Systemic Functional Perspectives*. London: Continuum.
- O'Halloran, Kay L. 2008. Systemic Functional-Multimodal Discourse Analysis (SF-MDA): Constructing Ideational Meaning using Language and Visual Imagery. *Visual Communication* 7(4). 443–475.
- O'Halloran, Kay L., Sabine Tan, Duc-Son Pham, John A. Bateman & Andrew Vande Moere. 2016. A Digital Mixed Methods Research Design: Integrating Multimodal Analysis with Data Mining and Information Visualization for Big Data Analytics. *Journal of Mixed Methods Research* 12(1). 11–30. <https://doi.org/10.1177/1558689816651015>.
- Özerdem, Mehmet Sıraç & Hasan Polat. 2017. Emotion Recognition Based on EEG Features in Movie Clips with Channel Selection. *Brain Informatics* 4(4). 241–252. <https://doi.org/10.1007/s40708-017-0069-3>.
- Pederson, Kaitlin & Neil Cohn. 2016. The Changing Pages of Comics: Page Layouts across Eight Decades of American Superhero Comics. *Studies in Comics* 7(1). 7–28.

- Piotrowski, Michael. 2018. Digital Humanities: An Explication. In Manuel Burghardt & Claudia Müller-Birn (eds.), *Inf-dh-2018*, Bonn: Gesellschaft für Informatik e.V.
- Priem, Karin & Geert Thyssen. 2013. Puppets on a String in a Theatre of Display? Interactions of Image, Text, Material, Space and Motion in The Family of Man (ca. 1950–1960s). *Paedagogica Historica* 49(6). 828–845. <https://doi.org/10.1080/00309230.2013.846925>.
- Reinhard, CarrieLynn D. & Christopher J. Olson (eds.). 2016. *Making Sense of Cinema: Empirical Studies into Film Spectators and Spectatorship*. New York and London: Bloomsbury.
- Rigaud, Christophe & Jean-Christophe Burie. 2018. Computer Vision Applied to Comic Book Images. In Alexander Dunst, Jochen Laubrock & Janina Wildfeuer (eds.), *Empirical Comics Research. Digital, Multimodal, and Cognitive Methods*, 104–124. London and New York: Routledge.
- Royce, Terry D. 1999. *Visual-Verbal Intersemiotic Complementarity in The Economist Magazine*. United Kingdom: University of Reading dissertation.
- Royce, Terry D. 2007. Intersemiotic Complementarity: A Framework for Multimodal Discourse Analysis. In Terry D. Royce & Wendy L. Bowcher (eds.), *New Directions in the Analysis of Multimodal Discourse*, 63–110. Mahwah, NJ: Lawrence Erlbaum Associates.
- Sahle, Patrick. 2015. Digital Humanities? Gibt's doch gar nicht! *Grenzen und Möglichkeiten der Digital Humanities. Sonderband 1 der Zeitschrift für digitale Geisteswissenschaften* https://doi.org/10.17175/sb001_004.
- Seizov, Ognyan & Janina Wildfeuer (eds.). 2017. *New Studies in Multimodality: Conceptual and Methodological Elaborations*. London and New York: Bloomsbury.
- Smith, Tim J. 2012. The Attentional Theory of Cinematic Continuity. *Projections: The Journal for Movies and the Mind* 6(1). 1–27.
- Stöckl, Hartmut. 1997. *Textstil und Semiotik englischsprachiger Anzeigenwerbung*. Frankfurt am Main: Peter Lang.
- Stöckl, Hartmut. 2004. In Between Modes: Language and Image in Printed Media. In Eija Ventola, Cassily Charles & Martin Kaltenbacher (eds.), *Perspectives on Multimodality*, 9–30. Amsterdam: John Benjamins.
- Stöckl, Hartmut. 2016. Multimodalität – Semiotische und textlinguistische Grundlagen. In Nina-Maria Klug & Hartmut Stöckl (eds.), *Handbuch Sprache im multimodalen Kontext* (Handbooks of Linguistics and Communication Science (HSK) 7), 3–35. Berlin: de Gruyter Mouton.
- Streeck, Jürgen. 2013. Interaction and the Living Body. *Journal of Pragmatics* 46(1). 69–90.
- Tan, Sabine, Kay L. O'Halloran, Peter Wignell, Kevin Chai & Rebecca Lange. 2018. A Multimodal Mixed Methods Approach for Examining Recontextualisation Patterns of Violent Extremist Images in Online Media. *Discourse, Context and Media* 21. 18–35.
- Terras, Melissa, Julianne Nyhan & Edward Vanhoutte (eds.). 2013. *Defining Digital Humanities: A Reader*. Farnham and Burlington: Ashgate.
- Tseng, Chiao-I, Jochen Laubrock & Jana Pflaeging. 2018. Character Developments in Comics and Graphic Novels: A Systematic Analytical Scheme. In Jochen Laubrock Alexander Dunst & Janina Wildfeuer (eds.), *Empirical Comics Research: Digital, Multimodal, and Cognitive Methods*, 154–175. London: Routledge.
- Unsworth, Len. 2007. Image/Text Relations and Intersemiosis: Towards Multimodal Text Description for Multiliteracies Education. In Leila Barbara & Tony Berber Sardinha (eds.), *Proceedings of the 33rd International Systemic Functional Congress (33rd ISFC)*, 1165–1205. Sao Paulo, Brazil: Pontifícia Universidade Católica De Sao Paulo (PUCSP).

- van Leeuwen, Theo. 2005. Three Models of Interdisciplinarity. In Ruth Wodak & Paul Chilton (eds.), *A New Agenda in (Critical) Discourse Analysis: Theory, Methodology and Interdisciplinarity*, 3–18. Amsterdam: John Benjamins.
- Waller, Robert. 2012. Graphic Literacies for a Digital Age: The Survival of Layout. *The Information Society: An International Journal* 28(4). 236–252.
- Wildfeuer, Janina. 2015. Bridging the Gap between Here and There: Combining Multimodal Analysis from International Perspectives. In J. Wildfeuer (ed.), *Building Bridges for Multimodal Research. International Perspectives on Theories and Practices of Multimodal Analysis*, 14–33. Frankfurt am Main: Peter Lang.
- Wildfeuer, Janina & Ognyan Seizov. 2017. Afterword: Towards a New Discipline of Multimodality. In Ognyan Seizov & Janina Wildfeuer (eds.), *New Studies in Multimodality. Conceptual and Methodological Elaborations*, 277–284. London: Bloomsbury.
- Yarbus, Alfred Lukyanovich. 1967. *Eye Movements and Vision*. New York, NY: Plenum Press.
- Zhao, G., Y. Ge, B. Shen, X. Wei & H. Wang. 2018. Emotion Analysis for Personality Inference from EEG Signals. *IEEE Transactions on Affective Computing* 9(03). 362–371. <https://doi.org/10.1109/TAFFC.2017.2786207>.



Part II: Disciplinary Thoughts

Hartmut Stöckl

Linguistic Multimodality – Multimodal Linguistics: A State-of-the-Art Sketch

Abstract: Distilling relevant recent publications, the present chapter seeks to sketch out the current state of the art in the burgeoning field of (linguistic) multimodality research. It does so by identifying, explaining, and inter-linking concepts that have come to constitute a consensual core in the area. The sketch also points out and discusses contested notions highlighting the nature of the differences in opinion. Wherever possible and useful, the explanations will be based on a trans-textual example of multimodal discourse involving image and text in two different genres. Finally, attention will be directed to three promising perspectives: multimodal rhetoric, cognitive semiotics, and trans-textuality.

Keywords: multimodality, multimodal linguistics, rhetoric, semiotics, trans-textuality

1 Introduction

Since its development on the basis of functional grammar in the 1980s (Halliday & Hasan, 1985; Halliday, 1978), social-semiotic multimodality research has taken a prolific and dynamic development. It is essentially based on two plausible arguments: the semiotic dictum that communication relies on a whole host of different signing modes and their combination, and the linguistic concerns evident since the advent of pragmatics and text linguistics with a gradual extension of context. The first idea drove the description of individual signing modes in structural-semiotic, functional-grammatical, and pragmatic terms (e.g., image, music, etc.). The second idea led to the realization that one mode is contextualized by another, so that modes mutually provide co(n)text and can often only function properly through this co-contextualization. Given these two central tenets, it is only natural that attention gradually focused on the genre-specific patterns of mode-linking and the ways in which multimodal ensembles, artifacts, or performances come to be constituted, structured, and used in all kinds of discourse domains and media—now the foremost concerns of linguistic multimodality studies.

Even though its inter-relations to other more traditional fields are numerous, most notably to linguistics and media/communication studies, multimodality research exhibits all the hallmarks of an independent discipline: a set of core concepts forming a more or less unified base-theory, a broad and widening area

<https://doi.org/10.1515/9783110608694-002>

for empirical study, suitable analytical methods of analysis and a lively exchange of ideas dynamically driving a re-development of theory and empirical tools. All this is attested by a large number of text-books, introductions, handbooks, and (edited) volumes on special subjects (cf., e.g., Ledin & Machin, 2018; Bateman et al., 2017; Tseronis & Pollaroli, 2018; Tseronis & Forceville, 2017; Jewitt et al., 2016; Klug & Stöckl, 2016; Bezemer & Kress, 2016; Jewitt, 2014b, 2009b; Norris & Maier, 2014; Bateman, 2014b, 2008; Kress, 2010; van Leeuwen, 2005, 1999; Kress & van Leeuwen, 2001; Kress & van Leeuwen, 2006 [1996]; Kress et al., 2000). Being young, however, it is only natural that large parts of multimodal theory are still solidifying and that ‘rivaling’ terms are in use and opinion divided over both basic and more subtle differences in interpretation. What could still be called a varied landscape of theory is also promoted by the existence of separate subdisciplines, such as multimodal interaction analysis, geo-semiotics, or systemic-functional multimodal discourse analysis, for example. Some of the variation is owing to differing objects of analysis and methodological approaches, another part is perhaps the result of an insufficiently integrated theory of multimodality. The rapid diversification of multimodality studies seems to also somewhat facilitate a lack of mindfulness towards the theoretical core and a tendency to re-invent (and thus perhaps skew) what is already known, labelled, and defined.

In how far the implied unification or reinforcement of multimodal theory is a pressing need is not for me to judge, but I am suggesting here that it may be useful to look back on the theory-building accomplished, raise awareness for the multimodal foundations, and reflect on some of the promising ways to move on. The present chapter seeks to do just this; it is intended as a bird’s-eye-view of the multimodal theoretical terrain, recording major landmarks, their position relative to one another, the connecting lines between them, and zooming in on some of them. Section 2 highlights four districts of the multimodal terrain, each consisting of a network of related key concepts. Wherever useful, the explanations will refer to multimodal discourse fragments, i.e., two different genre exemplars—an iconic painting as discussed on *Google Arts and Culture* (see Figure 1) and a *Volkswagen* advertisement using this and yet another iconic painting (see Figure 2). The key concepts will be explained and related to one another; unresolved issues will be pointed out—as much as this is possible within the limited scope. Section 3 is a summary outlining four simple hypotheses that also involve ways to move on in the field. The final section also spotlights the rhetorical, the cognitive-semiotic, and the trans-textual as three promising perspectives or foci in future work.



The birth of Venus

Sandro Botticelli 1483 - 1485



Aus der Sammlung von
Uffizi Gallery

The painting was commissioned by Lorenzo di Pierfrancesco de' Medici, a cousin of Lorenzo the Magnificent. The theme was probably suggested by the humanist Poliziano. It depicts Venus born from the sea foam, blown by the west wind, Zephyrus, and the nymph, Chloris, towards one of the Horae, who prepares to dress her with a flowered mantle.

This universal icon of Western painting was probably painted around 1484 for the villa of Castello owned by Lorenzo di Pierfrancesco de' Medici. Giorgio Vasari saw the work there in the mid-sixteenth century – along with Botticelli's other well-known *Primavera* – and described it precisely as "showing the Birth of Venus." The old idea that the two Botticelli masterpieces were created for the same occasion, in spite of their substantial technical and stylistic diversity, is no longer accepted. However, rather than a birth, what we see is the goddess landing on the shore of her homeland, the island of Cyprus, or on Kithera. The theme, which can be traced back to Homer and to Ovid's *Metamorphoses*, was also celebrated by the great humanist Agnolo Poliziano in the poetic verses of his *Stanze*. The Venus of the Uffizi is of the "Venus pudica" type, whose right breast is covered by her right hand and billowing long blond hair partially shrouds her body. The goddess stands upright on a shell as she is driven towards the shore by the breeze of Zephyrus, a wind god, who is holding the nymph, Chloris. On the right is the Hora of springtime, who waits to greet Venus ashore with a cloak covered in pink flowers.

The seascape, stunning for its metaphysical tone and almost unreal quality, is illuminated by a very soft, delicate light. Like Botticelli's other masterpiece,

Pallas and the Centaur, the *Birth of Venus* is painted on canvas – fairly unusual for its time – using a technique of thin tempera, based on the use of diluted egg yolk, which lends itself particularly well to give the painting that aspect of extraordinary transparency, which brings to mind the pictorial quality of a fresco. The figure recalls classical sculpture and is very similar to the famous Medici Venus found in the Uffizi, which the artist certainly knew. The real meaning of this dreamlike vision is still under scholarly debate and investigation but is undoubtedly linked with the Neo-Platonic philosophy, widely cultivated in the Medici court.

Like the *Primavera*, the *Birth of Venus* is also associated with the concept of *Humanitas*, or virtuous Humanity, a theory developed by Marsilio Ficino in a letter to the young Lorenzo. According to the interpretation by Ernst Gombrich, the work depicts the symbolic fusion of Spirit and Matter, the harmonious interaction of Idea and Nature. Nevertheless, the interpretations of this painting of extraordinary visual impact are numerous and diverse. The divine ethereal figure has been viewed as an allegorical representation of *Humanitas* upon her arrival to Florence, while the nymph holding out the cloak of flowers for the goddess may perhaps be identified as Flora, the same depicted in this masterpiece's "twin", the *Primavera*, where she may be seen instead as the personification of the city of Florence. From this work emerges clear evidence of Botticelli's strive to reach perfection of form that could rival with classical antiquity. It is for this reason that the humanist Ugolino Verino in his work *Epigrammata*, presented in 1485 to the King of Hungary, Matthias Corvinus, likened the Florentine painter to the legendary Apelles of Ancient Greece.

Details

Title: The birth of Venus

Arteller: Sandro Botticelli

Datierung: 1483 - 1485

Style: Renaissance

Provenance: Lorenzo di Pierfrancesco de' Medici family collections

Original Title: La nascita di Venere

Abmessungen: w2785 x h1725 mm

Typ: painting

Material: Tempera on panel



Sandro Botticelli

Florentine painting

Italienische Renaissance

Renaissance

Tempera

Fig. 1: Sandro Botticelli's *The Birth of Venus* on the *Google Arts and Culture* pages. Source: <https://artsandculture.google.com/asset/the-birth-of-venus/MQEeq50LABEBVg> (last accessed: 29 August 2019). Public Domain.

2 A Networking Account of Key Concepts

This section is based on a collection of key concepts from some of the most central and recent publications in multimodality (see Table 1). Rather than treat each of them separately, they will be grouped into four thematic sets (see Sections 2.1 to 2.4) in order to link them into internally coherent building blocks of multimodal theory and show the connections between them. The focus is thus not primarily on defining terms but on inter-relating them, weighing different interpretations, and discussing their merits. The treatment of multimodal theory proceeds from micro to macro: it starts with modes and materials/media (Section 2.1, then takes up meaning multiplication, i.e., the very concept of multimodality (Section 2.2), goes on to address mode linking (Section 2.3), and finally turns to multimodal genre and discourse (Section 2.4). My concern cannot be with detail nor with completeness here; instead my interest is in a generalized overview that affords orientation and enables judgement.

Tab. 1: Index of key terms in linguistic multimodality research

action	modal	resource
– communicative	... complexity	– semiotic
– social	... intensity	... integration
affordance	... logic	rhetorical
– modal	... reach	... process
canvas	mode	... situation
design	– centrality of	... strategy
dialogicity	... combination	... structure
... of signs	... family	semiotic(s)
discourse	... linking	... artifact
... semantics	– semiotic	– cognitive
genre	– signing	... material(s)
– multimodal	multimodal	... mode
indexicality	... coherence	... resource
intermodal harmony	... discourse	– social
intersemiotic	... ensemble	– structural
complementarity	... text	... work
relations	multimodality	status
materiality	mutual elaboration	... relations
meaning multiplication	relations	... relative
medium (mediality)	– conjunctive	transcription
medial variant	– logicosemantic	transtextual
	– text-image	... ity
	resemiotization	

The samples discussed below prejudice the two central signing modes language and image (or text-image relations) but allow insight into the workings of different media, genres, and discourses. They also aptly demonstrate how semiotic artifacts are strategically resemiotized in the rhetorical design of specific multimodal discourse domains: Botticelli’s *Birth of Venus* (1486, see Figure 1) first pictorially records a mythical story, is then explained and interpreted in an art-historical, didactic context provided by the *Google Arts and Culture* website, and finally—along with Edvard Munch’s *Cry* (1919)—finds its way into an advertisement carrying the central argument (see Figure 2). The evolving trans-textual multimodal discourse trajectory thus leads from illustrating mythology over art explanation to commercial persuasion, each stage in the cycle entailing different and distinct multimodal genres.



Fig. 2: “Just because it’s classic doesn’t mean it fits.” Volkswagen 2015, Grabarz & Partner, Hamburg. Lürzer’s Archive 4/2015, automotive 4.1545. Personal photo.

2.1 Shaping Semiotic Modes from Medial Materials

Undoubtedly, *semiotic/signing mode* (see Figure 3) is the most central concept in multimodality, at the same time, however it is perhaps hardest to pin down

satisfactorily. Traditionally, the organizing principles and *resources* of modes have been “understood as an outcome of the cultural shaping of a material” (Jewitt, 2014b, 464) so that Kress (2010, 155) simply refers to mode as “the material stuff of signs”. This signals a potentially risky conflation of modes and media (see below). A relatively uncontroversial notion of mode would therefore be to say it is “a set of resources, shaped over time by socially and culturally organized communities, for making meaning” (Jewitt et al., 2016, 15). A glossary definition in Norris & Maier (2014, 391) adds to this the orderly nature ascribed to codes by calling a mode a “semiotic system with rules and regularities attached to them as they are in use by social actors”.

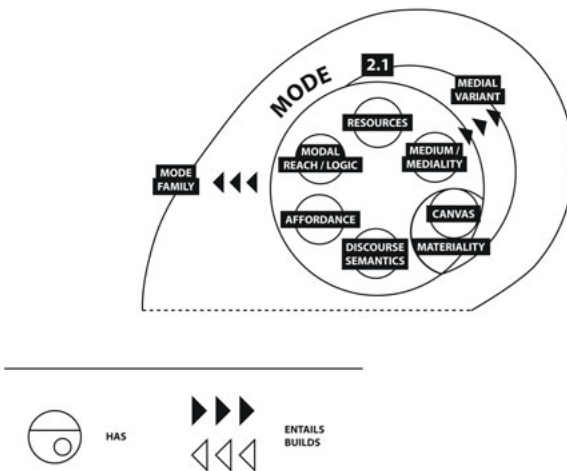


Fig. 3: A graphic model of a mode’s conceptual elements (Visualization: Jana Pflaeging & Hartmut Stöckl).

Writing and painting are clear examples of modes, the first with a very wide currency, the latter the domain of especially gifted and trained experts (artists). Written language has shaped the graphic material into sets of characters that can represent the spoken word and strings them together in clauses and paragraphs. But it also uses punctuation, spacing, font-styles, layout, and color to realize syntactic and textual units in the space of a graphic surface. Painting is similarly dependent on making visual marks on a suitable material but wields the brush to produce lines, shapes, and textures that ultimately make up recognizable gestalts and arrange these into configurations of objects, people, scenes, events, and actions. This ex-

emplification makes clear that it is plausible and preferable to think of modes as having or making available sets of *semiotic resources*, rather than equate mode with resource. Exactly which resources a mode has and how they work is a prime subject of semiotic or (multi-)modal analysis. As a result, it is possible to gauge and compare the differing *modal affordances*, i.e., the kinds of semiotic work the various modes enable (cf. Kress, 2010, 62–63) or “what is possible to express and represent easily with a mode” (Jewitt, 2014b, 456). Applied to the samples, painting affords showing (even fictitious) entities in full physical detail and visualizing a mythical narrative; writing affords telling, interpreting that story and explaining an image as well as explicitly construing a rational argument.

Further work on the conceptualization of mode has introduced various strata, i.e., levels on which modes need to function and on which they can be described. A potential, maximum stratification is suggested by Bateman (2016, 39), who—referring to Klug & Stöckl (2015, 243–247)—lists: *materiality/mediality*, codality, sensory modalities, processing mechanisms, and socio-cultural conventions. Narrowing this down, Stöckl (2016, 6–9) proposes to characterize any given mode by determining its medial realization, its perceptual channel, and its structural regularities of coding meaning. In this light, painting is a special *medial variant* (Stöckl, 2014, 11–16) of the image, characterized by a manual image-making using canvas, brush, and different kinds of paint/colors, creating a unique work of art that is potentially embedded in museum and collection contexts. Perceptually, paintings are visual media, but quite unlike writing, they function holistically and invite the viewer to explore all its material richness and semantic density (cf. Goodman, 1969). Along with speech, writing belongs in the *mode family* (Bateman et al., 2017, 122–123) of verbal language sharing with it most code characteristics, i.e., the semiotic resources of lexicogrammar, but differing from it in its graphic-spatial nature and linear visual perception. *Medial variant* and *mode family* are useful terms to do justice to the fact that modes may be related and close in terms of their *semiotic affordances* but different in their *material/medial affordances*. With regard to the classic example of speech vs. writing, Kress (2010, 104) says that “at a lower level of generality they do differ and at a ‘higher’ level they don’t”.

Bateman (2011, 2016) and Bateman et al. (2017) suggest a similar stratification of semiotic modes that differs in one important respect. This model suggests a material substrate comprising both medial and perceptual dimensions, a “mid-level” (Bateman, 2016, 46–47) providing the lexis and grammar components of a mode’s internal structure, and “finally, ‘above’, or ‘surrounding’ these levels of semiotic abstraction” a “more abstract stratum of (local) *discourse semantics*” (Bateman, 2016, 46–47). The neat suggestion in this model is that image and language, for example, already come equipped with “the mechanisms that govern the interpretation of semiotic modes in their context of occurrence” (Norris & Maier,

2014, 387). Taking this at face value, this would mean that the Botticelli painting (on the *Google Arts and Culture* website) has a “contextual interpretation of the semiotic resources” (Hiippala, 2014, 115) inscribed in it, namely discourse semantics, which “directs the reader towards the correct interpretation in a given context” (Hiippala, 2014, 115). Two questions might be raised about the elegant idea of modes containing their discourse semantics: first, is not the semantics of discourse something that would need to be placed at the level of text or semiotic artifact (or better even genre) and concerns both cohesive structures in it and assumed cognitive processes in the recipient? And, secondly, in the light of multimodality, is it perhaps not more adequate to say that discourse semantics operates across and between modes rather than separately within single modes? Applied to the samples, we would need to ask: Can a recipient arrive at an adequate, contextually appropriate interpretation of *The Birth of Venus* or *The Cry* (and their fusion in the advertisement) solely on the grounds of a mode-knowledge of painting? And, is not ultimately the continuous transduction of meaning from one mode to another (illustrating a myth, ekphrasis of painting, multimodal argumentation) actually generating the adequate contextual interpretation?

Differing views of mode also seem to determine the stances taken on the boundedness and distinctiveness of modes. Two ‘camps’ are discernible here: first, a heuristic position which argues that “a mode has no clear boundaries” (Norris, 2004, 11) and that modes are “construed by social, cultural, and historical factors” (Jewitt, 2009a, 22), and, second, a categorical position which advocates the possibility of delineating and characterizing different modes (Stöckl, 2016, 9–19). Essentially, a view of mode that emphasizes the creative shaping of materials will shy away from drawing up a definite list of modes, as they can be remade and new ones emerge. For example, shall we call the navigational and interactional resources of hypertextual and social media (as exhibited on the *Google Arts and Culture* pages) a mode? By contrast, a concept of mode that teases out its various organizational strata will facilitate mode recognition and description. Interestingly, these are not, as might perhaps seem, irreconcilable conflicts of opinion: You may very well claim that “just what semiotic modes there are is always a socio-historical question” (Bateman, 2014b, 18) and still be interested in identifying different modes and their typical structural and functional organization. Distinguishing between core or “full-blown modes” (Hiippala, 2014, 121) and peripheral modes or sub-modes (Stöckl, 2004, 11–18) seems a helpful way of coming to grips with the different nature, status, and currency of modes.

Whichever way we look at modes—whether as semio-culturally shaped materials, as stratified abstract entities with material, ‘grammatical’, and discourse-semantic properties or as ‘interpretative practices’ (Bateman, 2016, 39)—it is important not to conflate them with the notion of *medium* or ‘*semiotic materials*’

(Ledin & Machin, 2018, 3). (In fact, Ledin/Machin also wrongly equate *semiotic materials* with *semiotic artifacts*). Kress & van Leeuwen (2001, 21–22) have already warned against this and cautioned a view that keeps modes and media separate but related in the way that modes are realized in a medium and media as tools and materials facilitate and constrain modes. It is not surprising that given the increasing significance of diversifying materials in current communication design and given the traditionally contested nature of the concept medium/media that recent work in multimodality puts medial/material questions center-stage.

Bateman et al. (2017) make a fresh attempt in this regard by distinguishing *medium* and *canvas*. While there appears to be considerable conceptual overlap between the two and the term canvas may inadvertently suggest only certain materials, the main aim of the terminological split seems to be to single out canvas in order to arrive at a way of classifying “‘abstract’ or ‘generalized’ materialities that are “multimodality ready” (Bateman et al., 2017, 103). So rather than consider the whole host of features usually factored into a medium (such as the technological tools or the social/institutional context and the routine actions, for example) canvazooms in on the “material regularities [...] not just the material itself” (Bateman et al., 2017, 87). According to this logic, media have canvases and sub-canvases; they help to “break down larger complex communicative situations into smaller, component parts” (Bateman et al., 2017, 101) and are generally defined as “the material employed for meaning making” (Bateman, 2016, 40) or “all possible bearers of meaningful regularities” (Bateman et al., 2017, 86–87). The classification scheme suggested for canvases (Bateman et al., 2017, 103–110) is plausible and convincing; its dimensions take up many of the linguistic mainstream typologies for media or communicative situations (cf., e.g., Page et al., 2014, 16–28). Canvas classification may well “offer some fairly radical ways of recasting a host of existing problems” (Bateman et al., 2017, 198). It is to be feared, however, that canvas will have a hard time in current research contexts, as the concept overlaps with medium in many ways: i.e., as “material substrate to carry the semiotic resources” (Hiippala, 2014, 116–117), and as general “historically stabilized site for the deployment and distribution of some selection of semiotic modes for the achievement of varied communicative purposes” (Bateman et al., 2017, 123).

In the two genres of my samples, the medial distinctions that become meaningful are manual, artistic painting vs. mechanic photography vs. digitally manipulated images and a static, permanent surface vs. an immutable ergodic (Bateman et al., 2017, 108) hypertextual screen-resource. More importantly, and for clearly stated reasons not covered in the notion of canvas, the medial difference at work here can be seen as one of the interpretative practices of paintings or art-didactic materials vs. those of brand and product promotion. It is in this sense that Kress (2010, 114) argues that the concepts of mode—and this entails attention to its ma-

terial substrate—, medium/canvas, genre, and discourse are on the same level. This means they are inextricably bound up in practice and must, therefore, be thought of in a unity. Abstractions and conceptual definitions draw meaningful boundaries here, but there cannot be an easy consensus over how far we want to stretch a concept's content or how narrowly we feel we need to restrict it. It is more important in my view to see the connectedness of various analytical levels rather than their singular constitution.

2.2 From Modal Reach to Multimodality

The essential nature of multimodality (see Figure 4) resides in the textual combination of different modes and their integration in terms of structure, discourse semantics, and rhetorical function within contexts of social (inter-)action. Jewitt et al. (2016, 2–3) talk about “integrated, multimodal whole(s)” and emphasize that each mode and its resources offer “distinct potentials and limitations”. Such mode differences have been captured by the idea of *modal reach* (Kress, 2010, 83) and *modal logic* (Jewitt, 2014b, 464), which in my view are different terms for the same concept. What is ultimately “the communicative and representational potential” (Jewitt, 2014b, 464) of a mode is said to either describe “what can be *done* with a semiotic mode” (Bateman et al., 2017, 119) or what ‘areas’ or ‘terrain’ the mode “*covers*” (Kress, 2010, 11 & 83). So, whether the focus is on communicative action or on semantic representation, different modal reaches and logics derive from specific material/medial and code-structural properties (Bateman et al., 2017, 119) and are the result of socio-culturally and historically determined conventions.

The samples clearly demonstrate that language/writing and image (painting/photography) have different modal reaches and logics in many respects. In my view, such differences can best be explained by the semiotic resources that modes do or do not make available. Language can aptly name the artist, fix the mythical references but also describe the technique and style of the painting and interpret its impact in the arts and humanities. This reach is based on a fine-grained and fixed lexical repertoire and on speech act indicators. Through composition images afford foregrounding (actor and their symbolic attributes) and backgrounding (context) as well as salience and narrative action based on the positioning of image elements. It is another interpretation of semiotic reach to argue that what was originally a taboo for pictorial representation (i.e., mythical subjects) began to be painted and that what was usually described and explained through language (product claim/argument) is now realized in artful image-making. Here one could talk of shifting socio-cultural conventions regulating the use of modes.

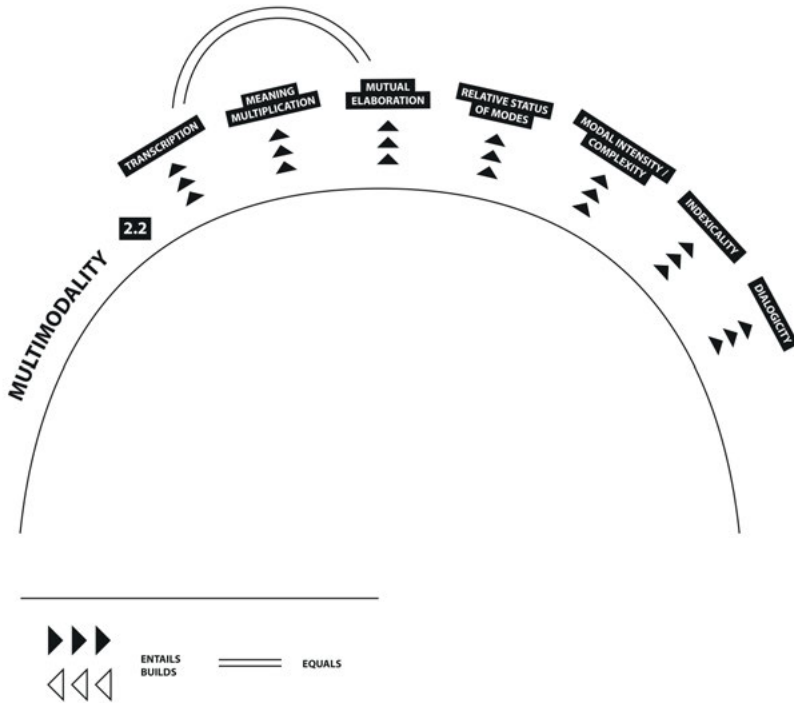


Fig. 4: A graphic model of multimodality (Visualization: Jana Pflaeging & Hartmut Stöckl).

One of the core questions in multimodality is to explain exactly what happens when various modes with different reaches combine. That they combine in the first place can conveniently be seen to be a direct result of the insufficient reaches or limitations of the individual modes, which can mutually be compensated for in mode-integration. If we elevate this kind of logic to a universal principle, no mode is sufficient onto itself and relies for adequate meaning-making on the elaboration or enrichment by other modes. The concept of *transcription* by Jäger (2002) has advanced this hypothesis of an essentially multimodal overall cultural semantics. This seems a fitting explanation of the relation between the painting and its underlying mythical story; it seems less obvious in the case of the advertisement, which could have spelt out the argument using just language.

In any case, mode-combinations are seen to result in the *multiplication of meaning* or the *mutual elaboration* of the modes. Mutual elaboration reinforces Jäger’s idea (see above) that “sets of semiotic resources [...] are partial and incomplete” (Jewitt et al., 2016, 158–159) and are consequently in need of joining other modes

and resources to be fully elaborated. Meaning multiplication was suggested by Lemke (1998) as a principle of mode integration that claims that the informational impact of the individual modes is not simply added one to another but is ‘more’ or different than the sum of the modal parts (cf. Bateman, 2014b, 6). This also means that combining modes is a “dynamic process of text construction and reception”, not a “static or pre-given inter-relationship across codes or semiotic resources” (Bateman, 2016, 54–55). Much of the work in multimodality studies seeks to find out about “the nature of this ‘more’” (Bateman et al., 2017, 16) implied by the idea of multiplication; this work leads straight to questions of multimodal cohesion and coherence and to multimodal genre (see below). It is interesting to ask in this connection how much of the necessary multimodal construal can be explained ‘from below’ as resulting from the structure of the available modes and media, and how much ‘from above’ as guided by the discourse semantics and textual structure of the overall text/artifact. It is my conviction that genre knowledge plays a crucial part in construing multimodal relations and establishing logical/argumentative ties between the modes: we know, for example, that art-didactic text conventionally delivers information about the painter, the art-period, the making and interpretation of the painting etc. in order to contextualize the piece of art. Our genre knowledge of advertisements contains the important multimodal fact that iconic paintings in advertising copy are not interpreted as such but used as *topoi* in an argument. (Multimodal) discourse semantics, therefore, would seem to me a crucial part of multimodal genreknowledge and of our awareness of logical relations, in addition to being essentially a component of a mode’s internal structure, as Bateman et al. (2017, 113–121) argue (see Section 2.1).

We can also borrow an idea from multimodal interaction analysis to account for the different relative importance a mode can have in a multimodal text, something traditionally called *status* in text-image relations (Bateman, 2014b, 191–195). What is at issue here is the centrality of a mode to the particular social and communicative action at hand. While the multimodal structures of both samples explicitly tie image and language together by demonstrative reference and lexical cohesion, the mix of paintings in the ad is less central or modally intense, as the commercial argument would also work without the image. Norris (2004, 90) calls “the intensity or weight a mode carries in the construction of a higher-level action” *modal intensity*. This intensity seems stronger in the art-didactic text, where any explanation, interpretation, or contextualization as the overall communicative or social actions of the genre would fail without the image. Multimodal texts may also be of different *modal complexity*, a concept similarly originating in the work of Norris (2004), which “refers to the interrelationship of modes in a particular social action” (Pirini, 2014, 83) or in a multimodal text/genre. The Arts and Culture page about *The Birth of Venus* has a higher modal complexity than the advertisement

in that it also involves navigation and interactive (ergodic) features that link the website info to art-didactic films. Film generally has a higher modal complexity than the magazine or newspaper.

Finally, it may be useful to consider an idea from geo-semiotics, a discipline that looks at “how the sociocultural and meaning-making resources of space and place shape communication” (Jewitt et al., 2016, 111). The *indexicality* of the painting (i.e., its “meanings derived from the placement of the signs in their context”) shifts when it is transposed from the original museum context to an art-didactic website and, finally, when it is mixed and utilized as an argument in a commercial context. More importantly, what geo-semioticians call *dialogicity* of signs, namely an “inter-semiotic, and inter-discursive relationship between signs, once they are placed in an environment” (Jewitt et al., 2016, 111) is also revealing in the example. Both paintings merged and placed in the ad carry with them parts of their original interpretations, which in the structure of the commercial argument are drastically reduced to notions of the classic iconic painting, and to ideas of beauty and fear. There is also, in this advertisement, a remarkable dialogicity between language and image, as the first realizes the discourse of spare parts for classic cars, whereas the latter triggers bits of discursive knowledge about the two paintings involved.

2.3 Mode Linking and Multimodal Coherence

The core idea of multimodality posits the linking of semiotic modes and their formal, semantic and functional integration (see Figure 5). *Mode linking* (Kress, 2010, 119) or *mode combinations* (Bateman et al., 2017; Lyons, 2018), therefore, are the essence of multimodal artifacts/ensembles/communication and lead to *intersemiotic relations*, i.e., an “interplay between modes” (Jewitt, 2014a, 27), where “each mode is (...) partial in relation to the whole of the meaning”. Exactly what entities are linked is not so clear: ‘lexical’ items of the various modes, e.g., words and visual image elements, or the information and its semantic, rhetorical and discursive impact; van Leeuwen (2005, 219–220), for instance explicitly talks about “information linking”. How the linking happens must be the focus of multimodal study and has produced a whole number of different accounts of *multimodal cohesion and coherence* (Bateman, 2014a, 151–164). Kress (2010, 119), who says “much of semiosis is about linking of various kinds: (...) by and through actions, by adjacency and proximity, temporal or spatial” argues that “modes differ in how their affordances for realizing such relations have been developed”.

There is a general consensus to assume that the structures resulting from mode linking or combining create *intermodal harmony* (Norris & Maier, 2014, 390) or *intersemiotic complementarity* (Royce, 1998). Norris & Maier (2014, 390) equate

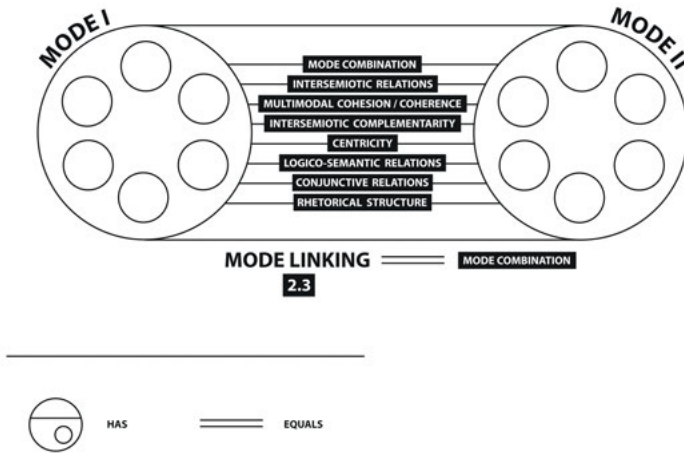


Fig. 5: A graphic model of mode linking (Visualization: Jana Pflaeging & Hartmut Stöckl).

intermodal harmony with *multimodal coherence*, “which derives from complementary connections in terms of form, meaning or function.” Exactly “how diverse combinations of semiotic modes can work together to form coherent communicative artifacts” (Bateman, 2014a, 171) is the subject of different analytical frameworks and theoretical approaches. Two very general issues are debated controversially in this respect: First, do we really assume there is something like mode equality as suggested by such terms as harmony or complementarity, in the sense that each modal contribution is equally significant? Or do we take differing modal reaches (see above) and differing structural arrangements of modes (through layout or sequence, for example) to affect different relative statuses or semantic and rhetorical weights of individual modes in any one multimodal artifact?—a notion very much favored in recent functional-grammatical and rhetorical models of multimodal coherence and expressed in such terms as *nuclearity* or *centricity* (cf. Caple, 2013; Stöckl et al., 2020). Second, do we take cohesive and coherent ties to be something materialized, explicitly expressed in the structure of a multimodal artifact, or can such coherence patterns also be created in the minds of the recipient on the basis of “knowledge, perception and commonsense reasoning” (Bateman, 2014a, 164)? There is much evidence from multimodal genre analysis to assume that “the material being interpreted must make a significant contribution” (Bateman, 2014a, 164) to multimodal meaning construal.

Bateman (2014a, 151–163) reviews four approaches to multimodal coherence, which I will briefly exemplify using Botticelli’s *Birth of Venus* on the *Google Arts*

and Culture pages (see Figure 1). Afterwards, I shall also introduce one of my own, simpler rationales for analysis (Stöckl, 2015, 2016; Stöckl, 2016), which integrates various approaches and generally looks at multimodal coherence as the recipient's cognitive engagement and semiotic work with combinations of semiotic modes according to an underlying hermeneutic. First, Bateman (2014a, 152) highlights the analysis of *multimodal cohesion* as an important way to “make more explicit just what kinds of semantic consequences follow from relating particular text-image (and other) elements”. In the art-didactic multimodal website, the text underneath the painting picks out and elaborates on some of its main visual elements: the painting's theme, its main actors or image participants (*Venus, Zephyr, Chloris* and one of the *Horai*), its materiality (technical/stylistic details, light, colors, effects), and the painting's provenance and contexts of interpretation. Multiple multimodal cohesive ties are thus established between a lexical element (e.g., *Venus/the goddess, this universal icon of Western painting*) and either the whole image or parts of it. What are called details on the webpage (e.g., *original title, measurements, material*, etc.) and what are identifiable as tags/links (e.g., *Sandro Botticelli, Uffizi Gallery, Florentine painting, renaissance, tempera*) serve as semantic categories to characterize the painting and add redundancy to the web of multimodal cohesive ties. This kind of analysis uncovers genre-typical, ekphrastic cohesive ties between text and painting, showing that the webpage affords both intense visual perusal of the painting (every possible detail may be magnified and studied) and its interpretative contextualization. What I have labelled ‘grammatical cohesion’ elsewhere (Stöckl, 2015, 67–68) and what could also be called multimodal deixis is realized by demonstrative and personal pronoun reference (e.g., *this, it*), which has lexical items point to the painting and its elements.

A second approach to multimodal coherence has been characterized as “modelled on grammar” (Bateman, 2014a, 154–157) because it treats units of text and image as *clause-relations* similar to how this is done in systemic-functional grammar (Martinec & Salway, 2005). Essentially two gains are made from this *clause-combinational approach*: first, modes are given unequal or equal status or relative importance, that is either one mode dominates or leads the other (e.g., *illustration, anchorage* in the case of text-image relations) or both modes have the same significance and equally and mutually contribute to the overall meaning (*relay*). In the present example, the webpage seems dominated and led by the painting—its main communicative functionality is to visually explore the painting as you would in a visit to the museum; none of the lexical references to the image would make sense in its absence. A second gain lies in adopting “specific logical relationships taken to hold between text and image regardless of their relative status” (Bateman, 2014a, 156). These relationships are described as *logicosemantic* and are ways of information linking. In the example, the text seems to mainly elaborate or restate

information contained in the painting (*elaboration*) but it also offers circumstantial info about the painting, e.g., time, place, manner (*enhancement*). In comparison with the analysis of cohesive ties, the ‘grammatical’, clause-combinational approach treats mode combinations as structural units which weigh their individual components semantically and arrange them in pre-determined semantic relations. An obvious way of merging both approaches would be to set up different classes or sets of multimodal cohesive ties (e.g., those to do with the painting’s theme, materiality, provenance, etc. or those that are collocational or meronymic) and individually explore their status and logicosemantics. What would seem to be a drawback of the grammatical approach is the apparent lack of specificity and precision in the available distinctions of logicosemantic relations, which seem hard to apply to different genres and media.

Two further approaches to multimodal coherence discussed by Bateman (2014a, 157–163), namely *conjunctive relations* (Martin & Rose, 2003, 119) and *rhetorical structure theory* (Mann & Thompson, 1988), locate mode linking not in the grammatical structure of the clause and its combinations but move it to the level of discourse or text structure. This means that recipients are seen to construct semantic links as they process a multimodal artifact from any discourse element they deem significant and suited to the semiotic work of establishing a general coherent text structure. Both approaches are essentially concerned with identifying recurrent and generalizable *relational propositions* (Mann & Thompson, 1986), i.e., standard ways in which statements and ideas may be connected semantically. While the conjunctive-relations approach (merely) inventories the relations by stating what textual or pictorial element performs what function in relation to another element (e.g., specify, summarize), the rhetorical-structure approach aims to formally model the discourse structure by arranging binary spans of propositions in relation (i.e., a combination of a nucleus and a satellite; e.g., nucleus — PURPOSE — satellite) on hierarchical levels.

In our example, the text *names* major image elements of the painting, *restates* the depicted scene, *specifies* technical details, *explains* temporal and local context, and *interprets* the icon’s impact on art, etc. Interestingly, this conjunctive-relations interpretation is not so different from what Bateman (2014a, 163) calls “approaches based on speech acts, interaction and action”, as essentially the multimodal relations are ascriptions to text elements of communicative actions they are taken to perform in relation to another element. In a rhetorical-structure perspective, the painting and its various image elements are discourse nuclei, to which the various propositions of the text relate as satellites. The relations that occur are exclusively of the subject-matter type (not of the presentational type): ELABORATION, CIRCUMSTANCE, INTERPRETATION, EVALUATION, and RESTATEMENT. This tells us that the sample is of the explanatory, expository kind, where “the intended effect

is that the reader recognizes the relations in question” (Mann & Thompson, 1988, 257). This is different in presentational relations (e.g., MOTIVATION, EVIDENCE, JUSTIFY), “whose intended effect is to increase some inclination in the reader, such as the desire to act or the degree of positive regard for, belief in, or acceptance of the nucleus” (Mann & Thompson, 1988, 257)—something more akin to the persuasive multimodal argument construed in the advertisement (see Figure 2).

In both discourse-semantic accounts of multimodal coherence, the nature of the sample’s art-didactic genre is highlighted adequately: conjunctive relations work to pick out and characterize parts and contextual aspects of the painting, and spans relate pictorial nuclei to verbal satellites. What emerges, therefore, is a classic image-centric genre, where text elaborates and contextualizes a painting. The layout of the webpage supports this reading: the top part presenting the painting affords the visual exploration of the image and its museum context, the mid-layer consists of neatly structured and paragraphed verbal explanatory text, and the bottom area delivers short labels and tags to summarize main properties and aspects of the painting and to point to related materials in the art-didactic context (i.e., other paintings and didactic videos). It is an obvious advantage of the conjunctive-relations and rhetorical-structure approaches that they duly attend to genre discrimination and patterns typical for their recognition and that they factor in spatial structures (or modes) such as layout, typography, or page- and image-flow. It is in this sense that “the space [...] is never neutral and may carry a variety of additionally charged ‘functional positions’ depending on genre and the constraints of the [...] artifact” (Bateman, 2014a, 166).

Finally, in an effort to simplify the analysis of mode linking and integrate major ideas about the description of multimodal coherence in one framework, I have suggested elsewhere (Stöckl, 2015; Stöckl, 2016; Stöckl, 2016) a hermeneutic rationale. It looks at mode linking as semiotic work performed by the recipient on the basis of the materials perceived, the structures processed and, most importantly, on the basis of conventionalized cognitive operations and contextual knowledge. The framework, therefore, places multimodal coherence not with formal cohesion, not with grammatical structure, nor with discourse semantics alone, but views it primarily as a potentially comprehensive, multi-level, purpose- or task-driven engagement with the informative and communicative import and impact of modes. Consequently, the descriptions produced by the framework are driven by the specificity of the genre at hand, entailing the strong suggestion that mode linking and multimodal coherence are determined by the constraints of the genre—in terms of function, structure of communicative actions, modal forms and styles. The two examples are a case in point; they construe multimodal coherence according to the generic conventions of explanatory, educational arts webpages and argumentative minimalist advertising, respectively.

The hermeneutic rationale comprises five recipient activities:

- (1) allocating the sign repertoires to modes and developing a rough internal (spatial or temporal) structure of the artifact (see multimodal configuration in Stöckl, 2015, 62);
- (2) constructing a structure of communicative actions;
- (3) constructing a thematic/logical structure of content;
- (4) identifying formal and propositional coherence relations;
- (5) realizing inter-textual and inter-discursive relations.

Again in application to the art-didactic webpage, the recipient will likely

- (1) differentiate image, text, and graphic elements serving navigation/ hyperlinking and realize the essentially three-part division of the text into painting, text, and summary/links;
- (2) understand that the text explains, interprets, and contextualizes the painting in many different ways;
- (3) realize that the info given on the painting neatly falls into types of themes, as for example, content, provenance, material, style/technique, art-historical impact, related works of art;
- (4) notice that the text points to the painting through demonstrative reference and establishes numerous cohesive ties through lexis that takes up all major aspects of the painting and falls into conventional frame- and script-knowledge of works of art (e.g., painter, motif, period/style, material, interpretation, etc.); and
- (5) comprehend that the painting in question is related in manifold ways to 'similar' paintings, relies on a well-known mythological scene and has entered into a rich art-historical and philosophical discourse.

Figure 2 demonstrates that renaissance/mythological and expressionist visual discourses can be made to connect and in turn be argumentatively utilized in an automotive-engineering discourse.

Beyond its integrational motives, the framework has some merits in its orientation on recipients' putative semio-cognitive actions and in its broad reach that can still accommodate specification or detailing from any kind of theoretical angle.

2.4 Rhetorical Design and Multimodal Discourse

In this final key-concepts section, we address those higher-level entities and processes that come into play when modes fashioned from materials and commanding specific reaches multiply their meanings through mode-linking and produce coherent multimodal artifacts. First, the idea of *design* (see Figure 6) has become useful for describing “the situated process in which a sign maker chooses and arranges semiotic resources to realize a particular social function or purpose” (Jewitt et al., 2016, 156). In this way, design links multimodal meaning making to *social and communicative action* (Bateman et al., 2017, 65), which involves interested rhetors who pursue their goals by adopting a textual form that best realizes their purposes in a given context. The advertisement (see Figure 2), for instance, has been designed by an agency functioning as a rhetor who acts on behalf of the commissioning company to promote original Volkswagen spare parts for classic car models. Its design adopts a special *rhetorical strategy*, i.e., an “established means of doing particular kinds of communicative work (commercial persuasion—H.S.) by deploying and making selections in the available semiotic modes” (Hiippala, 2014, 114). The selection and arrangement of the modes respond to regulating genre regimes (for instance, it is customary in current advertising to combine minimal text with large-format images) as much as to the specific situation, in which, apparently, the targeted audience and the advertised product made fusing two iconic paintings and their utilization in a simple multimodal argument (cf. Kjeldsen, 2012) seem a suitable strategy. Kress (2010, 121) places such rhetorical strategies in what he calls *rhetorical processes* that “underlie, precede and then become design processes. Bateman et al. (2017, 131) link rhetorical strategies to genre by designating genres as “bundles of strategies for achieving particular communicative aims in particular ways”. These views emphasize the importance of rhetorical approaches in the widest sense to multimodal communication (cf. Stöckl, 2014).

Concrete multimodal text exemplars are always representative of a genre; our use of the genre names art-didactic webpage and advertisement in reference to the samples has demonstrated this aptly. In my view, *multimodal genres* is one of the most central concepts in the field because—as Bateman (2016, 60) puts it—“without genre allocation, it is often not possible to provide a sensible description of a text, multimodal or not, at all”. In addition, I argued above that issues of modal discourse semantics only properly play out in co- and context, which are primarily determined by the genre. Based on definitions by Swales (1990, 58) and Martin & Rose (2003, 8), genres can be characterized as recurrent classes of texts, that

- (1) share a given communicative purpose which is achieved by
- (2) a series of semiotic actions (stages),

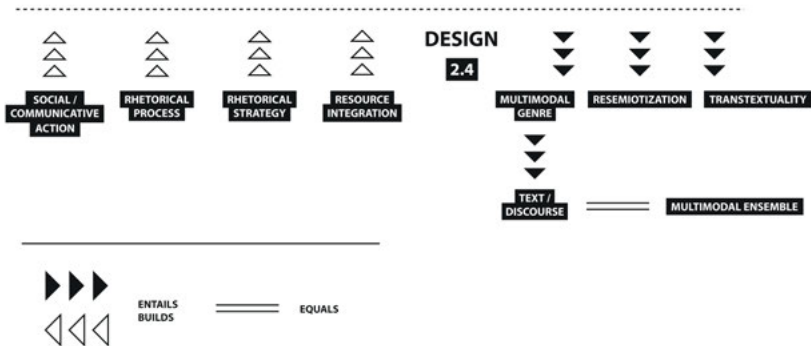


Fig. 6: A graphic model of the conceptual elements of design (Visualization: Jana Pflaeging & Hartmut Stöckl).

- (3) have more or less conventionalized forms of structure and organization, and
- (4) display expectable content and style of semiotic form that are constrained by
- (5) the communicative situation.

The different situations of both samples (arts webpages vs. advertisement) and their differing communicative purposes (explaining vs. persuading) give rise to different text structures (i.e., selection and ordering of stages), mode combinations, and styles of using semiotic resources (IMAGE: authentic vs. doctored; TEXT: matter of fact/explicit vs. playful/economical). Multimodal genre analysis must specifically address what has been called *orchestration* of modes (Kress, 2010, 157) or *semiotic resource integration* (Baldry & Thibault, 2006, 18). Both concepts highlight the interdependent combination of modes and the resulting multimodal discourse semantic structure, which can be teased out in cohesion/coherence analyses (see above). These aim to determine the communicative stages of a genre and look at the ways those stages are realized by the various modes and their interplay.

While genre is largely uncontroversial in its application to multimodal communication, the use of text is contested. Rejected for its narrowing associations with written language, it has been replaced by some others, most notably *multimodal ensemble* (cf., e.g., Bezemer & Kress, 2016, 28–30). On the other hand, text is justified by some (Bateman et al., 2017, 132) because it emphasizes a property common to all multimodal artifacts or actions, namely “that they are structured in order to be interpreted”. A snappy definition of text would then be: “a text is

what you get whenever you actually use the semiotic modes of a medium to mean something” (Bateman et al., 2017, 132).

Traditionally, texts as material artifacts are distinguished from *discourse* as “the higher level of abstractness” (Bateman et al., 2017, 133), so that it is correct to say discourses are realized through texts and texts in their turn realize “sets of semiotic modes (...) as well as the semiotic relations involved” (Bateman et al., 2017, 132). In any case, (*multimodal*) *discourse* is as controversial as text but reveals an interesting duality. From a macro-perspective it can be understood as “socially constructed knowledge of some aspect of reality” (Kress & van Leeuwen, 2001, 4) and thus feeds into the design of a multimodal artifact as part of its rhetorical strategy. On the other hand, such knowledges can only be produced through (a series of) multimodal texts. The advertisement strongly relies on some knowledge of the two iconic paintings (i.e., that they are classic paintings, that *The Birth of Venus* epitomizes beauty and *The Cry* fear) in its *multimodally constructed argument* (i.e., If you don’t like to be in for a nasty surprise and if you want your spare parts to fit and look beautiful on your classic car, use *Volkswagen classic parts*.). Roque (2017, 42) would classify the painting as a ‘visual flag’ that attracts attention to what he calls a ‘joint multimodal argument’ as text and image are closely intertwined in it. The art-didactic webpage seeks to provide and construe some of this knowledge, and *The Birth of Venus* itself relies on the knowledge of a fragment of mythical discourse.

From a micro-perspective, discourse may be seen as “local meaning-making mechanisms” (Bateman et al., 2017, 135) that operate between units of text or semiotic modes. Such a view is more akin to concerns with discourse interpretation as it sees discourse as “what happens when we make sense of these contributions to an unfolding communicative situation” (Bateman et al., 2017, 135). This “act-by-act unfolding of signifying activities” (Bateman et al., 2017, 135) that underlies discourse construction also usefully applies to a *trans-textual perspective* highlighting the fact that multimodal discourse may unfold in whole series of inter-textually or inter-medially related texts. Mythological narrative, iconic painting, art-didactic webpage, and multimodal commercial argument would be such an exemplary series of texts which are discursively connected. Iedema (2003, 296) uses the concept of *resemiotization* to describe “how semiotics are translated from one into the other as social processes unfold” and “why these semiotics (rather than others) are mobilized to do certain things at certain times” (Jewitt et al., 2016, 40).

3 Summary and Concluding Hypotheses

This brief, bird's-eye-view sketch has attempted to outline the conceptual foundations of current (linguistic) multimodality research by drawing up a network of core terms and using them to present a largely consensual rationale for multimodal studies and concrete analyses (see Figure 7). My exemplification of most of this conceptual and methodological framework with the help of a trans-textual, mixed-genres multimodal discourse has also striven to spotlight some approaches and views as particularly useful and recommendable. By and large, however, I have sought to present a balanced view doing justice to different interpretations and enabling the reader to see how key ideas have developed and relate to one another. In this final section, I will summarize some of the main points in the way of advancing cautious hypotheses or suggesting future ways to tread.

3.1 Semiotic Modes Need Delineating and Describing

I have advocated a view of mode(s) here that envisages a stratification of the concept into medial, perceptual, 'lexico-grammatical', and (discourse) semantic layers. Such a view may best promote a thorough investigation of individual modes and help to bring out mode differences—foundational semiotic work that still needs doing. While I expressed ambivalence towards including discourse semantics into the conception of individual modes, it should be clear that for modes to work in practice they need to have a specific kind of disposition for affording certain kinds of meanings and realizing special discourse functions. I feel that the idea of modal reach or logic already covers some of that ground.

Based on empirical mode descriptions, research needs to catalogue or map distinct modes and their relations, a job that can only be done by looking at a large variety of multimodal genres and discourses. Two tasks would be part and parcel of this undertaking: decide which complexity is required to talk about modes (rather than about individual resources of modes) and distinguish central, core, full-blown modes from less-developed, peripheral modes. So, for instance do we designate color as a mode in its own right or do we treat it as a semiotic resource that can be a part of different modes (e.g., image, typography, etc.) or social practices (advertising, interior design)? Such foundational social semiotic work must ultimately aim to develop an understanding of the differing structural nature, relative status and social currency of modes.

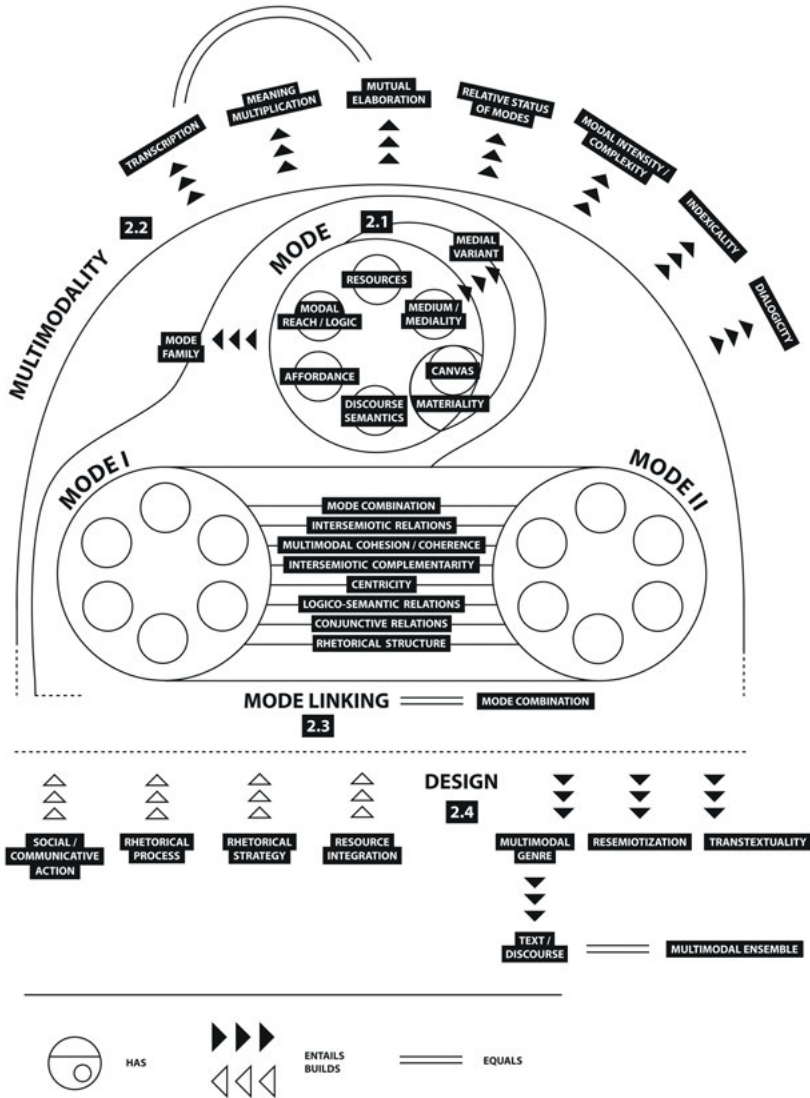


Fig. 7: A graphic overview of the concepts constituting a theory of (linguistic) multimodality (Visualization: Jana Pflaeging & Hartmut Stöckl).

3.2 Meaning Multiplication Happens Bottom-Up and Top-Down

It is a commonsense argument to say that the reach or logic of a mode is shaped and ultimately determined by its semiotic resources, i.e., essentially by the lexico-

grammatical regularities available in that mode. If we factor in something like a discourse semantics typical of a mode and medial/material peculiarities, we arrive at a fairly apt idea of what a given mode can contribute to meaning multiplication in multimodal ensembles. The specifics of any kind of meaning multiplication must, however, not only be described as deriving—bottom-up—from its modal ‘ingredients’ and dispositions but also—top-down—as being controlled and steered by the constraints of social/communicative action, best encapsulated in the notion of multimodal genre. So, if we are asking exactly how meaning is multiplied, we need to look both ways: at the constitutive properties of the modes in the multimodal ensemble and at the genre-induced patterns of mode integration. It would seem to be a pertinent task in this connection to outline what exactly a discourse semantics might be in relation to individual modes (Wildfeuer & Bateman, 2018, 19–21) and where and in which ways a discourse semantics operates across or between the modes—the latter a question adequately addressed already in accounts of multimodal coherence.

3.3 Mode-Linking is Socio-Cognitive Action Based on Rhetorical Design/Structure

The outline of various approaches to multimodal linking and coherence will have made it clear that the complementarity of modes must be described on different levels. Whatever model we favor, it would need to account for two interrelated multimodal facts: first, mode-linking and coherence-construction are cued by cohesive ties and by aspects of a given rhetorical structure that follows a more or less explicit design. Second, as they are involved in a rhetorical process with goals, tasks, and structures in a given situation, producers and recipients of multimodal artifacts alike will construe multimodal coherence from whatever textual cues available mainly on the basis of cognitive semiotic routines and knowledge acquired in the varied landscapes of communication. Both—rhetorical structures and semio-cognitive mechanisms of interpretative construal—must be highly genre-sensitive, i.e., rhetorical structure depends on the specifics of the genre in the same way as cognitive-semiotic work will have to be attuned to and uncover them. In other words, whatever the choice of framework for the description of multimodal coherence, it would do well to fuse structural aspects with plausible steps of interpretative practice.

3.4 Multimodal Discourse Interpretation is Genre-Based and Trans-Textual

Placing multimodal ensembles firmly in the context of a rhetorical situation means to look at them as comprising a rhetor's goal, rhetorical strategies as deliberate choices and combinations of semiotic resources, and a recipient's task-based communicative engagement in the resulting multimodal structure. Such a consistently rhetorical approach allocates genre a central role in shaping and constraining multimodal discourse interpretation. Any multimodal artifact would then first of all be an exemplar instantiating or realizing a genre's underlying functional, logical, structural, and stylistic regime. Uncovering multimodal genre—as I have argued—entails describing how textual structures and their stages are orchestrated by various modes and their combinations. Most importantly, as a consequence, local multimodal discourse interpretation must invariably be a genre-based and trans-textual venture, which requires adequate knowledge of genre-conventions, individual texts and discourses, all of which may enter and combine in the understanding of a multimodal artifact and its discourse trajectories.

If at all this sketch has stretched the boundaries of the field, it is by emphasizing and endorsing three perspectives or ways of doing multimodality research. These three might conveniently be labelled as 'rhetorical', 'semio-cognitive', and 'trans-textual'. First, a rhetorical view engenders a wide notion of context embracing situation, participants, social and communicative action, goals, tasks and design—but it also highlights multimodal structure as a genre-sensitive focal point for all descriptions of multimodal meaning-making. Second, a semio-cognitive perspective foregrounds the active construal of multimodal coherence from both structural cues as well as genre and discourse knowledge. Third, and finally, a trans-textual take of multimodal discourse highlights the need for sign-makers and consumers alike to have at their disposal a rich knowledge of genre conventions, discourses and individual texts so they can combine them, allocate them correctly and construe sense beyond the narrow confines of a given local multimodal discourse structure.

Bibliography

- Baldry, Anthony & Paul J. Thibault. 2006. *Multimodal Transcription and Text Analysis: A Multimedia Toolkit and Coursebook with Associated On-Line Course*. London and New York: Equinox.
- Bateman, John A. 2008. *Multimodality and Genre: A Foundation for the Systematic Analysis of Multimodal Documents*. Basingstoke: Palgrave Macmillan.

- Bateman, John A. 2011. The Decomposability of Semiotic Modes. In Kay L. O'Halloran & Bradley A. Smith (eds.), *Multimodal Studies: Multiple Approaches and Domains*, 17–38. London: Routledge.
- Bateman, John A. 2014a. Multimodal Coherence Research and its Applications. In Helmut Gruber & Gisela Redeker (eds.), *The Pragmatics of Discourse Coherence*, 145–177. Amsterdam: John Benjamins.
- Bateman, John A. 2014b. *Text and Image: A Critical Introduction to the Visual/Verbal Divide*. London and New York: Routledge.
- Bateman, John A. 2016. Methodological and Theoretical Issues for the Empirical Investigation of Multimodality. In Nina-Maria Klug & Hartmut Stöckl (eds.), *Sprache im multimodalen Kontext / Language and Multimodality* (Handbooks of Linguistics and Communication Science (HSK) 7), 36–74. Berlin: de Gruyter Mouton.
- Bateman, John A., Janina Wildfeuer & Tuomo Hiippala. 2017. *Multimodality – Foundations, Research and Analysis. A Problem-Oriented Introduction*. Berlin: Mouton de Gruyter.
- Bezemer, Jeff & Gunther Kress. 2016. *Multimodality, Learning and Communication: A Social Semiotic Frame*. New York and London: Routledge.
- Caple, Helen. 2013. *Photojournalism. A Social Semiotic Approach*. Basingstoke and New York: Palgrave Macmillan.
- Goodman, Nelson. 1969. *Languages of Art. An Approach to a Theory of Symbols*. London: Oxford University Press.
- Halliday, Michael A. K. 1978. *Language as Social Semiotic*. London: Edward Arnold.
- Halliday, Michael A. K. & Ruqaiya Hasan. 1985. *Language, Context and Text: A Social Semiotic Perspective*. Geelong, Victoria: Deakin University Press. (Language and Learning Series). Also: Oxford University Press, London, 1989.
- Hiippala, Tuomo. 2014. Multimodal Genre Analysis. In Sigrid Norris & Carmen Daniela Maier (eds.), *Interactions, Images and Texts: A Reader in Multimodality*, 111–123. Berlin: Mouton de Gruyter.
- Iedema, Rick. 2003. Multimodality, Resemiotization: Extending the Analysis of Discourse as Multi-Semiotic Practice. *Visual Communication* 2(1). 29–58.
- Jäger, Ludwig. 2002. Transkriptivität. Zur medialen Logik der kulturellen Semantik. In Ludwig Jäger & Georg Stanitzek (eds.), *Transkribieren. Medien/Lektüre*, 19–42. München: Wilhelm Fink.
- Jewitt, Carey. 2009a. An Introduction to Multimodality. In Carey Jewitt (ed.), *The Routledge Handbook of Multimodality*, 14–29. London, New York: Routledge.
- Jewitt, Carey (ed.). 2009b. *The Routledge Handbook of Multimodal Analysis*. London: Routledge.
- Jewitt, Carey. 2014a. An Introduction to Multimodality. In *The Routledge Handbook of Multimodal Analysis. Second Edition*, 15–30. London and New York: Routledge.
- Jewitt, Carey (ed.). 2014b. *The Routledge Handbook of Multimodal Analysis. Second Edition*. London and New York: Routledge.
- Jewitt, Carey, Jeff Bezemer & Kay O'Halloran. 2016. *Introducing Multimodality*. London: Routledge.
- Kjeldsen, Jens. E. 2012. Pictorial Argumentation in Advertising: Visual Tropes and Figures as a Way of Creating Visual Argumentation. In F. van Eemeren & B. Garssen (eds.), *Topical Themes in Argumentation Theory. Twenty Exploratory Studies*, 239–255. Amsterdam: Springer.

- Klug, Nina-Maria & Hartmut Stöckl. 2015. Sprache im multimodalen Kontext. In Ekkehard Felder & Andreas Gardt (eds.), *Handbuch Sprache und Wissen* (Handbooks of Linguistics and Communication Science (HSK) 1), 242–264. Berlin: de Gruyter Mouton.
- Klug, Nina-Maria & Hartmut Stöckl (eds.). 2016. *Handbuch Sprache im multimodalen Kontext*. Berlin: de Gruyter.
- Kress, Gunther. 2010. *Multimodality: A Social Semiotic Approach to Contemporary Communication*. London: Routledge.
- Kress, Gunther, Carey Jewitt, Jon Ogborn & Charalampos Tsatsarelis. 2000. *Multimodal Teaching and Learning. The Rhetorics of the Science Classroom*. London: Continuum.
- Kress, Gunther & Theo van Leeuwen. 2001. *Multimodal Discourse. The Modes and Media of Contemporary Communication*. London: Arnold.
- Kress, Gunther & Theo van Leeuwen. 2006 [1996]. *Reading Images: The Grammar of Visual Design*. London and New York: Routledge.
- Ledin, Per & David Machin. 2018. *Doing Visual Analysis: From Theory to Practice*. London: Sage.
- van Leeuwen, Theo. 2005. *Introducing Social Semiotics*. London: Routledge.
- Lemke, Jay L. 1998. Multiplying Meaning: Visual and Verbal Semiotics in Scientific Text. In James R. Martin & Robert Veel (eds.), *Reading Science: Critical and Functional Perspectives on Discourses of Science*, 87–113. London: Routledge.
- Lyons, Agnieszka. 2018. Combining Modes: Multimodality across Communicative Contexts. In Colleen Cotter & Daniel Perrin (eds.), *The Routledge Handbook of Language and Media*, 296–308. London, New York: Routledge.
- Mann, William C. & Sandra A. Thompson. 1986. Relational Propositions in Discourse. *Discourse Processes* 9(1). 57–90.
- Mann, William C. & Sandra A. Thompson. 1988. Rhetorical Structure Theory: Toward a Functional Theory of Text Organization. *Text* 8(3). 243–281.
- Martin, James R. & David Rose. 2003. *Working with Discourse: Meaning Beyond the Clause*. London and New York: Continuum.
- Martinec, Radan & Anthony Salway. 2005. A System for Image-Text Relations in New (and Old) Media. *Visual Communication* 4(3). 337–371.
- Norris, Sigrid. 2004. *Analyzing Multimodal Interaction: A Methodological Framework*. London and New York: Routledge.
- Norris, Sigrid & Carmen Daniela Maier (eds.). 2014. *Interactions, Images and Texts. A Reader in Multimodality*. Boston and Berlin: de Gruyter.
- Page, Ruth, David Barton, Joann W. Unger & M. Zappavigna. 2014. *Researching Language and Social Media. A Student Guide*. London and New York: Routledge.
- Pirini, Jesse. 2014. Introduction to Multimodal (Inter)Action Analysis. In Sigrid Norris & Carmen D. Maier (eds.), *Interactions, Images, and Text: A Reader in Multimodality*, 77–91. London and New York: Routledge.
- Roque, Georges. 2017. Rhetoric, Argumentation and Persuasion in a Multimodal Perspective. In Assimakis Tseronis & Charles Forceville (eds.), *Multimodal Argumentation and Rhetoric in Media Genres*, 25–49. Amsterdam: John Benjamins.
- Royce, Terry D. 1998. Synergy on the Page: Exploring Intersemiotic Complementarity in Page-Based Multimodal Text. *Japan Association for Systemic Functional Linguistics (JASFL) Occasional Papers* 1(1). 25–49.
- Stöckl, Hartmut. 2004. In Between Modes: Language and Image in Printed Media. In Eija Ventola, Cassily Charles & Martin Kaltenbacher (eds.), *Perspectives on Multimodality*, 9–30. Amsterdam: John Benjamins.

- Stöckl, Hartmut. 2014. Rhetorische Bildanalyse. In Netzwerk Bildphilosophie (ed.), *Bild und Methode: Theoretische Hintergründe und Methodische Verfahren der Bildwissenschaft*, 377–390. Köln: Halem.
- Stöckl, Hartmut. 2015. From Text Linguistics to Multimodality: Mapping Concepts and Methods across Domains. In Janina Wildfeuer (ed.), *Building Bridges for Multimodal Research: International Perspectives on Theories and Practices of Multimodal Analysis*, 51–75. Frankfurt am Main: Peter Lang.
- Stöckl, Hartmut. 2016. Multimodales Verstehen: Zwischen Zeichensystemwissen und Textsortenkompetenz. In Klaus Sachs-Hombach (ed.), *Verstehen und Verständigung: Intermediale, multimodale und interkulturelle Aspekte von Kommunikation und Ästhetik*, 88–112. Köln: Halem.
- Stöckl, Hartmut. 2016. Multimodalität – Semiotische und textlinguistische Grundlagen. In Nina-Maria Klug & Hartmut Stöckl (eds.), *Handbuch Sprache im multimodalen Kontext* (Handbooks of Linguistics and Communication Science (HSK) 7), 3–35. Berlin: de Gruyter Mouton.
- Stöckl, Hartmut, Helen Caple & Jana Pflaeging. 2020. *Shifts towards Image-centricity in Contemporary Multimodal Practices*. New York: Routledge.
- Swales, John M. 1990. *Genre Analysis: English in Academic and Research Settings*. Cambridge: Cambridge University Press.
- Tseronis, Assimakis & Charles Forceville. 2017. Argumentation and Rhetoric in Visual and Multimodal Communication. In Assimakis Tseronis & Charles J. Forceville (eds.), *Multimodal Argumentation and Rhetoric in Media Genres*, 1–24. Amsterdam: John Benjamins.
- Tseronis, Assimakis & Chiara Pollaroli. 2018. Pragmatic Insights for Multimodal Argumentation. *International Review of Pragmatics* 10. Special Issue.
- van Leeuwen, Theo. 1999. *Speech, Music, Sound*. London: Macmillan.
- Wildfeuer, Janina & John A. Bateman. 2018. Theoretische und methodologische Perspektiven des Multimodalitätskonzepts aus linguistischer Sicht. *IMAGE – Zeitschrift für Bildwissenschaft* 28. 5–45.

Martin Thomas

Making a Virtue of Material Values: Tactical and Strategic Benefits for Scaling Multimodal Analysis

Abstract: Multimodality has gained significance ‘wherever meaning is the issue’. Beyond the ancestral disciplines of linguistics, semiotics, and education there is an increasing intensity in our dialogue with people working in cognate fields, such as information design, media studies, human-computer interaction, and, broadly, the digital humanities. While multimodality seeks to assert itself as a discipline, it is through interdisciplinary activity that we are likely to produce socially relevant outcomes. Within multimodality, those who have long been calling for greater empirical rigor have been joined by scholars coming from more interpretive, social semiotic perspectives in recognizing the limitations of small-scale intensive analyses. We argue that it is precisely these limitations that constrain the application of multimodality to tackling real world problems. The proposed approach to this challenge is distinct in that: (1) it works from the ground up—on the assumption that it is easier to aggregate features in response to particular questions than it is to disaggregate them post hoc; (2) it aims to yield datasets automatically; and (3) it side-steps annotation by relying on forms of representation that exploit properties inherent to digital content. It does this by focussing on the material features of two-dimensional documents, which can be seen, essentially, as combinations of x-y coordinates and variation in values at each point.

Keywords: multimodality, materiality, scale, annotation, semiotic mode

1 Multimodality, Materiality, Scale, and Application

Multimodal analysis entails confronting the materiality of texts (Kress & van Leeuwen, 2001)—their composite look and feel, the contexts of their production and consumption. Rather than isolating words from their context and graphic expression, multimodality commits to looking at them as they are placed on the page or screen: their size, shape, color, and weight. It also seeks to account for images and any other graphic devices that contribute to meaning-making.

This concern with materiality resonates particularly in an age in which technology has become such a focus of attention. In this context it is little surprise that

<https://doi.org/10.1515/9783110608694-003>

in the years since Gunther Kress claimed that “wherever meaning is the issue, the concepts of mode and multimodality are rapidly gaining significance” (Kress, 2009, 54), multimodality has indeed become a buzzword across disciplines—including education (e.g., Early et al., 2015), media and communication studies (e.g., Pauwels, 2012), and translation studies (e.g., Pérez-González, 2014; Ramos Pinto & Gambier, 2018)—wherever the primary object of study is communication, situated as it must be in some social context.

Within multimodality there is a growing appreciation of the need for empirical methods for analysis that are capable of handling data at a scale large enough to support the kinds of generalization that might inform prediction, while also being sufficiently delicate to help make sense of the complex combinations of meaning-making resources (see Bateman, 2008; Jewitt, 2009; O’Halloran et al., 2016). This chapter responds by offering a concrete proposal to inform methods for the representation, searching, and analysis of multimodal data as they are found in the wild. This proposal is founded on a clearing of the ground—a fundamental re-examination of some methodological practices and conceptual assumptions of multimodality. This constitutes the bulk of the chapter.

The ground clearing aspect of this chapter seems especially timely, as multimodalists claim their disciplinary independence. The conference which inspired the current volume aimed to lay the foundation for the formation of a stand-alone discipline dubbed ‘multimodality’ as opposed to the widespread interdisciplinary view (see Wildfeuer et al. in this volume). While I remain agnostic about this move, it does suggest that multimodality has reached a certain stage in the disciplinary life-cycle. This is reflected also in other research communities, such as machine translation (see, e.g., Shah et al., 2016), who perceive potential benefit in exploiting multimodal information in pursuit of their own goals. At the same time, those of us working in multimodality are receiving expressions of interest in collaboration from colleagues working in disciplines whose primary focus is not communication, but who are associated with communities of practice in which effective communication is vital and in which failure entails significant risks, such as healthcare and finance. It is perhaps worth stating explicitly that, as we seek to apply multimodal analysis with the aim of generating real-world benefits, we will need to be confident of the validity of our findings.

Given the breadth of multimodality as a field of interest, and in order to meet the need for reliable generalizations, I specify two-dimensional documents as the object of study for the approach proposed here. The approach is distinctive in that: (1) it works from the ground up—on the assumption that it is easier to aggregate features in response to particular questions than it is to disaggregate them post-hoc; (2) it aims to yield datasets automatically; and (3) it side-steps annotation by relying on forms of representation that exploit properties inherent to documents, whether

they take digital form throughout their life-cycle or are designed and/or consumed in print. It does this by focussing on the material features of two-dimensional documents, which can be seen, essentially, as combinations of x-y coordinates and variation in values at each point (see Bertin, 2010). Rather than aiming to interpret meaning as such, my intention is to move towards a situation in which we can reliably identify patterns in the use of meaning-making resources across large sets of data, thus providing the means to respond to diverse research agendas.

2 State of the Art: Current Limitations and a Way Forward

There is a growing acknowledgement that multimodal analysis is currently limited in terms of scale, even among scholars established at the more interpretive end of social semiotics, and that this “can restrict the potential of multimodality to comment beyond the specific to the general” (Jewitt, 2009, 27). Jewitt goes on to recognize multimodal corpora as a possible means of overcoming this limitation. In this section, I will develop Jewitt’s observation in two directions. Firstly, I seek to unpack the limitation of scale—in particular, to problematize the weakness of current approaches to multimodality in generalizing from specific cases. Secondly, I look at some of the specific technical and conceptual difficulties in developing multimodal corpora.

2.1 The Need for Empirical Methods that Scale

The inability to conduct multimodal analysis at scale leads to a number of problems, not least of which is the lack of evidence on which to base claims about generalized patterns (Forceville, 2007; Bateman, 2011). Moreover, where claims are made that go beyond those obvious to the “moderately attentive viewer-reader” (Forceville, 2007, 1236), they often lack empirical basis and are open to question through counter-examples. A case which has attracted particular criticism is Kress and van Leeuwen’s (1996) extension of the given-new thematic structure typical in spoken and written English to account for information values that hold in two-dimensional layout (see, e.g., Forceville, 1999; Bateman et al., 2004; Thibault, 2000; Thomas, 2009b, 2014; Waller et al., 2012).

Thus, approaches based on ‘close reading’ of hand-picked examples have left the field of multimodal analysis open to critiques that focus on the lack of empirical grounding and generalizability of results (e.g., Bateman et al., 2004; Thomas, 2014).

Aside from providing material for academic discussion, this weakness critically undermines the usefulness of multimodal analysis in real-world applications.

Calls for approaches which benefit from a more rigorous empirical grounding are not new. Perhaps most notably, Bateman et al. (2004, 66–67) make the case for approaches which distinguish between ‘occasionally plausible’ and ‘reliable property’, and a move away from ‘impressionistic interpretive’ accounts featuring ‘appealing social interpretation’ and ‘post-hoc rationalization’. They outline the desiderata of a methodological program in which the frames of analysis “can be expressed as predictive and falsifiable claims about document design and meaning-making” and subject to “more detailed and systematic investigation, varying types of documents, types of consumers, types of presentation medium, and purposes so that we can get a finer grip on the meaning-making possibilities of the various semiotics in play” (Bateman et al., 2004, 68).

Crucially, then, the frames of analysis themselves need to be transparent, open to question, and the scope of their validity testable and tested.

2.2 Corpus-Based Approaches

Like Jewitt, Bateman and colleagues propose the use of corpora as offering a solution to the problem of scalability in multimodal analysis. However, despite the consensus that seems to have grown over the past decade or so within the multimodality community, we have not moved beyond (1) prototype systems that lack the robustness to scale due to the degree of manual remedial intervention needed to correct errors in automatic processing (e.g., Thomas, 2009b); (2) tools that are “likely to be most effective for generating a baseline for manual annotation” (Hiippala, 2015, 88); or (3) approaches that rely on manual qualitative analysis to feed in to quantitative methods downstream (e.g., O’Halloran et al., 2016). Thus, moving from “manual analysis and discursive interpretation of a limited number of multimodal texts toward automated recognition of multimodal meanings across large data sets” remains the “latest challenge in the field” (O’Halloran et al., 2016, 17).

Looking beyond the multimodality community, there is relevant activity in several areas. The last decade or so has seen the rather niche activity of humanities computing metamorphose into digital humanities. This shift coincides with the adoption of computational techniques in parts of the mainstream of research in the arts—and, more or less, with the growth of interest in multimodality. In particular, the combination of cultural studies with data science and visualization techniques has been developed, perhaps most famously by Lev Manovich and colleagues, as *Cultural Analytics* (Manovich, 2009).

Cultural Analytics provides insights into the flows of cultural innovation: helping to shed light on what content is communicated where and through what channels. As such, there are parallels with the ‘distant reading’ approach to patterns in literary publishing by Moretti (2013). While yielding striking results, Cultural Analytics does not attempt to account for the meanings that are made through the combination of the various semiotic resources in play. To extend the analogy with literary studies, doing this still largely relies on close reading.

Work identified with more traditional areas of computer science also combines scale with analysis of specific meaning-making resources. For example, Reinecke & Gajos (2014) have undertaken work on website aesthetics on an impressive scale, though to make the approach tractable their scope is deliberately limited in terms of the visual features considered, i.e., ‘colorfulness’ and ‘complexity’. These are taken as proxies for aesthetic appeal. As such, while it provides significant quantities of data, it lacks the breadth and nuance of a more interpretive, general framework for the multimodal analysis of aesthetics in digital texts, such as that proposed by Adami (2015).

With broader ambitions, Kumar et al. (2013) took computer vision techniques and machine learning methods and applied them to web page design. Their system, *Webzeitgeist*, allowed the user to build queries around complexes of design features. For example, querying for pages featuring “roughly centered (vertically and horizontally) text INPUT elements, and fewer than 50 visual elements on the page” retrieves examples of search engine-like page designs (Kumar et al., 2013, 7). Thus, *Webzeitgeist* could be used to identify specific instances of complex aspects of design, such as layout, by taking combinations of features as proxies.

However, the project is now defunct and the system unavailable (personal communication with Kumar and Talton, 10-11/2016). While it seems to have offered great potential for those interested in data-driven approaches to web design, *Webzeitgeist* was subject to limitations. Firstly, it was engineered specifically to work with web pages, rather than other sorts of (digitized) documents. Secondly, Kumar and colleagues highlight the inherent problem presented by the lack of render-time stability in dynamically-generated web pages. Their solution was to capture a screenshot to establish a ground-truth for each web page analyzed. Clearly this would hold only for a particular instance of content generated under current conditions, layout having been rendered by particular software on a specific device. Thirdly, Kumar et al.’s system, being developed with graphic design as its focus, did not integrate the full range of semiotic resources used in web-pages; written language, for example, was not accommodated. Finally, and perhaps because it has not developed beyond a prototype, their system appears to have lacked a user-friendly interface of the sort that would be needed to support multimodal analysis.

While Reinecke & Gajos (2014) and Kumar et al. (2013) both have shortcomings, not least in terms of the limited range of semiotic resources they cover, they also offer inspiration—specifically in the use of complexes of computationally tractable, low-level features as proxies for more abstract phenomena and higher-level features.

Within the recent literature on multimodality, we find calls for a similar methodological tack for handling audiovisual data:

Since full automation of analysis for detection of more abstract visual and audiovisual patterns is still well beyond the state of the art, methods need to be developed for investigating abstract research questions with as much support as possible from less abstract levels of coding that are approachable automatically. (Bateman et al., 2016, 139–140)

In sum, across the field of multimodal analysis, in work inspired to varying degrees by social semiotics (e.g., Jewitt, 2009) and systemic-functional linguistics (e.g., O’Halloran et al., 2016), there are signs of a growing acceptance of the limitations of relying on the authority and capacity of the ‘interpretive subject’. The task at hand, then, is to operationalize scalable empirical approaches to multimodality. In addition to developing predictive and falsifiable claims, such approaches require the explicit articulation of parameters of constraint in order to take what Bateman et al. (2004, 68) call a “finer grip”. Finally, we have noted the value in developing an approach which automates the recognition of low-level, material features and then considers whether and how these might be used—by aggregating low-level features taken as proxies for higher-level features—in order to get traction on specific research questions.

3 What is (Not) a Mode?

I return to this perennial question again here briefly to re-examine the conceptual underpinnings of multimodality and to establish the theoretical basis for the approach I propose, which unapologetically assumes as little as possible.

Kress usefully points to a distinction between “socially oriented” and “formally oriented” responses to the question (see Kress, 2010, 84–92), a distinction which has been taken up broadly by the community and which will frame the discussion below.

3.1 Formally Oriented Definitions

A number of formal definitions of mode have been proposed, several of which invoke the three metafunctions as described by Halliday (1978): ideational, interpersonal, and textual (see, e.g., Kress & van Leeuwen, 1996; Kress, 2009; Stöckl, 2004). We might note that, not only did Halliday develop the metafunctional model for language, but it is founded explicitly on properties specific to language:

Typically each sentence embodies all functions, though one or another may be more prominent; and most constituents of sentences also embody more than one function, through their ability to combine two or more syntactic roles. (Halliday, 1973, 108)

Models predicated on the performance of all three metafunctions are open to question both in terms of the depth of insight they offer (Bateman, 2013) and the scope of their applicability. For example, Kress's account of layout (Kress, 2010, 88–92), which invokes commutation as a test for both the ideational and interpersonal metafunctions, could be seen essentially in textual terms (for a fuller discussion of multimodal texture, see Thomas 2009a).

Bateman (2011, 20–21) proposes a model of mode, whose three strata might be glossed thus: (1) material substrate; (2) non-material component comprising content and expression planes with contribution characterized in terms of paradigmatic and syntagmatic axes of organization; (3) distinct discourse semantics. This finer granularity in the description of the internal organization of mode is seen as necessary in order to be “responsive to empirical results without prejudging what is occurring” (Bateman, 2011, 18–19). While (3) might be the privileged criterion for the identification of a mode, both (2) and (3) rely on patternings in (1) for their realization and discernibility.

In sum, in formal terms, an approach to multimodal data based on patternings of such material features would thus seem compatible with Bateman's model, while at the same time being amenable to the empirical investigation of the scope and applicability of those of Kress (2009, 2010) and Stöckl (2004).

3.2 Socially Oriented Definitions

There is broad agreement in the literature that what might be considered a mode—and hence its scope and affordances—is context-bound and culturally constructed. Mode is relative. Kress et al. capture this succinctly, thus:

the question of whether X is a mode or not is a question specific to a particular community. As laypersons we may regard visual image to be a mode, while a professional photographer

will say that photography has rules and practices, elements and materiality quite different from that of painting and that the two are distinct modes. (Kress et al., 2000, 43)

The social, or functional, aspect of distinguishing between modes can also be framed around questions of reception, cognition, and analysis. Fundamental to multimodality is the observation that modes for making meaning operate in combination—indeed a key focus of discussion has been the ways in which such combination occurs, since these are understood not to be readily apparent to the casual reader. As Stöckl has it:

When ‘reading’ a multimodal text, average recipients will normally become only dimly aware of the fact that they are processing information encoded in different modes. The manifold inter-modal connections that need to be made in order to understand a complex message distributed across various semiotics will go largely unnoticed. All modes, then, have become a single unified gestalt in perception, and it is our neurological and cognitive disposition for multimodal information processing that is responsible for this kind of ease in our handling of multimodal artifacts. (Stöckl, 2004, 16)

Thus, only through a process of dissection can we explore cases of multimodal ‘overlapping’ and ‘mixing’ or compensation (Stöckl 2004, 19; for a survey of approaches to ‘intersemiosis’ see O’Halloran, 2008). Furthermore, in order to perform such a dissection, we need to be confident about what it is that we are seeking to isolate—and this cannot be taken for granted. In response to the well-known observation by Lemke (1998) that the combination of meaning-making resources in multimodal texts entails a multiplication of the ‘set of possible meanings’, Bateman asserts the need to establish precisely what it is we are multiplying:

The assumption of particular modes holding even prior to empirical investigation is one major reason why the vast majority of multimodal ‘analyses’ still go little beyond detailed description. (Bateman, 2011, 18)

In sum, then, mode is socially constructed and therefore a relative concept; it is necessary to be able to identify distinct semiotic modes before proceeding to analyzing their combination; the presence of modes is itself an empirical question the answer to which will vary from context to context and genre to genre.

It is here that the seemingly esoteric question of what constitutes a mode has very practical consequences—for multimodal analysis and its potential for application in real-world scenarios. On the one hand, if we are to construct datasets enriched to support automated retrieval of patterns of use in multimodal analysis, we need to settle on a finite set of computationally tractable features. On the other hand, we cannot assume a priori that particular modes hold in a given context.

In this light, the proposal to use proxies, i.e. features of a finer degree of material granularity, aggregations of which might support context-sensitive empirical approaches to mode, seems to resolve this paradox in a way that is both theoretically sound and operationally expedient. In other words, just as the focus on the material proposed here supports the investigation of different formal approaches to the concept of mode, it also supports the investigation of the extent to which specific modes are present or relevant for different communities or in different domains or genres.

4 To Transcribe or Not to Transcribe?

Having established the methodological motivation for a focus on material resources—thus making progress in identifying the categories of interest—, it remains to identify suitable means for their representation, which allows for their retrieval.

Manovich identifies two properties which follow directly from digitalness: digital data “can be described formally” and is “subject to algorithmic manipulation” (Manovich, 2001, 27). Here, ‘manipulation’ should not be confused with ‘understanding’ (Bateman et al., 2016, 132). Software might provide support and ways into data, but ‘understanding’ and ‘interpretation’ of results, as well as the questions asked, rely on humans. While digitalness might give a helping hand over the first hurdle of data collection, it does not resolve the issue of representation, which is needed to make data available for analysis.

Perhaps reflecting practice in linguistics, much of the literature assumes that multimodal corpora involve transcription. Given our commitment to materiality—and that “transcription leads necessarily to a separation of content from form” (Bertin, 2010, 4), issues around multimodal transcription are particularly fraught and have been given significant attention. As with the ‘mode’ question, here some assumptions might usefully be unpacked.

4.1 Transcription and Decision-Making

Ochs’s seminal paper *Transcription as Theory* raises some fundamental concerns that pre-date both widespread computational analysis in the humanities and interest in multimodality:

the problems of selective observation are not eliminated with the use of recording equipment. They are simply delayed until the moment at which the researcher sits down to transcribe the material from the audio- or videotape (Ochs, 1979, 44).

Transcription entails decision-making and information loss. In making a “systematic attempt to ‘transcode’ the data” (Bateman et al., 2016, 135), the transcriber is inevitably shaping it. Some of both the deliberate and unintended consequences of this may be known, while others will not be. If we are to remain as neutral as possible with regard to the uses of the data we collect, it seems sensible to take steps to minimize the effect of this in so far as we can.

Ochs subsequently notes: “one of the consequences of ignoring transcription procedure is that researchers rarely produce a transcript that does reflect their research goals and the state of the field” (Ochs, 1979, 45). This observation resonates strongly today, as we see an emerging recognition of the potential of arranging data in ways such that it is of value across disciplinary boundaries:

the division between archives, databases and corpora is fast becoming one of disciplinary access and research methods rather than reflecting technical distinctions. But these differences in disciplinary usage can readily become misaligned with the functionalities that are actually required of the systems so described (Bateman et al., 2016, 131).

Clearly, then, we need to give active thought to the intended audience/user communities and applications/research questions for which datasets—and any tools to support their analysis—are intended. Focussing on material properties and leaving scope for humans to “imagine the relationships” (Bertin, 2010, xiv) can thus be seen as a strategic choice to maintain neutrality, rather than being merely expeditious in tactical terms.

At the same time, Ochs’s points about transcription are compounded in the multimodal analysis of documents by a number of factors: (1) the desire to avoid logocentrism; (2) the need somehow to accommodate intermodal relations, i.e., relations between elements of different semiotic modes; (3) the inherent two-dimensionality of the objects of study; (4) the problem of segmentation. While (1) might appear primarily to be a doctrinal concern, arguably the implications of logocentrism give rise to very real problems in relation to (2), (3), and (4). Moreover, points (3) and (4) are very closely inter-related. We will take each in turn below.

4.2 Logocentrism

Perhaps paradoxically, the issue of logocentrism is often raised by multimodalists whose formative academic experiences were within language studies and linguistics-

tics. For instance, Flewitt et al. (2009, 47) ask: “If [...] the aim in the analysis is not to prioritize spoken (or written) language, how can we transcribe in accessible, ‘readable’ ways?” (see also Thibault 2000 and Knight et al. 2009 for discussions on this theme).

In practical terms, this raises a question about for whom we are transcribing—should the transcription be human- or machine-readable? As Manovich points out:

Text is unique among media types. It plays a privileged role in computer culture. On the one hand, it is one media type among others. But, on the other hand, it is a metalanguage of computer media, a code in which all other media are represented: coordinates of 3-D objects, pixel values of digital images, the formatting of a page in HTML. (Manovich, 2001, 74)

Of course, even when formed of character strings taken from human writing systems, this metalanguage is not always human-readable—or, at least, meaningful to most of us. Indeed, as long as we have some means of validating the description, e.g., through representations of the results, there seems to be no need for it to be human-readable. In this sense, it is perhaps not so much logocentrism in terms of the form of transcription that is the heart of the issue.

4.3 Intermodal Relations

The handling of intermodal relations is an area of particular concern to those working with multimodal corpora. There often seems to be an assumption that the relations themselves should be encoded, for example:

It would be more useful [...] to create an alternative manual coding system that is transferable across the different data streams in the multimodal corpus in order to allow us to connect directly the pragmatic and semantic properties of the two modes and to enable cross referencing between the two. (Knight et al., 2009, 12)

This follows previous work by Baldry & Thibault (2001, 88–90), which identifies a need to move beyond the ‘intramodal’ perspective held hitherto in corpus linguistics toward an ‘intermodal’ orientation.

While this might sound attractive, I argue that such a move entails a number of questionable assumptions. Firstly, it becomes necessary to establish, a priori, which particular modes hold in a given body of data, which is problematic for the reasons discussed above. Secondly, it seems to assume that there is single formalism that can be applied across modes—or at least that the forms of representation for each mode are somehow mutually compatible. Thirdly, we need to posit the kinds of semantic relationships that may pertain between these modes, which in

turn implies that we need to anticipate the affordances of different modes. Fourthly, the move towards thinking ‘intermodally’ necessarily involves a combination of realization/form with semantics/function. Finally, it raises further questions about segmentation to which we will turn below.

While it shares the assumption of a single, encompassing formalism, as a representational form the *GeM* (Genre and Multimodality) model proposed by Bateman et al. (2004) has a fundamentally different architecture to those proposed by Knight et al. and Baldry and Thibault. Stand-off layers of XML annotation afford the analytical separation of features into orthogonal layers. Features of form (or realization) are separated from functional (or semantic) values, regardless of their status as semiotic modes. This offers crucial advantages in relation to the mitigation of the risk of analytical circularity. For example, it allows the investigation of the (different) resources deployed in the expression of messages with similar functions, the kinds of rhetorical relations that hold between different modes or the (lack of) correlation between physical (spatial) and semantic (conceptual) proximity (see Thomas, 2014). However, the labour-intensity of GeM annotation has limited its deployment in terms of scale and, as we shall see below, as currently specified, the GeM scheme has some limitations in relation to handling two-dimensionality.

4.4 Two-Dimensionality

The two-dimensionality of documents presents complex problems for multimodal analysis and transcription. Again, it is worthwhile here briefly considering recent work on this question. Kress (2009, 56) talks of the “‘logics’ of time and space”, noting that they “offer distinct potential for making meaning”. He goes on to explain that, while the “spatial display of writing gives rise to the sense that it works in some ways at least like an image”, “the reading of writing is governed by linearity and directionality” and “is not, dominantly and finally, organized by the logic of space” (Kress, 2009, 56).

While this again sounds plausible, the ways in which writing—and printed language—combines with other modes through layout is inherently two-dimensional—and any apparent linearity is open to question. As Waller has it:

At one level reading a page is a little like recognizing a face—you don’t inspect the eyes, the nose, and the mouth separately but in one take. This makes layout challenging for technologies or analytical frameworks that fail to go beyond the linear default. (Waller, 2012, 243)

Bateman outlines some of the consequences of the lack of linearity for multimodal analysis:

A substantial set of problems is raised by the fact that the object of study is not linear, either temporally or in terms of the principles for its consumption; moreover, its multichannel nature makes it difficult to reconcile and peg together the methods of recording, transcription, analysis and annotation that have been developed separately for each mode. (Bateman, 2008, 272)

The GeM scheme does not assume linearity in principle. Layout elements and blocks can be located in two-dimensional space and the base layer, which serves essentially to ‘peg together’ the other annotation layers, is conceived as a set of items, rather than a sequence of items.

The specification of grid-like layout structures is built into the GeM area model—suggesting that layout is decomposable into rectangular blocks. This is also common to much of the work on layout analysis from computer vision perspectives. Indeed, the grid is well-established as a design tool and still very much in use Waller (2017). Indeed, as Bateman (2008, 134–143) and Waller (2017, 185–188) provide compelling examples in which the grid underpins the placement of elements even where is an appearance of arbitrariness or “the superficial impression of freedom” (Bateman, 2008, 141).

While rectangular structures may still be prevalent in creative practice, the use of color, shape and placement often give the illusion of more varied and fluid layouts for the consumer. Furthermore, and as noted by Kumar et al. (2013), the rendering of digital content is generated by specific software on specific hardware. As such, imposing a static grid as a frame for analysis may be too reductive. In this regard, it should be noted that the GeM scheme is open to further development for the representation of structures of other kinds. Indeed, subsequent work which builds on GeM, such as Bateman et al. (2017) on comics, has developed such schemes (see also Stamenković and Jaćević in this volume).

Beyond specific questions about grids and rectangles, we could see this more fundamentally as a problem of segmentation—how do we get an analytical grip on the elements placed in two-dimensional space? Indeed, what are those elements and how would we know when we found them?

4.5 Segmentation

Previous work which has sought to adapt the GeM model in ways which make it feasible to implement at scale have highlighted some of the specific technical challenges involved in appropriate segmentation for source data which is not organized in linear fashion (see, e.g., Thomas, 2009b; Thomas et al., 2010).

From a conceptual perspective—and echoing Stöckl’s formulation, the “single unified gestalt in perception” (Stöckl, 2004, 16), as well as the need for meticulous

dissection in support of multimodal analysis—Waller points specifically to the tension between gestalt structures and markup:

Will computer pattern recognition ever be sophisticated enough to emulate the gestalt structures seen by human readers, to use physiognomic recognition, or to spot generic resonances—those visual, holistic features of a visual display that go beyond what is defined in the markup languages (such as XML) that lie behind many modern documents? (Waller, 2012, 243–244)

Thus, we come full circle, as Waller brings together the processes of production, perception and analysis—and, in so doing, raises a question as to whether a form of representation which has proven suitable for production and distribution is also appropriate as a format for storage of research data supportive of the kinds of analysis we envisage.

Indeed, this goes beyond questions about the current capabilities of ‘computer pattern recognition’ to core assumptions on which we base empirical approaches to multimodal analysis: the apparent tension between the gestalt structures perceived by readers and the analyst’s need for the separation into component parts. This in turn might be seen to derive from a logocentric assumption—that, like written and spoken language, all semiotic modes are realized through the combination of elements.

Manovich too notes the significance of this tension between the analytical desire for segmentation and the reality of specific media, such as photographs. He links this directly to the development of semiotics from Saussure through Barthes with characteristic pithiness: “The key assumption of modern semiotics is that communication requires discrete units. Without discrete units, there is no language” (Manovich, 2001, 28–29). In a rather more nuanced account of the foundations on which structuralist approaches are based, Bateman (2014, 32) points to “patterns of contrast that give rise to meaning”, which leads to questions about how such contrasts, or indeed their patternings, might be identified. And here Bateman emphasizes the need for analyses to be grounded both in the professional and the academic fields devoted to them—art history, graphic design and so on—as well as in the context of a specific text or artifact under analysis (Bateman, 2014, 46). For example, there seem to be qualitative differences between “notions of ‘form’, ‘masses’, ‘weight’, ‘balance’, ‘movement’, and so on from art criticism and the psychology of vision” on the one hand, and “the possibilities of lines and curves, of colours and shadings of black and white, of a few font sizes or interline-spacing, and so on that a linguist might come up with” on the other (Bateman, 2014, 46). Again, this suggests that different combinations of features are likely to be useful, or informative, for different analyses and different objects. Here, the notion of

dynamic derivation of elements of analysis is particularly interesting (see Bateman, 2014, 248).

In light of the transcription-related problems identified above—namely, segmentation and decomposability of images, (lack of) linearity, and the questionable appropriateness associated with the forms of reduction inherent in transcoding the visual as verbal—I argue here that we need to explore new forms of representation for graphic data.

In the next section, I propose an approach to handling multimodal data that draws on developments in computer vision, while sidestepping the need for formal segmentation and, with it, the assumptions on which this is necessarily based.

5 Proposed Approach

Essentially, the approach proposed here—for the automated collection and multimodal interrogation of documents—follows Occam’s razor, i.e., the principle that in explaining a thing no more assumptions should be made than are necessary. Thus, we do not assume the ways in which resources combine to make meanings; rather, we focus on the material properties of documents. In turn, this allows us to remain agnostic about the kinds of research questions and domains of application for which the approach might be useful. I present an example of a concrete use case for the approach towards the end of this section.

Having explored possibilities for scaling the GeM-based approaches to corpus development by integrating optical character recognition (OCR) into the process of data preparation (see, e.g., Thomas, 2007, 2014)—and having considered some of the problems associated with annotation—I have sought inspiration elsewhere. Recent work founded on principles from computer vision (e.g., Kumar et al., 2013; Reinecke et al., 2013) has developed methods for the interrogation of sets of graphic documents that do not rely on markup to encode two-dimensional ground data.

The approach proposed here does not rely on developing or implementing elaborate annotation that aims at a comprehensive representation of document form or content. While the design of the GeM scheme has several key advantages over other corpus-based approaches as outlined above—not least in its orthogonal layers and extensibility—schemes such as this are inevitably over-engineered for some purposes, while remaining inadequate for others. Furthermore, the manual labour associated with implementing such schemes inhibits their deployment at scale (Thomas, 2009b; Hiippala, 2015). Rather, I propose to collect and store images of scanned print documents or screenshots of rendered digital content and to assemble a toolkit which combines existing technologies that support ways into

specific questions pertinent to particular genres. Thus, the visual features that are deemed useful in a given case—and technologies required to handle them—will depend on the interests of the analyst. For example, text spotting might be needed for cases in which characters are embedded within pictures, or found on three-dimensional artifacts or scenes. In any case, I would restrict these features to properties which are material in nature. Mindful of the insights provided by Ochs (1979), in other words, I seek deliberately to delay interpretation as long as possible.

Specifically, such properties could include: the size, weight, and style of fonts used in strings of characters; the presence and distribution of colors across the document image; the presence of writing and embedded pictorial or schematic graphics. Rather than identifying a priori a set of properties deemed useful, this approach is founded on an acceptance that not all features would be present or relevant for all genres or for all questions. I would also capture the location of features that are recognized algorithmically within the two-dimensional space, in terms of x-y coordinates. Such features should be recoverable using existing techniques available off-the-shelf. Recent advances in convolutional neural networks (CNNs) have led to significant advances in image recognition and classification, as well as text spotting (e.g., Jaderberg et al., 2014). The novelty and value will be in combining these in ways that support multimodal document analysis.

As we move towards interrogation and analysis, the principle challenge will be how to go from this low, material level to higher levels of abstraction. This could be seen, in Peircean terms, as moving from regularities of form and perceptual qualities identified with a qualisign towards something that might in a given context be considered a candidate semiotic mode or conventionalized legisign (Bateman, 2018). Here we might usefully keep in mind two crucial questions recently posed by Bateman et al. (2016, 140): (1) “how to select automatic processing methods that are relevant for more abstract research questions”; (2) “how to relate the results of automatic processing to characterisations necessary for formulating more abstract research questions.” My point of departure is that it is easier to build upwards, by aggregating material features, than it is to disaggregate hypotheses post-hoc. This reflects the Peircean notion of prescinding or entailment between the categories of Thirdness, Secondness and Firstness (see Bateman, 2018). The approach also supports the understanding from Peirce that sign types operate simultaneously and that signs are “made by interpreters” (Bateman, 2018, 6).

There remain open many questions as to how precisely this bottom-up approach would be complemented by top-down interpretations, not least, to explore the respective roles of the human analyst, who designs the research questions and the integration of any feedback in CNNs that identify patterns. Low-level, material features of typographic realization—(relative) type size, weight, style,

color, alignment, etc.—might be seen to combine in conventional patterns to make recognizable structures of a second order of abstraction (‘itemized list’, ‘headline-newsbite’, ‘ragged-right aligned’, ‘table’, and so on). In turn, these typographic structures combine in a third order of abstraction with different types of non-verbal graphic/pictorial (photograph, painting, line drawing, pictograph/icon) and schematic (pie chart, line graph, scatter plot, histogram, flow chart). Could these third-order combinations help us to identify specific layout ‘modes’, in the conception by Bateman (2008) (i.e., text flow (as in classic novel), page flow (seen in many user manuals), image flow (as in some text books, encyclopedias and tourist guides—and often as components within page flow)?

Given the discussion of two-dimensionality above, two key methodological problems remain: (1) how to ‘peg together’ elements that differ in type; and (2) how to generalize across instances in order to identify patterns. In terms borrowed from corpus linguistics, how might we find multimodal collocations? What might a multimodal concordance look like (see for an early discussion Thomas, 2007)? In addressing these questions it is in the very two-dimensionality of the documents that we will look for answers.

In discussing “the levels of abstraction that might be appropriate for expressing spatial realizations for layout”, Bateman (2008, 221) explains “the human perceptual system is as unlikely to work with absolute values in this domain as in any other”. The same is true of our need to identify instances of phenomena them from sets of data. Qualitative spatial reasoning (QSR; Cohn & Renz, 2007) provides computational techniques by which spatial information is represented, not by numerical coordinates but by logical statements (such as, above, below, to the right of, aligned with, in between, touching, overlapping). Significantly, such relations are at a level and of a kind which humans often use when we communicate with one another about spatial information in everyday, informal, situations. These relations also allow us to remain agnostic about the presence or absence of an underlying layout structure, such as a grid.

While Bateman et al. (2017) report on the use of qualitative categories for the representation of comics, their scheme is genre-specific in its design. The proposal here is to make use of relationships of a higher level of abstraction, that are sufficiently general to apply and, crucially, to make comparisons across genres. In effect this explores the possible complementarities suggested in Bateman (2008, 223) between qualitative spatial representations of layout constraints as used in some automatic page generation systems and empirical approaches to genre. Figure 1 gives an illustrative example of the kind of claim or question the approach could be used to explore. It takes as its starting point a specific claim made elsewhere in the literature Waller et al. (2012) – i.e. that we have identified an instance of a document design pattern NEWS HEADLINES being transported from one genre

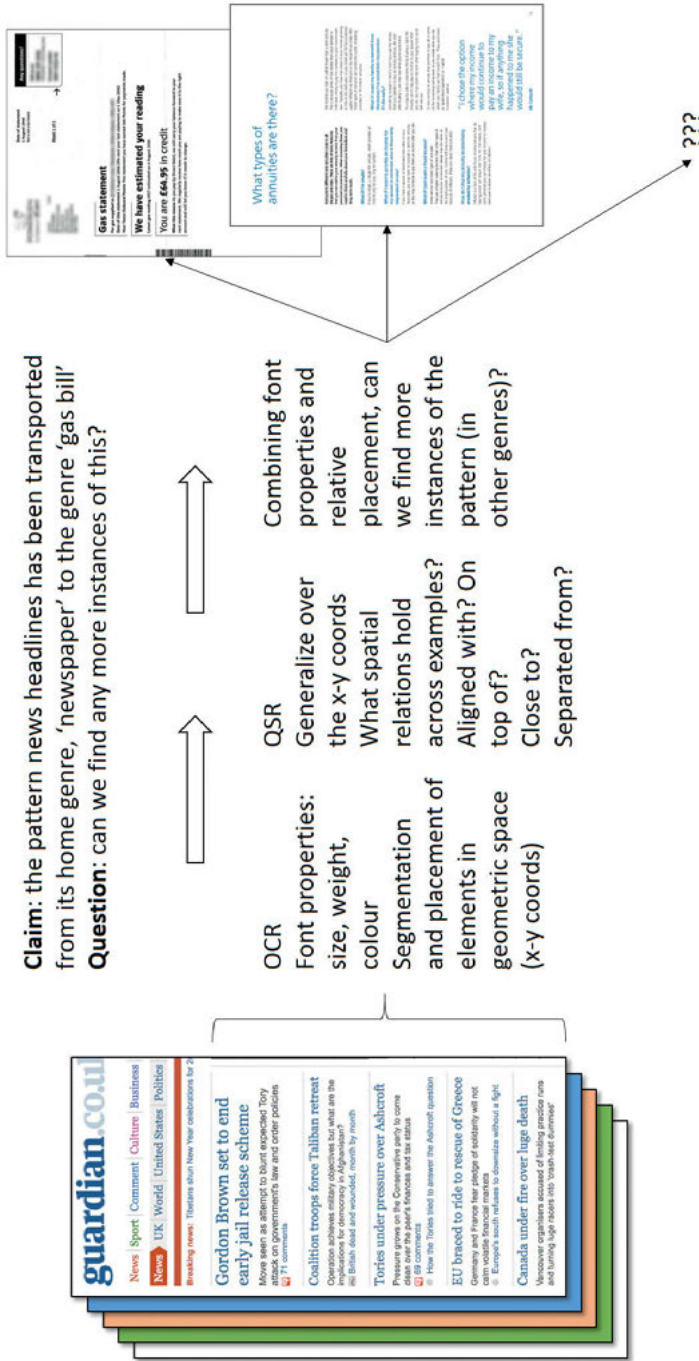
(online newspaper) to another (utility bill). We can imagine myriad analogous hypotheses around document design and genre variation, hybridization and shift, the validity and extent of which cannot be established without empirical testing. The approach identifies material typographic properties of the putative instance of the pattern – such as font size, color, and placement. These features are then modeled in relational or qualitative terms (for example, ‘repeated iterations of a string of characters in larger font appearing above and left-aligned-with a string in smaller font’). Instances of this pattern are then sought in a larger corpus of documents to support empirical investigation of its distribution across genre or other dimensions of variation.

In sum, then, the proposal here is to exploit the fact that space is the integrating principle in two-dimensional documents, which follows from observations about the contribution of layout to the textual metafunction (see Thomas, 2009a). More specifically, I propose that we explore the relational nature of QSR to support generalization of patterns of features across sets of visual features.

6 Conclusions

The widespread interest in and potential applications of multimodality justify, indeed require, greater rigor and investment of effort in developing robust conceptual frameworks and reliable methods. If multimodality is to gain traction in the real world, we cannot afford analytical circularity or fuzziness. In this context, I have identified an emerging consensus across different positions with multimodality around the need for quantitative methods, which might usefully complement qualitative and interpretive approaches. Specifically, I have aligned the hypothesis that we can exploit the modular structure and manipulability of digital media to build more or less abstract representations of their configurations as rendered in the wild with data which is readily available. I propose to exploit the fact that this data is not naturally organized in terms of semiotic mode to enable us to ask questions that do not assume the presence of particular modes. In this sense, the novelty of the approach taken here can be seen primarily in conceptual terms. By focussing on resources and their co-deployment by certain communities for particular purposes I seek to operationalize a social, fluid approach to semiotic mode that is empirically grounded.

While the digitalness of data in some ways makes harvesting and storage more straightforward, it does not of itself solve very concrete problems of representation and retrieval. Data is not tractable by current methods in its raw state. These problems remain to be explored, but I can outline the benefits of the approach



(Claim, Guardian screenshot and Powergen scan from Waller et al. 2012)

Fig. 1: Sketch of an approach to the exploration of a particular claim.

proposed here. At this time, the prioritization of material features—and building up from there—has intrinsic merit in logistical terms, as well as in terms of theory and principle. It is intellectually honest as well as pragmatic.

This approach offers both tactical and strategic advantages for multimodal analysis. In tactical terms, focussing on material values provides a way into multimodal description that is tractable in terms of existing techniques in computer vision, thus easing us through the bottle-neck identified in current work (e.g., Hiipala, 2015; O'Halloran et al., 2016). Strategically, the key virtue is that interpretive decisions are delayed. A number of methodological benefits follow from this: (1) we mitigate the risk of analytical circularity, while making the most of the “rather weak signal” identified by Bateman (2008, 13); (2) we operationalize a dynamic, social conception of semiotic mode, which is responsive to genres and communities of discourse, practice and analysis—and compare this across common sets of data; (3) we remain agnostic about the kinds of research questions, disciplinary assumptions and interests that people using the framework might bring. In terms of theoretical development, the approach provides a means through which to explore empirically the reach and robustness of existing conceptual frameworks when applied to data sets controlled for genre or context of production/consumption. Thus, we can establish their reliability and nuance their claims.

In terms of its contribution to multimodality, the approach outlined here remains empirically grounded, while offering the potential to scale beyond that which we have achieved as a community so far. Whether or not we identify as a discipline, these two features seem to be necessary conditions for using multimodality to approach, and suggesting reliable answers to, questions involving communication and meaning making, particularly in domains where the real-world consequences are significant.

Acknowledgements

I would like to thank John Stell for recent conversations around Qualitative Spatial Representation which have inspired some of the ideas presented here. I am also grateful to the editorial team for their critical engagement with the version originally submitted and their suggestions for its improvement. Any remaining errors or misrepresentations are mine.

Bibliography

Adami, Elisabetta. 2015. Aesthetics in Digital Texts beyond Writing: A Social Semiotic Multimodal Framework. In Arlene Archer & Esther Breuer (eds.), *Multimodality in Writ-*

- ing: *The State of the Art in Theory, Methodology and Pedagogy*, 43–62. Leiden: Brill.
https://doi.org/10.1163/9789004297197_004.
- Baldry, Anthony P. & Paul Thibault. 2001. Towards Multimodal Corpora. In Guy Aston & Lou Burnard (eds.), *Corpora in the Description and Teaching of English*, 87–102. Bologna: CLUEB.
- Bateman, John A. 2008. *Multimodality and Genre: A Foundation for the Systematic Analysis of Multimodal Documents*. Basingstoke: Palgrave Macmillan.
- Bateman, John A. 2011. The Decomposability of Semiotic Modes. In Kay L. O'Halloran & Bradley A. Smith (eds.), *Multimodal Studies: Multiple Approaches and Domains*, 17–38. London: Routledge.
- Bateman, John A. 2013. Book Review: Theo van Leeuwen, *The Language of Colour: An Introduction*. *Linguistics and the Human Sciences* 9(1). 1–8. <https://doi.org/10.1558/6545748635>.
- Bateman, John A. 2014. *Text and Image: A Critical Introduction to the Visual/Verbal Divide*. London and New York: Routledge.
- Bateman, John A. 2018. Peircean Semiotics and Multimodality: Towards a New Synthesis. *Multimodal Communication* 7(1). <https://doi.org/10.1515/mc-2017-0021>.
- Bateman, John A., Judy L. Delin & Renate Henschel. 2004. Multimodality and Empiricism: Preparing for a Corpus-based Approach to the Study of Multimodal Meaning-Making. In Eija Ventola, Cassily Charles & Martin Kaltenbacher (eds.), *Perspectives on Multimodality*, 65–87. Amsterdam: John Benjamins.
- Bateman, John A., Chiao-I Tseng, Ognyan Seizov, Arne Jacobs, Andree Lüdtke, Marion G. Müller & Otthein Herzog. 2016. Towards Next-Generation Visual Archives: Image, Film and Discourse. *Visual Studies* 31(2). 131–154. <https://doi.org/10.1080/1472586X.2016.1173892>.
- Bateman, John A., Francisco O.D. Veloso, Janina Wildfeuer, Felix HiuLaam Cheung & Nancy Songdan Guo. 2017. An Open Multilevel Classification Scheme for the Visual Layout of Comics and Graphic Novels: Motivation and Design. *Journal of Digital Scholarship in the Humanities* 32(3). 476–510. <https://doi.org/10.1093/llc/fqw024>.
- Bertin, Jacques. 2010. *Semiology of Graphics: Diagrams, Networks, Maps*. California: Redlands. Translated by William J. Berg. Originally published in French 1967.
- Cohn, Anthony G & Jochen Renz. 2007. Qualitative Spatial Representation and Reasoning. In Frank van Harmelen, Vladimir Lifschitz & Bruce Porter (eds.), *Handbook of Knowledge Representation*, 551–596. Oxford: Elsevier.
- Early, Margaret, Maureen Kendrick & Diane Potts. 2015. Special Issue: Multimodality: Out From the Margins of English Language Teaching. *TESOL Quarterly* 49(3).
- Flewitt, Rosie, Regine Hampel, Mirjam Hauck & Lesley Lancaster. 2009. What are Multimodal Data and Transcription? In Carey Jewitt (ed.), *The Routledge Handbook of Multimodal Analysis*, 40–53. London: Routledge.
- Forceville, Charles J. 1999. Educating the Eye? Kress and Van Leeuwen's *Reading Images: The Grammar of Visual Design* (1996). *Language and Literature* 8(2). 163–178. <https://doi.org/10.1177/096394709900800204>.
- Forceville, Charles J. 2007. Book Review: *Multimodal Transcription and Text Analysis: A Multimedia Toolkit and Coursebook* by Anthony Baldry and Paul J. Thibault. *Journal of Pragmatics* 39(6). 1235–1238.
- Halliday, Michael A. K. 1973. *Explorations in the Functions of Language*. London: Edward Arnold.
- Halliday, Michael A. K. 1978. *Language as Social Semiotic*. London: Edward Arnold.
- Hiiippala, Tuomo. 2015. *The Structure of Multimodal Documents: An Empirical Approach*. London: Routledge.

- Jaderberg, Max, Andrea Vedaldi & Andrew Zisserman. 2014. Deep Features for Text Spotting. In *Proceedings of the 13th European Conference on Computer Vision (ECCV 2014)*, 512–528.
- Jewitt, Carey. 2009. An Introduction to Multimodality. In Carey Jewitt (ed.), *The Routledge Handbook of Multimodality*, 14–29. London, New York: Routledge.
- Knight, Dawn, David Evans, Ronald Carter & Svenja Adolphs. 2009. HeadTalk, HandTalk and the Corpus: Towards a Framework for Multimodal, Multimedia Corpus Development. *Corpora* 4(1). 1–32.
- Kress, Gunther. 2009. What is Mode? In Carey Jewitt (ed.), *The Routledge Handbook of Multimodal Analysis*, 54–67. London: Routledge.
- Kress, Gunther. 2010. *Multimodality: A Social Semiotic Approach to Contemporary Communication*. London: Routledge.
- Kress, Gunther, Carey Jewitt, Jon Ogborn & Charalampos Tsatsarelis. 2000. *Multimodal Teaching and Learning. The Rhetorics of the Science Classroom*. London: Continuum.
- Kress, Gunther & Theo van Leeuwen. 2001. *Multimodal Discourse. The Modes and Media of Contemporary Communication*. London: Arnold.
- Kress, Gunther & Theo van Leeuwen. 1996. *Reading Images: The Grammar of Visual Design*. London and New York: Routledge.
- Kumar, Ranjitha, Arvind Satyanarayan, Cesar Torres, Maxine Lim, Salman Ahmad, Scott R. Klemmer & Jerry O. Talton. 2013. Webzeitgeist: Design Mining the Web. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2013)*, 3083–3092. Paris.
- Lemke, Jay L. 1998. Multiplying Meaning: Visual and Verbal Semiotics in Scientific Text. In James R. Martin & Robert Veel (eds.), *Reading Science: Critical and Functional Perspectives on Discourses of Science*, 87–113. London: Routledge.
- Manovich, Lev. 2001. *The Language of the New Media*. Cambridge, MA: MIT Press.
- Manovich, Lev. 2009. Cultural Analytics: Visualizing Cultural Patterns in the Era of ‘More Media’. *DOMUS* 923. 112–115.
- Moretti, Franco. 2013. *Distant reading*. London and New York: Verso.
- Ochs, Elinor. 1979. Transcription as Theory. In Elinor Ochs & Bambi B. Schieffelin (eds.), *Developmental Pragmatics*, 43–72. New York: Academic Press.
- O’Halloran, Kay L. 2008. Systemic Functional-Multimodal Discourse Analysis (SF-MDA): Constructing Ideational Meaning using Language and Visual Imagery. *Visual Communication* 7(4). 443–475.
- O’Halloran, Kay L., Sabine Tan, Duc-Son Pham, John A. Bateman & Andrew Vande Moere. 2016. A Digital Mixed Methods Research Design: Integrating Multimodal Analysis with Data Mining and Information Visualization for Big Data Analytics. *Journal of Mixed Methods Research* 12(1). 11–30. <https://doi.org/10.1177/1558689816651015>.
- Pauwels, Luc. 2012. A Multimodal Framework for Analyzing Websites as Cultural Expressions. *Journal of Computer-Mediated Communication* 17. 247–265.
- Peréz-González, Luis. 2014. Multimodality in Translation and Interpreting Studies. In Sandra Bermann & Catherine Porter (eds.), *A Companion to Translation Studies*, 119–131. Chichester: Wiley-Blackwell.
- Ramos Pinto, Sara & Yves Gambier. 2018. *Audiovisual Translation: Theoretical and Methodological Challenges*. Amsterdam: John Benjamins.
- Reinecke, Katharina & Krzysztof Z. Gajos. 2014. Quantifying Visual Preferences Around the World. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI’14)*, 11–20. Toronto.

- Reinecke, Katharina, Tom Yeh, Luke Miratrix, Rahmatri Mardiko, Yuechen Zhao, Jenny Liu & Krzysztof Gajos. 2013. Predicting Users' First Impressions of Website Aesthetics With a Quantification of Perceived Visual Complexity and Colorfulness. In *Proceedings of the sigchi conference on human factors in computing systems (ch' 13)*, Paris.
- Shah, Kashif, Josiah Wang & Lucia Specia. 2016. SHEF-Multimodal: Grounding Machine Translation on Images. In *Proceedings of the First Conference on Machine Translation, Volume 2: Shared Task Papers*, 660–665. Berlin: Association for Computational Linguistics.
- Stöckl, Hartmut. 2004. In Between Modes: Language and Image in Printed Media. In Eija Ventola, Cassily Charles & Martin Kaltenbacher (eds.), *Perspectives on Multimodality*, 9–30. Amsterdam: John Benjamins.
- Thibault, Paul J. 2000. The Multimodal Transcription of a Television Advertisement: Theory and Practice. In Anthony P. Baldry (ed.), *Multimodality and Multimediality in the Distance Learning Age*, 311–385. Campobasso, Italy: Palladino Editore.
- Thomas, Martin. 2007. Querying Multimodal Annotation: A Concordancer for GeM. In *Proceedings of the Linguistic Annotation Workshop*, 57–60. Prague: Association for Computational Linguistics.
- Thomas, Martin. 2009a. Developing Multimodal Texture. In Eija Ventola & Arsenio J.M. Guijarro (eds.), *The World Shown and the World Told: Multisemiotic Issues*, 39–55. Houndmills: Palgrave Macmillan.
- Thomas, Martin. 2009b. *Localizing Pack Messages: A Framework for Corpus-Based Cross-Cultural Multimodal Analysis*. Leeds: Centre for Translation Studies, University of Leeds PhD Thesis.
- Thomas, Martin. 2014. Evidence and Circularity in Multimodal Discourse Analysis. *Visual Communication* 13(2). 163–189. <https://doi.org/10.1177/1470357213516725>.
- Thomas, Martin, Judy Delin & Robert H. W. Waller. 2010. A Framework for Corpus-Based Analysis of the Graphic Signalling of Discourse Structure. In *Proceedings of Multidisciplinary Approaches to Discourse (MAD 2010)*, Moissac, France.
- Waller, Rob, Judy Delin & Martin Thomas. 2012. Towards a Pattern Language Approach to Document Description. *Discours: Revue de linguistique, psycholinguistique et informatique* 10. <https://doi.org/10.4000/discours.8673>.
- Waller, Robert. 2012. Graphic Literacies for a Digital Age: The Survival of Layout. *The Information Society: An International Journal* 28(4). 236–252.
- Waller, Robert. 2017. Practice-Based Perspectives on Multimodal Documents: Corpora vs Connoisseurship. *Discourse, Context and Media* 20. 175–190.

Kay L. O'Halloran, Sabine Tan, Peter Wignell, Rui Wang, Kevin Chai, and Rebecca Lange

Towards a Discipline of Multimodality: Parallels to Mathematics and Linguistics and New Ways Forward

Abstract: Multimodality, involving the study of meaning arising from the integration of language with images and other resources in multimodal texts, interactions and events, addresses the fundamental need to understand human communication in the current age of digital technology. However, multimodality is not considered to be a discipline per se at present. By drawing parallels between mathematics and linguistics, it is proposed that if multimodality is to become a discipline, then abstract context-based frameworks for modeling multimodal semiotic resources and methodologies for investigating patterns of human communication are required. An example of how this could be achieved is provided. From here, multimodality has the potential to provide the foundations for a range of multimodal sciences, in much the same way that mathematics and linguistics underpin the mathematical and language sciences respectively. In doing so, it may become possible to track the changes in human communication arising from digital technology and the resultant impact on thought and reality.

Keywords: multimodality, discipline, mathematics, mapping, computational tools, context

1 Introduction

A discipline is generally understood to be an organized and systematic body of knowledge that is typically studied at university level. For example, the Oxford English Dictionary defines 'discipline' as "a branch of knowledge, typically one studied in higher education".¹ Much of multimodality as it is researched today evolved from branches of linguistics and social semiotics in the 1990s (e.g., Tan et al., 2019) as a means for studying the ways in which human beings use a whole range of different semiotic resources (including language, sound, gesture, images, and so forth) for meaning making and communication.

¹ <https://en.oxforddictionaries.com/definition/discipline>, last accessed: 29 August 2019.

Today, multimodality can be viewed beneficially in the same light as mathematics and linguistics; namely, as a field that addresses a fundamental problem in contemporary society, in this case the need to understand human communication, particularly in the current age of digital technology (i.e., Internet, social media, and mobile devices). At the present time, multimodality is studied in various courses and has increasingly become the focus of postgraduate research. Despite these advances, however, multimodality is not considered to be a discipline per se at this stage. If multimodality is to become a discipline, it is proposed that generalizable, abstract context-based frameworks for modeling multimodal semiotic resources and analytical methods for investigating patterns of human communication over space and time are required.

In what follows, the ways in which multimodality may become a discipline in the future are explored by drawing parallels with mathematics and linguistics. These two fields developed in order to address key issues in the world—i.e., the modeling of the material world and the human world of language—and in doing so, developed abstract and generalized knowledge that could be applied in different real-life contexts. Firstly, mathematics flourished during the Renaissance when mathematical innovations in the form of abstract structures were linked to scientific discoveries. From there, mathematics became the science of number, quantity, and space for modeling and predicting the material world, giving rise to a range of mathematical sciences.

Secondly, modern linguistics turned to the study of grammatical systems in the 20th century to address key questions about human language in terms of language change, language structure, and language use. These developments gave rise to different language sciences informed by various branches of linguistics. From there, the two fields developed as disciplines that underpin a range of mathematical and language sciences respectively, as illustrated below. In the following sections, we discuss mathematics and linguistics respectively, before turning to multimodality.

2 Mathematics: Mapping the Physical World

Mathematics developed relatively slowly until mathematical innovations were linked to scientific discoveries in the Renaissance (e.g., Eves, 1990). Galileo (1623 [1957]), as a pioneer of the scientific method, understood the significance of mathematics for advancing science:

The universe cannot be read until we have learned the language and become familiar with the characters in which it is written. It is written in mathematical language, and the letters are triangles, circles and other geometrical figures, without which means it is humanly impossible

to comprehend a single word. Without these, one is wandering about in a dark labyrinth. (Galileo, 1623 [1957], *Opere Il Saggiatore*)

Today, mathematics is described as “the abstract science of number, quantity, and space, either as abstract concepts (pure mathematics), or as applied to other disciplines such as physics and engineering (applied mathematics)”.² From this perspective, pure mathematics is viewed as mathematics for its own sake without any pre-determined applications, although applications are often found later on. On the other hand, applied mathematics is designed to solve specific problems, sometimes leading to new fields of mathematics (e.g., statistics and game theory). The simple division of mathematics into two categories is seen to create barriers, however, whereas in reality there are many commonalities across the mathematical sciences (National Science Council, 2013).

Whichever way mathematics is considered, it is a human construction developed for certain purposes. As Kline (1980, 312) explains: “What then is mathematics if it is not a unique, rigorous, logical structure? It is a series of great intuitions carefully sifted, and organized by the logic men [sic] are willing and able to apply at any time”. In other words, mathematics is “a human construction with all that implies” (Little, 1981, 159). Although mathematics is not an empirical science, many of the ideas originate in empirical results from which further concepts and areas are developed. As Neumann explains:

I think that it is a relatively good approximation to truth—which is much too complicated to allow anything but approximations—that mathematical ideas originate in empirics, although the genealogy is sometimes long and obscure. But, once they are so conceived, the subject begins to live a peculiar life of its own and is better compared to a creative one, governed by almost entirely aesthetical motivations, than to anything else and, in particular, to an empirical science. (Neumann, 1956, 2063)

The National Science Council (2013) defines the mathematical sciences in broad terms: namely, those areas which “aim to understand the world by performing formal symbolic reasoning and computation on abstract structures” (National Science Council, 2013, 62).

The various areas of the mathematical sciences are displayed in Figure 1. This includes the traditional areas (e.g., engineering, economics, computer science, geoscience, astronomy, physics, chemistry, and biology) and areas that are concerned with building mathematical models and exploring them computationally through the analysis of datasets (e.g., medicine, social networks, information processing, communications, defense, manufacturing, marketing, and finance). All

² <https://en.oxforddictionaries.com/definition/mathematics>, last accessed: 29 August 2019

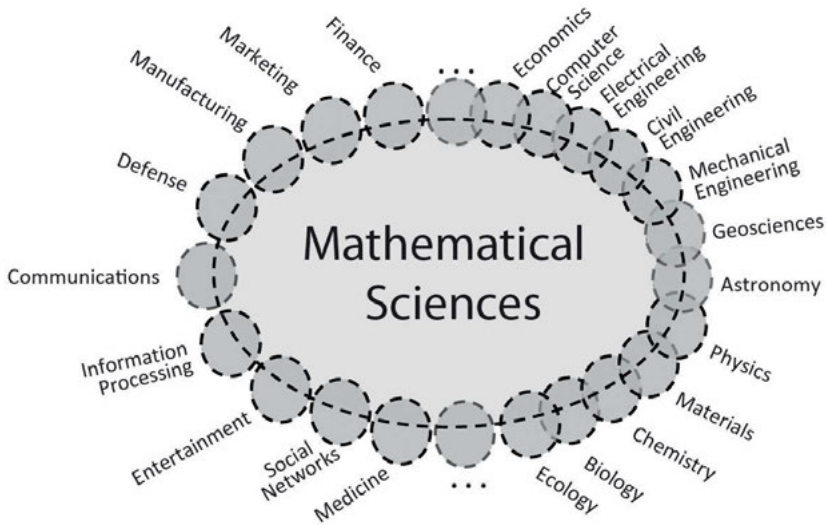


Fig. 1: The mathematical sciences (National Science Council, 2013, 63).

these fields are mathematical in nature, regardless if they are part of computer science or part of the discipline for which the modeling or analysis are performed. The various activities in the mathematical sciences aim to:

- discover relationships between the abstract structures;
- capture features of the world in abstract structures through modeling and formal reasoning or by using abstract structures as a framework for computation to make predictions about the world;
- use abstract reasoning, models, and structures to make inferences about the world through data science.

As the National Science Council explains, these activities are “linked to the quest to find ways to turn empirical observations into a means to classify, order, and understand reality—the basic promise of science” (National Science Council, 2013, 62).

In a similar fashion, multimodality needs to be seen as originating in empirical results from which abstract concepts and ideas are formed in order to classify, order and understand human (rather than physical) reality (Bateman, 2014a; O’Halloran et al., 2016; Tan et al., 2018). In this regard, we can view multimodality as a sci-

ence with the potential to be applied to other areas, henceforth referred to as *the multimodal sciences*. Before exploring these propositions further, we first consider linguistics as a discipline which shares similarities with mathematics and multimodality.

3 Linguistics: Mapping Human Language

Language has been an object of enquiry since antiquity; for example, logic, rhetoric, and grammar were studied in ancient Greece. Modern linguistics is the scientific study of language (Halliday, 2003)³ which aims to answer key questions about human language in terms of language change, language structure, and language use. Linguistics, like mathematics, can be grouped into two main areas: pure (or theoretical, or general) linguistics, and applied linguistics, with various subfields in each category. However, many branches of linguistics do not fit easily into either category, given that they are concerned with developing theory in order to understand how language is used (e.g., systemic-functional linguistics, psycholinguistics, sociolinguistics, and computational linguistics).

Following Halliday (1978), linguistics is primarily concerned with ‘language as system’ in terms of substance (phonic or graphic) and form (vocabulary, grammar and semantics), as shown in Figure 2 (see central triangle). In addition, linguistics is concerned with the study of ‘language as behavior’ (e.g., socialization and sociolinguistics), ‘language as knowledge’ (psycholinguistics), and ‘language as art’ (e.g., literary studies) (see Figure 2). The different areas of linguistics are related to other disciplines: for example, sociology, psychology, literature, and physics and physiology. Beyond this, linguistics is involved in other areas such as archeology, philosophy, logic and mathematics, communications engineering, culture, social anthropology, for example.

Linguistics mirrors mathematics in that core areas provide the basis for other areas of study, in this case, the language sciences. In order for this to occur, it was necessary to develop abstract structures to explain how language works as a system. For example, Halliday (1973) describes the grammatical systems through which language fulfills certain functions (see Figure 3), which he later developed into a comprehensive lexicogrammar of English (e.g., Halliday & Matthiessen, 2014). In Halliday’s model, the grammatical systems are organized into ranks according to three metafunctions: (a) *ideational meaning* consisting of experiential meanings to capture happenings in the world, and logical meaning to capture

³ <https://en.wikipedia.org/wiki/Linguistics>, last accessed: 29 August 2019

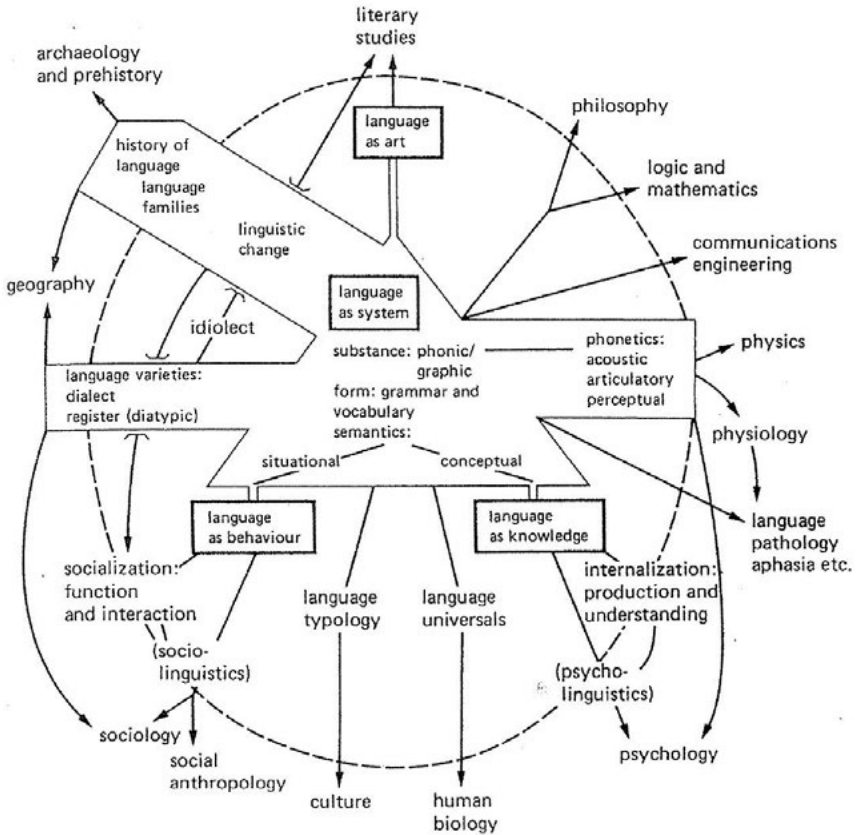


Fig. 2: Domains of language studies and their relations to other fields (Halliday, 1978, 11).

the logical relations between those happenings; (b) *interpersonal meaning* to map social interactions and relations; and (c) *textual meaning* to organize the message. The grammatical systems are used to map choices in linguistic texts and to examine relations between system choices according to roles which language is playing in different contexts.

The meaning of human language is dependent on context, however, so that the mapping between language choices and meaning is not straightforward. For this reason, it has not been possible to implement computational models for the full range of linguistic systems (e.g., see discussion in Bateman & O'Donnell, 2015). Indeed, most natural languageprocessing algorithms focus on lexical items, rather than grammatical systems which underlie the functional organization of language. However, the potential for modeling context has been greatly enhanced in the

rank		metafunction		ideational		experiential		interpersonal		textual	
		[class]	[process]	logical		(process type)		MOOD MODALITY POLARITY	THEME CULMINATION VOICE	(cohesive)	
clause	[complexe]										
	[prepositional]										
phrase	[verbal]										
	[nominal]										
group	[complexe]										
	[prepositional]										
word	[verbal]										
	[nominal]										
information unit	[complexe]										
	[prepositional]										

A "function / rank matrix" for the grammar of English, where (i) rows show rank and primary class, (ii) columns show metafunction, and (iii) capitals show system(s) in each cell

Fig. 3: Halliday's function /rank matrix for the grammar of English (Halliday, 1973, 141).

digital age, due to the large datasets of text, images, and videos which are now available with metadata (e.g., URLs, date of postings, source materials, and references). In addition, semantic categorizations of (nearly) every domain of human activity are now available through socially evolved knowledge classification systems such as Wikipedia. Therefore, it is possible to analyze human communication in relation to context in new ways that have not existed before, given the metadata for large datasets of multimodal texts which is available today.

For this reason, we propose that multimodality is poised ready to follow a similar trajectory to mathematics and linguistics in terms of developing generalizable, abstract structures that can be applied to different contexts of communication in the real world. If this does occur, then multimodality will become a scientific discipline, giving rise to the multimodal sciences for modeling and mapping the human universe. The multimodal data is available now, but we do not have the necessary abstract models and methodologies yet. In what follows, a possible path forward in this direction is discussed.

4 Multimodality: Mapping the Human Universe

As mentioned above, much of modern multimodality originated in linguistics, particularly in social semiotics and systemic-functional linguistics (Bateman et al., 2017; Jewitt, 2014; Jewitt et al., 2016; Tan et al., 2019). Multimodality is concerned with the entire range of semiotic resources which humans use for meaning making, including language, image, symbolism, gaze, gesture, space, architecture, and so forth. In particular, multimodality is concerned with the integration of language with other systems of meaning and mapping the interaction of semiotic choices in texts, interactions and events in different contexts. Tan et al. (2019) provide a comprehensive account of recent theoretical, methodological, and analytical trends in multimodality, and this review is not repeated here. Rather, the focus of this discussion is how multimodality may become a science which provides the foundations for other fields of study, with the leading question: “what are the requirements for multimodality in terms of mapping the human world?”

Mathematics succeeded by providing semiotic tools for formulating abstract structures, which could be used for modeling and formal reasoning, and as frameworks for computation to make predictions about the physical world. The interrelations between concepts, systems, and processes are made explicit, reasoning and computation are made as efficient as possible, and the limits of the findings are characterized. Moreover, the abstract structures hold, regardless of the context in

which they are applied and used in the physical world. The definition of ‘abstract structure’ makes this clear:

An abstract structure may be represented (perhaps with some degree of approximation) by one or more physical objects—this is called an implementation or instantiation of the abstract structure. *But the abstract structure itself is defined in a way that is not dependent on the properties of any particular implementation.* [emphasis added]⁴

Similarly, (Halliday, 2008, 7) views systemic-functional linguistics as an “applicable science” with a “comprehensive and theoretically powerful model of language” designed to address problems associated with language use. The systemic-functional model of language incorporates context (i.e., the context of situation and the context of culture derived from Malinowski (1923)) but it has not been possible to formalize contextual parameters to the same extent as the lexicogrammar as yet. As a result, instances of language use could not be fully accounted for, given that the meanings of linguistic choices arise from their context of use.

Nonetheless, Halliday’s systemic-functional model of language is a comprehensive description of how language is organized to create meaning (as a system of meanings), and how these systems are activated to fulfill certain functions in relation to context. Significantly, the basic principles of language as a social semiotic system can also be applied to images, videos, and other resources, resulting in frameworks with common theoretical concepts of metafunctions, systems, and ranks (e.g., see Figure 4), despite the different resources which are involved. This provides a common foundation upon which to model and analyze different semiotic resources in terms of their underlying organization in the form of metafunctionally based systems, organized according to different ranks, as displayed in Figure 4.

Following Halliday (2008), multimodality is conceptualized an “applicable science” that is designed to address problems associated with the use of language, images, and other semiotic resources. Furthermore, it is proposed that *abstract context-based models of semiotic resources* and *semiotic interactions* are required in order to map patterns of meaning in human communication, and to trace those patterns over space and time. These abstract models are designed to

- discover relations between semiotic resources;
- map patterns of semiotic choices in texts, interactions, and events;
- provide an overarching framework for computational models for mapping patterns and making predictions about the human world. This includes making inferences through reasoning, and models and structures using data science.

⁴ https://en.wikipedia.org/wiki/Abstract_structure, last accessed: 29 August 2019.

Genre			Register (Field, Tenor and Mode)		
Language			Text/Image		
Metafunction	Rank	System	Description		
Experiential	Clause	Processes; Participant Roles; Circumstance	Happenings, actions and relations		
Logical	Clause Complex	Logico-Semantic Relations	Relations between happenings, actions and relations		
Interpersonal	Discourse Clause	Appraisal Speech Function	Evaluation in terms of attitude, emotion and judgment Exchange of information (e.g. statements and questions) and goods & services (e.g. commands and offers)		
Textual	Clause Discourse	Information Focus	Organisation of information, with points of departure for what follows		
Images			Text/Image relations across metafunctions and ranks		
Metafunction	Rank	System	Description		
Experiential	Work Episode	Narrative Theme; Representation; Setting Processes; Participant Roles; and Circumstance	Nature of the scene Visual happenings, actions and relations		
Logical	Figure Work Episode	Posture; Dress Logical Relations Logical Connections	Characteristics of the participants Relations between process and participant configurations (e.g. temporal, spatial, causal)		
Interpersonal	Work Episode	Angle; Camera Distance; Lighting Proportion in Relation to the Whole Image; Focus; Perspective	Relations between parts of figures and objects Visual effects Happenings, actions and relations with respect to the whole image		
Textual	Figure Work Episode	Gaze-Visual Address Compositional Vectors; Framing Relative Placement of Episode; Framing	Direction of participant's gaze as internal to image or external to viewer The organisation of the parts as a whole, with the visual marking (e.g. framing) of certain parts Position of the happenings, actions and relations in relation to the whole image, and the visual marking of certain aspects		
	Figure	Relative Placement of the Figure within the Episode; Arrangement; Framing	Position of figures in relation to happenings, actions or relations, and the visual marking of certain aspects of those figures		

Fig. 4: Language and images (O'Halloran et al., 2016).

The scientific view of multimodality mirrors that of mathematics (for example, as formulated by the USUS National Science Council (2013, 62)) and linguistics (for example, as developed by Michael Halliday) with the goal of mapping the human world that includes and extends beyond language. Rephrasing the National Science Council (2013, 62), “this is linked to the quest to find ways to turn empirical observations into a means to classify, order, and understand *human* reality—the basic promise of *multimodality*”. However, there are two major problems: first, mathematics deals with abstract structures which are independent of applications in the real world (unlike language and multimodality which are context-dependent), and secondly, semiotic formulations beyond language are required in this model. These issues are discussed below.

5 Multimodality: Context and Semiotic Resources Beyond Language

Mathematics is the study of abstract structures, which are defined by laws, properties, and relationships which hold, regardless of the context. That is, the abstract structures can be represented in the physical world but the abstract structures themselves are independent of the properties of any particular instantiation. Computer language, for example, is considered to be an abstract structure because it can be implemented with the same result in any context, but natural language is not generally perceived to be an abstract structure because it can be used with different results according to the context of use. For example, “I like it” can mean different things (i.e., ‘I really do like it’, or ‘I really don’t like it’) according to how it is said and/or written and the context of the use. This same argument applies to multimodal texts, but the problem is exacerbated because the meaning arises from combinations of interacting semiotic choices which are interpreted in relation to the context.

Therefore, in order for language, images, and other resources and semiotic interactions to be modeled as abstract structures, these abstract structures need to incorporate context to account for the variations in meaning which occur in the instantiations of multimodal choices. But how can this be done? The problem is foregrounded in context-enhanced information fusion (Snidaro et al., 2016) where context is taken into account at different levels of abstraction; for example, the low-level data, feature extraction, patterns between features, and decisions and relationships and high level descriptions, as displayed in Figure 5. Furthermore, horizontal and vertical heterogeneity are incorporated in the model, as displayed in Figure 5. In what follows, we discuss possible approaches to modeling and

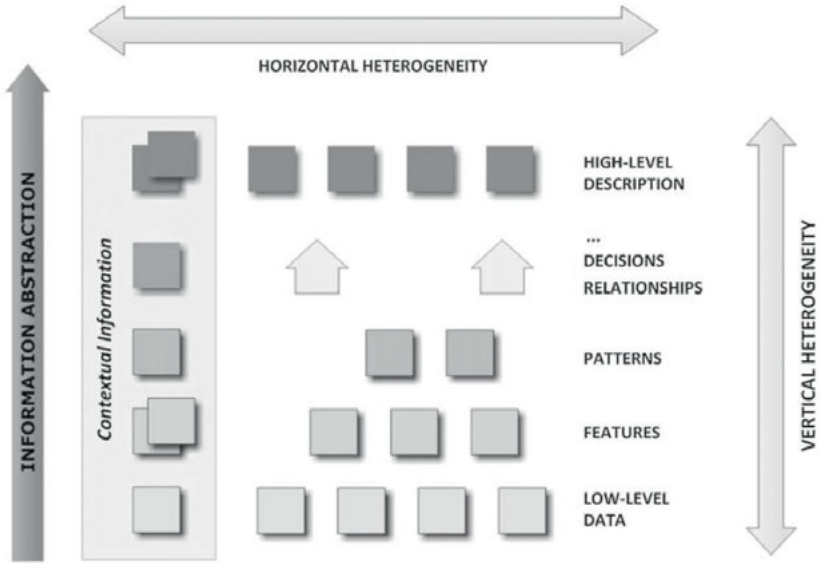


Fig. 5: Context-enhanced information fusion (Snidaro et al., 2016, 434).

analyzing low level data, features, and patterns, while recognizing that decisions regarding relationships and high-level descriptions involve further contextual parameters (e.g., decisions involving the emerging patterns and high-level descriptions, such as prediction).

One possible way forward is to use computational models for language, images, videos, and other resources because these models are already formulated in terms of abstract structures. However, computational models were typically developed for one resource: for example, natural language processing, image processing, and video processing. Moreover, these computational approaches identify low-level data features and use machine learning (e.g., neural networks) to identify lexical items in written texts and objects and events in images and videos.

Recent developments in information fusion aim to combine different modalities (e.g., Arevalo et al., 2017; Kiela et al., 2018). However, three major challenges exist: namely, “feature learning and extraction, modeling of relationships between data modalities and scalability to large multimodal collections” (Arevalo, 2018, 1). Given this situation, multimodality offers an exciting opportunity to contribute to this field, as evidenced by work which is underway (see overview in Bateman et al., 2019). In what follows, we describe how various computational tools can

be integrated into a multimodal framework for big dataanalytics of multimodal communications.

6 Integrating Computational Tools with Multimodal Theory and Context

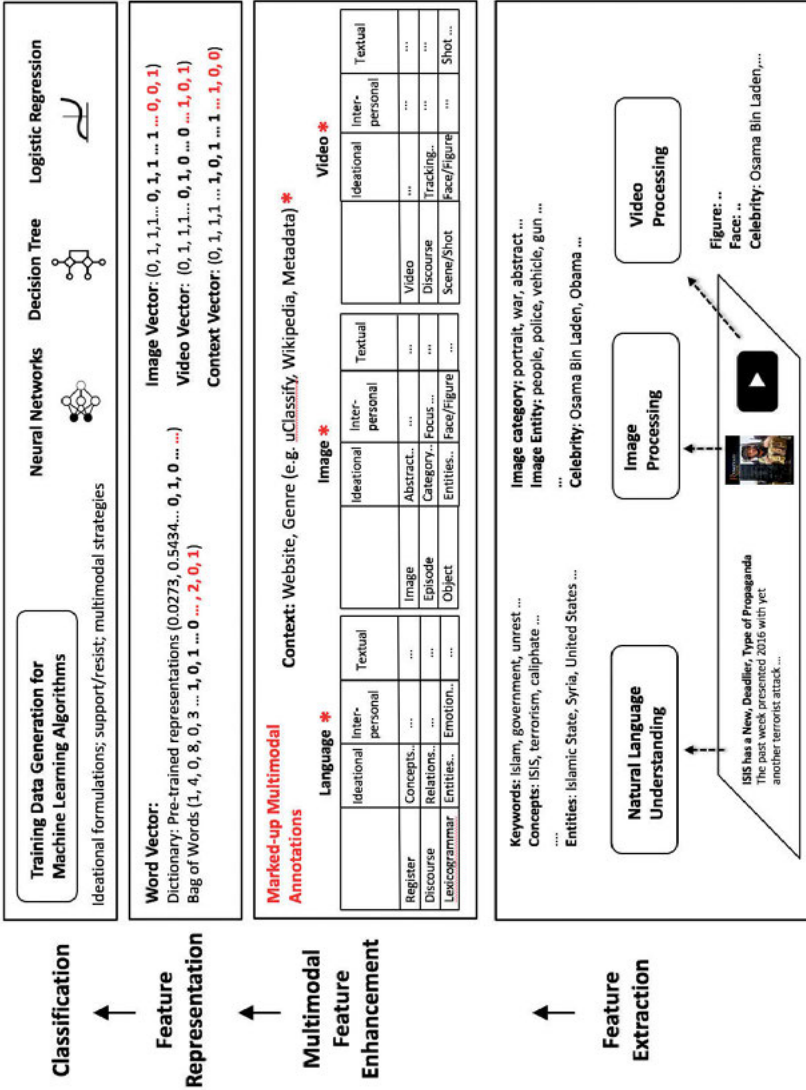
In order to formulate multimodal systems and choices as generalizable abstract structures which can be applied to any instance of use, it is proposed that state-of-the-art automated computational techniques (e.g., text, image, and video processing) are embedded in a multimodal framework which incorporates contextual parameters provided by various forms of metadata. Indeed, research efforts along these lines are already underway.

For example, mixed methods approaches involving the integration of multimodal analysis, data mining, and information visualization for modeling patterns of multimodal communications in large-scale data have been proposed (e.g., O'Halloran et al., 2016; Tan et al., 2018). The computational techniques provide the necessary abstract structures and foundations for scientifically investigating multimodal discourses across media platforms and contexts, building a basis for the future development of multimodality as a discipline.

As a further step forward with this initiative, an approach which incorporates automated computational techniques (e.g., text, image and video processing algorithms) within a multimodal framework and uses machine learning techniques for the analysis of big multimodal datasets is proposed. In this approach, the basic methodology involves representing the multiple dimensions of multimodal texts (e.g., language, image, video, and context) using vectors which record the presence or absence of each feature. Machine learning algorithms are then used to classify multimodal texts according to ideational formulations which are realized and the multimodal strategies which are used. As displayed in Figure 6, the approach has four steps: 1. Feature Extraction; 2. Multimodal Feature Enhancement; 3. Feature Representation; and 4. Classification. These steps are explained in turn below.

1. Feature Extraction: Natural language understanding (NLU) algorithms (for example, IBM Watson) are applied to the language components of the multimodal texts. For example, the NLU models in IBM Watson are *categories*, *concepts*, *emotion*, *entities*, *keywords*, *metadata*, *relations*, *semantic roles*, *sentiment*, *metadata* and *relations*⁵. General descriptions of the language models are provided by the

⁵ <https://natural-language-understanding-demo.ng.bluemix.net/>, last accessed: 29 August 2019



developers of the NLU tools, but as with most commercial NLU models, the criteria, rationale, and algorithms are not provided so these models are black boxes. Nonetheless, the usefulness of the language models when integrated in a multimodal analysis framework, supplemented by text tagging, have been demonstrated (Wignell et al., 2018).

Similarly, visual processing models (for example, DenseCap⁶, Clarifai⁷, Google Cloud Vision⁸, and IBM Watson Visual Recognition⁹) are applied to the images in the multimodal texts to extract semantic information.

For example, the image processing models include *object labelling*, *object bounding*, *face detection*, *face bounding*, *face analysis*, *logo detection*, *logo bounding*, *celebrity detection*, *celebrity bounding*, *apparel labelling* and *web detection*. These image models are also black boxes which nonetheless have proved useful for extracting information from the images (Cao & O'Halloran, 2015; O'Halloran et al., 2014; Podlasov & O'Halloran, 2014). In addition, video processing models (e.g., Amazon Rekognition¹⁰) are applied to extract information from the videos in the multimodal texts. The models include *object*, *scene*, and *activity detection*, *facial recognition*, *facial analysis*, *pathing*, *celebrity recognition* and *text-in-image recognition* for identifying participants, objects, events, and text in the videos.

2. Multimodal Feature Enhancement: The various computational models for language, images, and videos are integrated within a multimodal analysis framework, so that each model is categorized according to semiotic resource, meta-function (experiential, logical, interpersonal, and textual), and rank, as displayed in Figure 6. This means that the results from the computational models are marked up according to the multimodal theoretical framework, which also indicates the gaps where there is missing information. For example, NLU models are largely concerned with experiential meaning at the rank of word group, and neglect the textual organization, where certain elements have a greater semantic input due to the functions of those element (e.g., headline, caption, lead paragraph) (Wignell et al., 2018).

Furthermore, the multimodal feature enhancement takes context into account by incorporating metadata (e.g., URLs, date of postings, source materials, and references) so that the multimodal texts are annotated according to location, time, text type, and other attributes. For example, the text and URLs of websites are

6 <https://cs.stanford.edu/people/karpathy/densecap/>, last accessed: 29 August 2019

7 <https://clarifai.com/>, last accessed: 29 August 2019

8 <https://cloud.google.com/vision/>, last accessed: 29 August 2019

9 <https://www.ibm.com/watson/services/visual-recognition/>, last accessed: 29 August 2019

10 <https://aws.amazon.com/rekognition/>, last accessed: 29 August 2019

analyzed using various algorithms (e.g., uClassify¹¹) which classifies the website into different types of news stories and topics. In this way, multimodal theory and the context are incorporated in the descriptions of the multimodal texts.

3. Feature Representation: The various features of the multimodal analysis and the context are represented by a series of vectors. Dummy variables are used for representing images, videos, and context, i.e. using “0” and “1” to indicate the absence and presence of a feature respectively. Texts are represented using the bag-of-words representation, where a text is a dictionary vector and values corresponds to frequencies of the words appearing in the text. In addition to bag-of-words, texts can be also represented as word embeddings for training neural networks.

Word embeddings provide a series of dense, real-number vectors that are pre-trained over large amount of texts, e.g., a Wikipedia snapshot, which enables embedding vectors to encode semantics of words (Mikolov et al., 2013; Pennington et al., 2014). By coupling representations of images, videos, context, and texts, the multiple dimensions of the multimodal analysis (semiotic resource, metafunction, system, rank, and context) are incorporated into the model, resulting in a multidimensional description of the features of the multimodal texts, consisting of thousands (or more) dimensions. In this way, the complexity of the multimodal analysis is accounted for in the approach.

4. Classification: The multimodal texts are classified using machine learning algorithms which have been trained using previously classified data. Examples of machine learning algorithms include neural networks, decision trees, logistic regression, and other statistical methods. For example, K-modes clustering (Huang, 1997, 1998) and an interactive visualization application are being used to analyze the reuse of images from online terrorist propaganda across different media platforms. The multimodal texts are clustered according to similarities and differences derived from the vectors with the list of features for each multimodal text (see O'Halloran et al., 2016). The proposed techniques and methodologies for integrating multimodal theory with computational models for text, image, videos, and contextual information result in large-scale mapping of the semantic space of multimodal texts, together with classifications of the ideational formations which are created and the multimodal strategies which are used.

In addition, it is necessary to display the results using some form of interactive visualization in order to explore the results (O'Halloran et al., 2016; Tan et al., 2018) The methodology presented here is currently being tested with real-life data and an interactive visualization explores the usefulness of the approach for analyzing large datasets of multimodal texts. That is, the approach is being empirically tested. This

¹¹ <https://www.uclassify.com/browse>

is the basic proposition advocated in the current discussion: i.e. to test concepts, systems, and processes in a rigorous fashion so that the relations are made explicit, reasoning and computation are made as efficient as possible, and the limits of the findings are specified. As such, this discussion presents possible steps towards the scientific study of multimodality, with a view to paving the way for the multimodal sciences.

7 Conclusions

Multimodality holds great promise for addressing serious problems in the world today where truth itself is at stake, given the current era where private corporations are employed to spread false information to influence the outcome of political processes. If multimodality continues to develop by building abstraction upon abstraction without an empirical basis, the result will be “abstract inbreeding”, leading to the possible degeneration of the field (see discussion of mathematics in Neumann, 1956, 2063). Indeed, Bateman and colleagues (Bateman, 2014a,b, 2016; Bateman et al., 2004, 2019) have also called for an empirical basis to multimodal research in order to provide firm foundations for the future development of the field.

Looking back, mathematics and linguistics developed in order to address specific problems at the time. From here, each area developed into disciplines which provided the basis for mathematical and language sciences respectively. In much the same way, multimodality has the potential to address key issue of human communications, leading to a range of multimodal sciences. Recent studies in the United States reveal that there is an increasing number of students completing degrees in linguistics, particularly at undergraduate level, although the rate has slowed down in recent years (The Linguistic Society of America, 2017). The increased interest may be related to the changes in the communication landscape resulting from digital technology. Whatever the reason, this trend offers a promising scenario for multimodality which moves beyond the study of language in isolation to the study of language as it combines with other semiotic resources in human communication.

Galileo’s view of the critical role of mathematics for understanding and predicting the physical world can be extended to multimodality in terms of providing tools for understanding the human world. Paraphrasing Galileo makes this connection explicit:

The human universe cannot be read until we have learned the language and become familiar with the characters in which it is written. It is written in *multimodal language*, and the *signs* are *multidimensional in nature*, without which means it is humanly impossible to comprehend a single *dimension*. Without these, one is wandering about in a dark labyrinth. Based on Galileo (1623 [1957], *Opere Il Saggiatore*).

If multimodality becomes a discipline which provides the foundations for the multimodal sciences, it may become possible to understand the changes in human communication arising from digital technology and the resultant impact on thought and reality.

Bibliography

- Arevalo, John. 2018. *Multimodal Representation Learning with Neural Networks*: Engineering School, Systems and Industrial Engineering Department, National University of Colombia dissertation.
- Arevalo, John, Tharmar Solorio, Manuel Montes-y-Gomez & Fabio A. Gonzalez. 2017. Gated Multimodal Units for Information Fusion. In *Proceedings of the workshop of the 5th international conference on learning representations (iclr)*, Workshop of the 5th International Conference on Learning Representations (ICLR).
- Bateman, John A. 2014a. Looking for What Counts in Film Analysis: A Programme of Empirical Research. In David Machin (ed.), *Visual Communication*, 301–330. Berlin: Mouton de Gruyter.
- Bateman, John A. 2014b. Using Multimodal Corpora for Empirical Research. In Carey Jewitt (ed.), *The Routledge Handbook of Multimodal Analysis. Second Edition*, 238–252. London: Routledge.
- Bateman, John A. 2016. Methodological and Theoretical Issues for the Empirical Investigation of Multimodality. In Nina-Maria Klug & Hartmut Stöckl (eds.), *Sprache im multimodalen Kontext / Language and Multimodality* (Handbooks of Linguistics and Communication Science (HSK) 7), 36–74. Berlin: de Gruyter Mouton.
- Bateman, John A., Judy L. Delin & Renate Henschel. 2004. Multimodality and Empiricism: Preparing for a Corpus-based Approach to the Study of Multimodal Meaning-Making. In Eija Ventola, Cassily Charles & Martin Kaltenbacher (eds.), *Perspectives on Multimodality*, 65–87. Amsterdam: John Benjamins.
- Bateman, John A., Daniel McDonald, Tuomo Hiippala, Daniel Couto-Vale & Eugeniu Costechi. 2019. Systemic-Functional Linguistics and Computation: New Directions, New Challenges. In Geoff Thompson, Wendy L. Bowcher, Lise Fontaine, Jennifer Y. Liang & David Schönthal (eds.), *The Cambridge Handbook of Systemic Functional Linguistics*, 561–586. Cambridge: Cambridge University Press.
- Bateman, John A. & Michael O'Donnell. 2015. Computational Linguistics: The Halliday Connection. In Jonathan J. Webster (ed.), *The Bloomsbury Companion to M.A.K. Halliday*, 453–466. London and New York: Bloomsbury.
- Bateman, John A., Janina Wildfeuer & Tuomo Hiippala. 2017. *Multimodality – Foundations, Research and Analysis. A Problem-Oriented Introduction*. Berlin: Mouton de Gruyter.

- Cao, Yanpeng & Kay L. O'Halloran. 2015. Learning Human Photo Shooting Patterns from Large-Scale Community Photo Collections. *Multimedia Tools and Applications* 74(24), 11499–11516. <https://doi.org/10.1007/s11042-014-2247-0>.
- Eves, Howard. 1990. *An Introduction to the History of Mathematics*. New York: Saunders College.
- Galileo, Galilei. 1623 [1957]. Opere II Saggiatore. In *Discoveries and opinions of galileo*, 237–238. New York: Doubleday and Company.
- Halliday, Michael A. K. 1973. *Explorations in the Functions of Language*. London: Edward Arnold.
- Halliday, Michael A. K. 1978. *Language as Social Semiotic*. London: Edward Arnold.
- Halliday, Michael A. K. 2003. *Collected Works of M. A. K. Halliday. Volume 3. On Language and Linguistics*. London and New York: Continuum.
- Halliday, Michael A. K. 2008. Working with Meaning: Towards an Applicable Linguistics. In Jonathan J. Webster (ed.), *Meaning in Context: Strategies for Implementing Intelligent Applications of Language Studies*, 7–23. London: Continuum.
- Halliday, Michael A. K. & Christian M. I. M. Matthiessen. 2014. *Halliday's Introduction to Functional Grammar. Fourth Edition*. London and New York: Routledge.
- Huang, Zhexue. 1997. Clustering Large Data Sets with Mixed Numeric and Categorical Values. In Hiroshi Motoda & Lu Hongjun (eds.), *Proceedings of the First Pacific Asia Knowledge Discovery and Data Mining Conference, Singapore*, 21–34. Singapore: World Scientific Publishing Co Pte Ltd.
- Huang, Zhexue. 1998. Extensions to the K-Modes Algorithm for Clustering Large Data Sets with Categorical Values. *Data Mining and Knowledge Discovery* 2(3), 283–304.
- Jewitt, Carey (ed.). 2014. *The Routledge Handbook of Multimodal Analysis. Second Edition*. London and New York: Routledge.
- Jewitt, Carey, Jeff Bezemer & Kay O'Halloran. 2016. *Introducing Multimodality*. London: Routledge.
- Kiela, Douwe, Edouard Grave, Armand Joulinand & Tomas Mikolov. 2018. Efficient Large-Scale Multi-Modal Classification. In *Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18)*, 5198–5204. Palo Alto, California USA: AAAI Press.
- Kline, Morris. 1980. *Mathematics: The Loss of Certainty*. New York: Oxford University Press.
- Little, John. 1981. Review: Mathematics: The Loss of Certainty. *New Scientist*, 159–159.
- Malinowski, Bronislaw. 1923. *The Problem of Meaning in Primitive Languages* 451–510. Harcourt, Brace, and Co., Inc. Supplement I to C.K. Ogden and I.A. Richards *The Meaning of Meaning*.
- Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S. Corrado & Greg Dean. 2013. Distributed Representations of Words and Phrases and their Compositionality. In Christopher J. C. Burges, Leon Bottou, Max Welling, Zoubin Ghahramani & Kilian Q. Weinberger (eds.), *Advances in Neural Information Processing Systems 26 (NIPS 2013)*, 3111–3119. Lake Tahoe, Nevada, USA.: NIPS Foundation.
- National Science Council. 2013. *The Mathematical Sciences in 2025*. Washington D.C.: The National Academies Press.
- Neumann, John. 1956. The Mathematician. In James R. Newman (ed.), *The World of Mathematics*, vol. 4, 2053–2063. New York: Dover Publications Inc.
- O'Halloran, Kay L., Alvin Chua & Alexey Podlasov. 2014. The Role of Images in Social Media Analytics: A Multimodal Digital Humanities Approach. In David Machin (ed.), *Visual communication*, 565–588. Berlin: Mouton de Gruyter.
- O'Halloran, Kay L., Sabine Tan, Duc-Son Pham, John A. Bateman & Andrew Vande Moere. 2016. A Digital Mixed Methods Research Design: Integrating Multimodal Analysis with Data Mining

- and Information Visualization for Big Data Analytics. *Journal of Mixed Methods Research* 12(1). 11–30. <https://doi.org/10.1177/1558689816651015>.
- Pennington, Jeffrey, Richard Socher & Christopher D. Manning. 2014. GloVe: Global Vectors for Word Representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543. Doha, Qatar: Association for Computational Linguistics.
- Podlasov, Alexey & Kay L. O'Halloran. 2014. Japanese Street Fashion for Young People: A Multimodal Digital Humanities Approach for Identifying Socio-Cultural Patterns and Trends. In Emilia Djonov & Sumin Zhao (eds.), *Critical Multimodal Studies of Popular Culture*, 71–90. London and New York: Routledge.
- Snidaro, Lauro, Jesus Garcia, James Llinas & Erik Blasch. 2016. *Context-Enhanced Information Fusion: Boosting Real-World Performance with Domain Knowledge*. Cham: Springer International Publishing AG Switzerland.
- Tan, Sabine, Kay L. O'Halloran & Peter Wignell. 2019. Multimodality. In Anna D. Fina & Alexandra Georgakopoulou (eds.), *Handbook of Discourse Studies*, Cambridge: Cambridge University Press.
- Tan, Sabine, Kay L. O'Halloran, Peter Wignell, Kevin Chai & Rebecca Lange. 2018. A Multimodal Mixed Methods Approach for Examining Recontextualisation Patterns of Violent Extremist Images in Online Media. *Discourse, Context and Media* 21. 18–35.
- The Linguistic Society of America. 2017. The State of Linguistics in Higher Education: Annual Report. Tech. rep. The Linguistic Society of America.
- Wignell, Peter, Kevin Chai, Sabine Tan, Kay L. O'Halloran & Rebecca Lange. 2018. Natural Language Understanding and Multimodal Discourse Analysis for Interpreting Extremist Communications and the Re-Use of these Materials Online. *Terrorism and Political Violence* <https://doi.org/10.1080/09546553.2018.1520703>.



Part III: Diversity

Axel Schmidt and Konstanze Marx

Multimodality as Challenge: YouTube Data in Linguistic Corpora

Abstract: A large database is a desirable basis for multimodal analysis. The development of more elaborate methods, data banks, and tools for a stronger empirical grounding of multimodal analysis is a prevailing topic within multimodality. Prerequisite for this are corpora for multimodal data. Our contribution aims at developing a proposal for gathering and building multimodal corpora of audio-visual social media data, predominantly YouTube data.

Our contribution has two parts: First we outline a participation framework which is able to represent the complexity of YouTube communication. To this end we 'dissect' the different communicative and multimodal layers YouTube consists of. Besides the video performance YouTube also integrates comments, social media operators, commercials, and announcements for further YouTube videos. The data consists of various media and modes and is interactively engaged in various discourses. Hence, it is rather difficult to decide what can be considered as a basic communicative unit (or a 'turn') and how it can be mapped. Another decision to be made is which elements are of higher priority than others, thus have to be integrated in an adequate transcription format. We illustrate our conceptual considerations on the example of so-called *Let's Plays*, which are supposed to present and comment computer gaming processes.

The second part is devoted to corpus building. Most previous studies either worked with ad hoc data samples or outlined data mining and data sampling strategies. Our main aim is to delineate in a systematic way and based on the conceptual outline in the first part necessary elements which should be part of a YouTube corpus. To this end we describe in a first step which components (e.g., the video itself, the comments, the metadata, etc.) should be captured. In a second step we outline why and which relations (e.g., screen appearances, hypertextual structures, etc.) are worth to get part of the corpus. In sum, our contribution aims at outlining a proposal for gathering and systematizing multimodal data, specifically audio-visual social media data, in a corpus derived from a conceptual modeling of important communicative processes of the research object itself.

Keywords: multimodality, YouTube, multimodal corpora, Let's Plays, social media

<https://doi.org/10.1515/9783110608694-005>

1 Introduction

For multimodal analyses, it is desirable to have a large database (in our case of YouTube data). Within multimodality, the development of more elaborate methods, data banks, and tools to allow a stronger empirical grounding of multimodal analysis is currently an important topic (cf. Bateman et al., 2017, 152–155). Corpora of multimodal data are a prerequisite for this. Our contribution aims at developing a proposal for gathering and building multimodal corpora of audio-visual social media data, predominantly YouTube data. One of the main challenges in constructing corpora is to make them useful for pursuing specific questions that are relevant within a given scientific approach. Within social semiotics, semiotic products like websites, clips on YouTube or comments are viewed as constituting communicative acts (Kress & van Leeuwen, 2006 [1996]; Kress, 2010). Our aim in this contribution is to ground the process of data gathering and corpus building in assumptions about the research object itself. The questions we are asking are: What does a corpus of YouTube data have to look like to allow researchers to tackle questions that are relevant for a study of communicative acts? What components should such a corpus contain?

Although we do think that multimodality is becoming more important because communication has become more diverse and multimodal, we are skeptical about transferring multimodality into a discipline in its own right. This is mainly because it is difficult to identify a unique research object, which would be crucial for a new discipline. The meaning-making functions of different modal resources and their specific relations in communicative processes are not just a topic of multimodality, but of all disciplines that are concerned with reconstructing social meaning (cf. Habermas, 1967). Moreover, a focus on media communication within multimodality, which could be considered a specific focus, seems not to offer a systematic grounding because so-called non-mediated communication (such as, for example, face-to-face-interaction) is also organized multimodally. In sum, it is not clear to us how multimodality could be differentiated from other disciplines, especially sociology, media and communication studies, and semiotics.

Our contribution has two parts: First, we outline a participation framework¹ which can represent the complexity of YouTube communication, drawing mainly

¹ If semiotic material on websites is understood as communication, i.e., as something people are doing with another, they establish in and through their communication something Goffman has called a *participation framework* (Goffman, 1981, 137). The notion *participation framework* describes the relations which participants accomplish in and through their communication. Participation frameworks are on the one hand constrained and enabled by situational and technological parameters (for instance, by whether a communication is face-to-face or technically mediated,

on suggestions by Adami (2009b), Dynel (2014), and Eisenlauer (2014). To this end we ‘dissect’ the different communicative and multimodal layers that YouTube consists of. Besides the video component, YouTube also integrates comments, social media operators, commercials, and suggestions for further YouTube videos. The data consists of various media and modes and is interactively engaged in various discourses. Hence, it is difficult to decide what can be considered the basic communicative unit (or ‘turn’). We illustrate our conceptual considerations with an example, the so-called ‘Let’s Plays’. In this genre, which has become very popular in a very short period of time (Hale, 2013, 3), gamers document their gaming in films and present it to a (potential) mass audience via upload on video hosting websites like YouTube.

The second part of this chapter is devoted to corpus-building. Most previous studies of YouTube and similar media either work with ad hoc data samples or outline data mining and data sampling strategies (for references see Section 3). Our main aim is to identify necessary elements that should be part of a YouTube corpus in a systematic way based on the conceptual outline in the first part. To this end we initially describe which components should be captured (e.g., the video itself, the comments, the metadata, and so on). In a second step we outline which relations ought to be part of the corpus and why (e.g., screen appearances, hypertextual structures, etc.). Another decision to be made is which elements are of higher priority than others and, thus, have to be integrated in an adequate transcription format (Beißwenger, 2009; Recktenwald, 2017; Marx & Schmidt, forthcoming).

In sum, our contribution aims to outline a proposal for gathering multimodal data and making it accessible in a systematic way, specifically audio-visual social media data, via building a corpus that is derived from the conceptual modeling of important communicative processes of the research object itself. What is important, thus, results from a description of the communicative structures and the participation framework on YouTube.

whether it is written or oral, etc.; cf. Meyrowitz 1990). On the other hand, participation frameworks are indexed by the ongoing activities of the participants, and therefore are in constant flux (cf. Goodwin, 1986; Goodwin & Goodwin, 2004; Arminen et al., 2016). Goffman has termed the contribution of a single utterance (or parts of it) to the reflexive accomplishment of participation frameworks, *footing* (Goffman, 1981). In our case (communication on YouTube) we are interested in how technological parameters ‘afford’ the possibilities of accomplishing participation frameworks on YouTube.

2 Modeling Communication on YouTube

The social media service YouTube is described in many different ways, for example as post-television (Tolson, 2010), as creating a distinct aesthetic, often referred to as ‘Youtubeness’ (Burgess & Greenberg, 2014) or as an alternative business model in comparison to traditional media like television (Vonderau, 2016).

However, our main interest is to consider YouTube as a *specific* medium facilitating a *specific* form of communication which, in turn, enables specific kinds of participation. Forms of communication specify situational and technological conditions (like written/oral/audio-visual; one-way/reciprocal; public/private, etc.) without determining communicative uses (cf. Holly, 1997; Habscheid, 2000; Dürscheid, 2005; Schmitz, 2015). There is a growing body of studies concerned with the specific form of communication made possible by YouTube. Situated somewhere between mediated interpersonal communication and so-called mass media communication like television (Dynel & Chovanec, 2015), the communicative form YouTube enables is described as polylog (Bou-Franch et al., 2012), as video interaction (Adami, 2009a,b, 2015; Schmidt, 2011), as enabling viewer involvement (Frobenius, 2013, 2014), or as dialogical exchanges (Jones & Schieffelin, 2009).

Our theoretical and methodological background is interactional pragmatics (cf. D’hondt et al., 2009). Instead of focusing on language structures or systems, we are interested in how language and other modal resources are used to constitute activities. The starting point, therefore, is the purposeful *doing* of participants and the establishment of a *participation framework* in and through communicative exchanges. We start with the observation that YouTube establishes mediated interactional exchanges, similar to mediated interpersonal communication (Konijn et al., 2008). Although YouTube-communication is asynchronous, physically distant and involves an indeterminate viewership, it holds the possibility of producing ‘turns’, e.g., posting a video, and reacting to ‘turns’, e.g., by writing a comment or by posting an answering video (cf. Adami, 2009a; Schmidt, 2011). Since every interaction is situated, exchanges on YouTube can be described as distant situations spliced together via media technology (cf. Thompson, 1995). Those “synthetic situations” (title of a paper by Knorr-Cetina 2009) create a “response presence, without needing to be in one another’s physical presence” (Knorr-Cetina, 2009, 69). Thus, Social Media and YouTube generate a specific kind of participation framework modifying familiar forms of interactional participation and involving different communicative levels (Frobenius et al., 2014).

Our interest is to describe YouTube’s participation framework as a multimodal form of communication involving different levels of participation. This description can then be used as a basis to identify which elements should be part of a YouTube

corpus. Background to this study is a larger project of developing a standard for multimodal corpora of audiovisual YouTube data. Corpora, in general, are compiled to provide an empirical basis for research. A very important aspect of corpus construction, and in particular of the development of a standard for future corpus formation, is the question of which data should be integrated into a corpus (standard) and what form this data should take. Simply gathering data in a corpus without a thought-through design is not a promising strategy. For this reason, we first aim to develop a basic understanding of our research object in order to specify which data are needed and what format the data should take to allow research within the framework of the approach outlined above. A crucial aspect of this basic understanding is the participation framework and its levels, which are established in communication on YouTube.

Before explicating these levels, we introduce an example for illustrative purposes from the data we are currently analyzing. Our data consists of so-called ‘Let’s Plays’ (in the following referred to as LPs). In the literature LPs are defined as “playing videogames for the internet” (Hale, 2013, 3). In the simplest case, a user records his/her playing of a video game and his/her simultaneously produced comments and uploads the result to video hosting websites like YouTube (cf. Ackermann, 2016; Stephan, 2014; Marx & Schmidt, 2018, forthcoming). In addition, the player usually appears in a facecam (see Figure 1).² Single player LPs are videos between 30 mins and 2 hours that are accessible via video hosting websites. LPs may be watched in an embedded mode (see Figure 2) or a full screen mode (see Figure 1). Usually (as Figure 1 shows) they consist of the game play that fills almost the entire screen with the player giving comments and appearing in a facecam.

There are variants: In addition to single player LPs as described above, LPs can also be produced by several players; then they are called *Let’s Play Together* or *Multiplayer Let’s Plays* (for a more detailed description of such *Multiplayer Let’s Plays* see below). In addition, LPs can be recorded as described above, or they can be viewed as a live stream as on platforms like Twitch³ (cf. Recktenwald, 2017). The first Let’s Play appeared in 2006 on the website ‘*something awful*’⁴. Nowadays, Let’s Play videos on YouTube as well as the channels of so-called Let’s Players achieve high click rates. The German Let’s Player Gronkh, for instance, has about 4.8 million YouTube channel subscribers as of August 2018.

² Let’s Players usually use a so-called facecam, which conveys a visual image of the player’s face.

³ Twitch is a live streaming video platform that specializes in live streams of Let’s Plays.

⁴ See <https://www.somethingawful.com/>, last accessed: 29 August 2019.

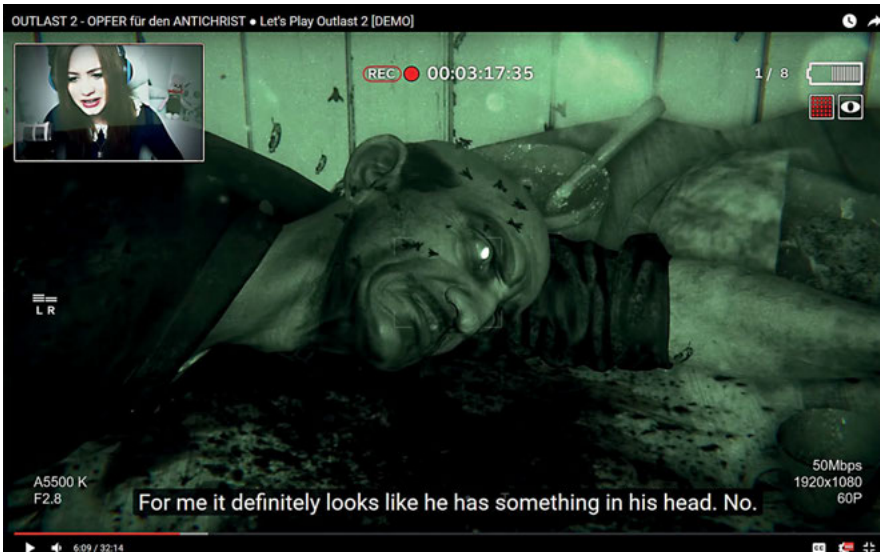


Fig. 1: A screenshot of a Single Player Let's Play in an embedded view.

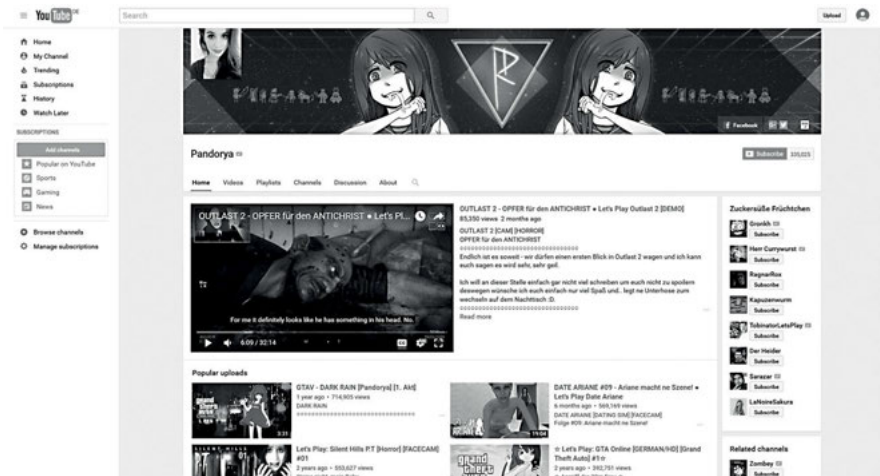


Fig. 2: A screenshot of a Single Player Let's Play in a full screen view.

Our example data consists of an extract lasting 1:25 minutes of a *Let's Play Together* (lasting a total of 23 mins) where four people (Gronkh, Curry, Tobi, and Pan) in spatial distance from each other play a video game together while at the

same time talking to each other (via Teamspeak⁵) and commenting their play moves. The game they are playing together, *Dead by Daylight*⁶ (Starbreeze Studios, 2017), is shown on the respective screens of each player. Three of the four screens show an integrated facecam capturing the face of the respective player (see Figure 3).

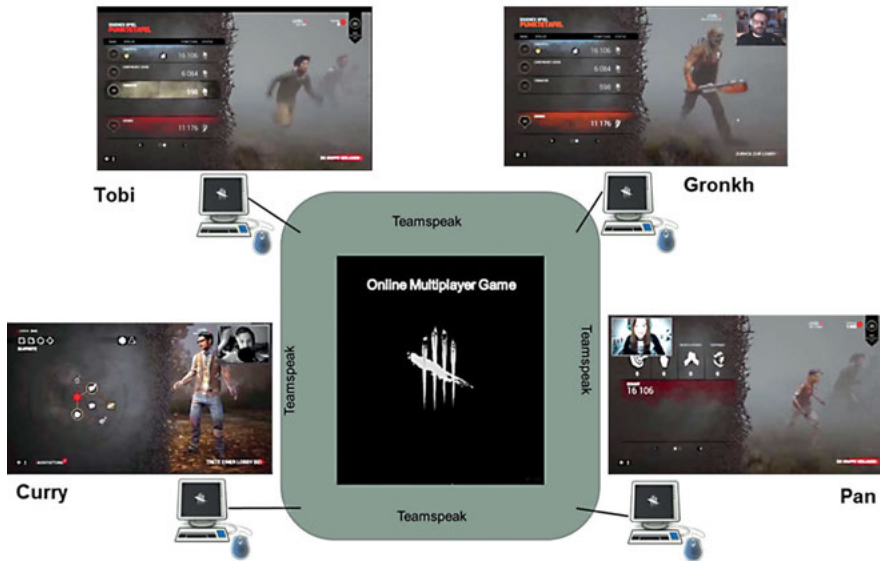


Fig. 3: A schematic diagram of a Let's Play Together.

Each play represented on the respective screen is individually recorded and afterwards uploaded on YouTube. In the following, we rely mainly on the version produced by *Gronkh* (Gronkh, 2016). Extract 1 in Figure 4 is from the beginning of the game session; the four players have just entered the lounge and are now trying to get the game started. In the transcript, we focus on the verbal comments by the participants (in black font), on status displays (appearing after the abbreviation *StaD*), game events (appearing after the abbreviation *GE*) and on game sounds (after *GS*). Talk is assigned to speakers by capital letters (A = Gronkh, B = Pan, C =

⁵ Teamspeak is an application for audio communication between users on a chat channel, similar to a telephone conference call.

⁶ *Dead by Daylight* is an asymmetric survival horror game which is played exclusively as a one versus four online multiplayer game.

Curry, D = Tobi). The participants speak German (original language is marked in bold). An indicative English translation is given in the line below. The non-verbal events are aligned with talk and pauses via special characters like %, &, etc. (cf. Mondada, 2014). Letters after abbreviations (a, b, c, d) indicate on which screen the events appear. Standalone small letters (a, b, c, d) indicate activities conveyed by the facecams. Stills are not used as they are not necessary for our argumentation here.

Extract: “okay i’m going to press REAdy now”/Let’s play together DbD #2/10.6.2016. Four LPers (Gronkh = A; Pan = B; Curry = C; Tobi = D) play the game *Dead by Daylight* together online.

```

1   A      *(.) also ich drück jetz ma auf FERTIG,
      okay I'm going to press ready now
GS   >>threatening music--->
Stada *button READY changes to NOT READY; red tick above C1
2   A      wenn ihr jetz auch auf fertig drückt unten RECH*TS,
      if you too press ready now at the bottom right
Stada                                     *tick above C3,
      line on CS
3   ma gucken was pas *[SIERT,]
4   let's see what happens
5   D      [ja, ]
      yes
Stada                                     *tick above C4, line on CS, button NOT READY
      disappears, countdown 0.06 starts
6   D      (0.37)
7   das is ne (xxx [xxx]      ]
      that is a (xxx xxx)
8   A      [SPIELbeginn; ]
      start of play
9   D      (.) [oh es geht      ] [LOS;      ]
      oh it's starting
10  C      [okay das geht auch] [SO;      ]
      okay it also works this way
11  B      [oh;      ]
      oh
12  B      (.) ja ja oh [GEIL; ]
      yes yes oh great
13  A      [ja      ] [nice; ]
      yes nice
14  D      [ach du] [SCHEISse; ]
      oh my gosh
15  A      [halt halt] halt) *YE[A: ]:H;
      wait wait wait yeah
16  B      [go-]
      [go-]
GE/STAda                                     *countdown:
      0.00, lettering disappears, DS turns black
17  B      [dann m      ]oderier ma AN;
      then start the moderation

```

```

18 C      [<<lachend> eHE,>]
          [<<laughing> eHE,>]
19 A      +(.) #[ihr seid TOT;      ]
          you are dead
20 C      [((räuspert sich  ))]
          [((clears his throat))]
GEC      +black DS and FC appear
GED      #black DS appears
21      (0.32)
22 A      $äh JA;
          äh YES;
GEB      $black DS appears
23 A      HALlo %un herzlich willKOMmen
          HELlo and a warm welcome
GEa-d    %game screen appears
24 A      $^bei ~dead ^^by %DAY~~light;
          to dead by daylight
STADb-d  %game symbols appear
GEB      $FC appears
b        ^waves, smiles, looks into FC
c        ~looks into FC, raises his eyebrows
25 A      %^hh un wir gehn_n die nächste RUNde,
          %^hh and we start with the next LAP,
STADb,d  %text fields appear
26 A      heute mit & (0.27)*weiterem beSUCH- &
          today with& (0.27)*some more Visitors-&
GS       &rustling plonk sound--->>
STADa    *game symbol appears

```

Fig. 4: Extract 1: “okay i’m going to press REAdy now”/Let’s play together DbD #2/10.6.2016.

At the beginning of the transcript (lines 1–16), the four players are concerned with the technical aspects of getting the multiplayer game started. Their talk, although already transmitted to the public, is obviously designed to achieve a joint start of the game. So A’s announcement in line 1 okay I’m going to press ready now is subsequently expanded to an encouragement addressed to his co-players to do the same, with if you too press ready now at the bottom right (line 2). What follows is a successive start of the game which is accompanied by comments of the four players. The comments in this phase merely serve the purpose of mutual coordination. Only at line 17, after having established and announced the start of the game (most obviously by A’s exclamation start of play in line 8), an intro moderation is requested (with B’s then start the moderation in line 17), which is subsequently mainly delivered by A (lines 19–26).

In this extract, three kinds of communication are at work at the same time: First, we (as the observers) watch the events on screen (how the videogame gets started) and listen to the accompanying verbal talk between the players. Second, we may feel addressed by the players welcoming the viewers to their joint game event (for instance, by saying hello and a warm welcome in line 23). Third, we

may divert our attention from watching the video to reading the comments below the video slot on the screen (not part of the transcript), which may give us an impression of how other audience members perceive the video.

Accordingly, Dynel (2014) differentiates three basic levels in the case of communication on YouTube: (1) the video interaction, (2) the sender-recipient interaction, and (3) the comments. Figure 5 summarizes these three possible forms of communication with respect to the example introduced above. Dark grey is used for level one, black for level two and light grey for level three.

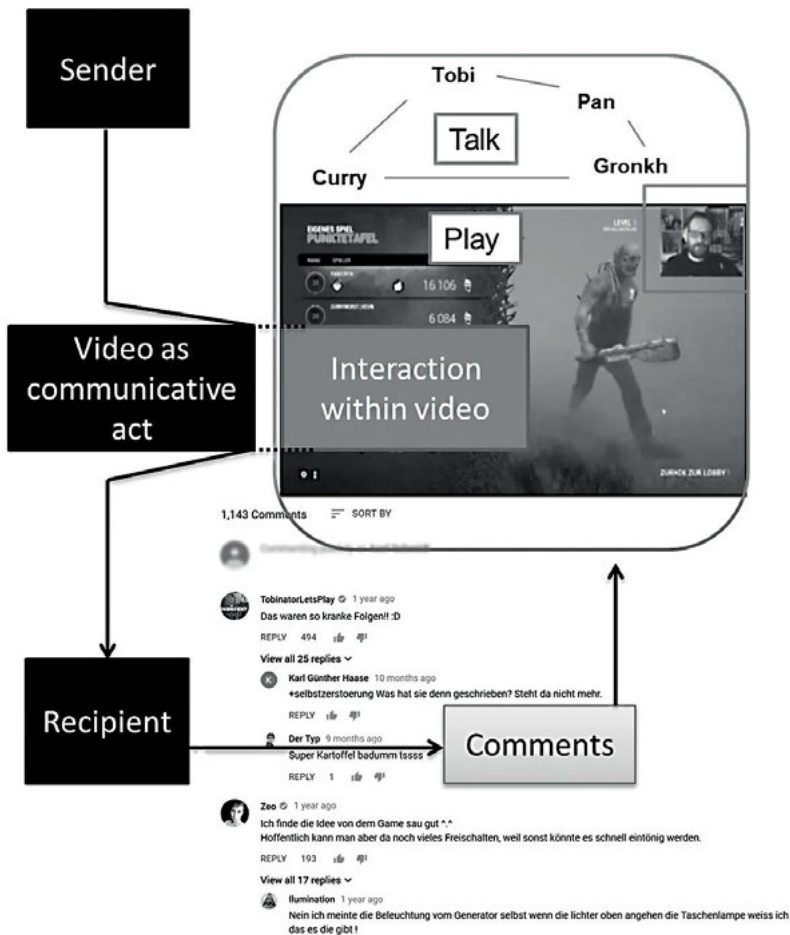


Fig. 5: Basic forms of communication on YouTube.

In the following, the three levels of communication on YouTube are discussed in more detail. Let us turn to the first level, the video interaction level.

2.1 Video Interaction Level

Videos can show people talking and interacting. Depending on the content of the video, recipients witness forms of participation they are familiar with either from face-to-face interaction or from mass media communication. Videos can show face-to-face interactions (like in talk shows) or pure verbal interactions (like in our case between the four players). In this case the basic participation framework with speaker and hearer categories as outlined first by Goffman's paper on footing (1981; cf. Levinson, 1988) is adaptable, namely the distinctions between ratified and non-ratified and addressed and not addressed participants.⁷ In other words: first-level interactions depicted by video resemble face-to-face-interactions with respect to possible participation roles. This, of course, holds only for depicted interactions within the video (and not for other kinds of communication on the platform). In our extract above, for example, we can perceive the content of the video as a focused interaction between four people made possible by technology.

In addition, the idea of different roles within a participation framework is adaptable in a next step to mass media communication (cf. O'Keeffe, 2007; Scannell, 1991). In this sense, videos on YouTube can either show para-social interaction, in which speakers within the video address an audience directly (cf. Horton & Wohl, 1956; Vorderer, 1996), or a video may indirectly target its viewers, thereby creating what has widely been understood as an 'overhearing audience' (cf. Heritage, 1985; Hutchby, 2006; Clayman, 2006).

In the above example, both of these roles are present: When Gronkh is welcoming the viewers with hello and a warm welcome in line 23, he is addressing the

⁷ Due to lack of space the basic categories of the participation framework by Goffman (1981) are only briefly reviewed in this footnote: Goffman distinguishes on the side of the 'hearers' between ratified and unratified participants. The former build what he calls a focused interaction, that is a certain number of people sharing a common attentional focus for a certain amount of time. The latter, the unratified participants, are all others who are in response presence but not officially part of the interaction. Ratified participants are further divided into addressees, who are addressed by a particular utterance, and side participants, who are part of the focused interaction but are not addressed. Unratified participants are understood as bystanders and subdivided into overhearers, who witness ongoing talk accidentally, and eavesdroppers, who purposefully eavesdrop a conversation. Goffman's model is the basis for most approaches dealing with participation. It has been widely taken up and extended, reformulated, and adapted (cf. Levinson, 1988; Goodwin, 1986; Dynel, 2014).

audience directly. In contrast, when he is playing and talking with other players (for instance in lines 1–16), he is engaged in an interaction which is both carried out for the purpose of coordination with his co-players and produced for viewers as he is talking in a public space. Therefore, this kind of interaction targets the audience indirectly. In both cases we are dealing with podium or platform formats (Goffman, 1981, 1983) which involve unequally distributed possibilities to participate.

Accordingly, YouTube viewers can focus on the content of a video (e.g., a TV series, parts of a movie or, as in our case, a commented video game), or they can treat it as a release of a YouTube sender who communicates something with it. This is often recognizable in the comments, which either deal with the content or with the sender (cf. Dynel, 2014, 44). This brings us to the second communicative level, which is the sender-recipient level.

2.2 Sender-Recipient Level

Beginning with the *reception end*, the second level is constituted by the fact that videos are uploaded for viewers, who are, in any case, ratified participants. This holds even if recipients are unregistered, not logged in or remain passive, as each of these participation statuses is possible and legitimate on YouTube. In all cases recipients are external to the interaction shown in the video at the first level because they are not able to intervene and, thus, unable to alter the symbolic flow of the video itself. In other words: Interaction depicted *within* the video is not contingent on the activities of the recipients. Thus, *depicted* participation frameworks within the video remain unaffected by recipients' activities. In our extract above, for example, recipients are able to listen to the conversation of the four players and they may comment on it *afterwards*, but they are not able to participate or intervene in the interaction depicted within the video. This is due to the recorded character of the video.

In addition, recipients are distant and distributed. As long as they do not post anything, it is not possible to determine who is watching. The “response presence” (Knorr-Cetina, 2009, 69) of a (potentially) mass audience only becomes evident via website metrics such as number of views, subscribers, and so on. In our case, we know from the metrics shown below the video that the video has had over one million views and over one-thousand comments. Therefore, YouTube ‘senders’⁸ are to a large extent always oriented towards imagined (mass) audiences (Androutopoulos, 2014).

⁸ The term *sender* is chosen in accordance with (Dynel, 2014, 42–45). The term should indicate the mass media-like aspect of communication on YouTube.

As opposed to the reception end, there is a video author and/or releaser at the *production end* (technically termed as sender). While the term *author* indicates an involvement in the production process of a video, the term *releaser* merely refers to the distribution of a video (in this case the process of uploading).⁹

Dynel (2014, 43–45) differentiates between three production formats: first, *re-publications* of mass media content; second, *modified versions* of existing videos; third, self-made videos termed as *vlogging*. Depending on the production format, the authorial status of the sender varies. In re-publications, a collective professional sender is embedded¹⁰; the releaser does not gain authorial status. Modified videos have an original author (the ‘first sender’) and the releaser (the ‘new sender’) becomes a ‘top layer author’ of the new version. Finally, *vlogging* videos are authored by individuals, who are usually the owner of an account or a channel on YouTube.

Furthermore, videos are not presented in isolation. They are embedded within a website containing manifold cues for understanding, partly authored by the releaser, like the title of the video or its short description. In our case, for example, the video is entitled “GRONKH macht GYROS! – DEAD BY DAYLIGHT #002 – Gronkh”, which contains information about the sender (communicating under the pseudonym *Gronkh*), the game that is played (*Dead by Daylight*), the episode (#002) and a satirical description of the content (*GRONKH macht GYROS!*, to be translated as *GRONKH makes GYROS!*, which alludes to the fact that the game belongs to the horror/slasher genre). Independent of the video’s content and its production format, the releaser and thus the owner of the account is seen as the (last) sender (but not necessarily as its author). In our case, the Let’s Player *Gronkh* is simultaneously the author and the releaser of a self-made video uploaded on his eponymous YouTube channel.

The roles of speaker, sender, author, and account-owner coincide particularly in those cases where vloggers speak to the camera directly (as in our case *Gronkh*). In such cases, it is more likely that the video author/releaser is personally addressed by recipients increasing the likelihood of a verbal exchange between sender and recipients.

⁹ We adopt both terms—*author* and *releaser*—from (Dynel, 2014, 42–45). *Author*, in addition, is a term used by Goffman to indicate who is responsible for constructing an interactional move (who has chosen the words, the images, etc.).

¹⁰ *Embedded* means that original content—for example a Hollywood movie or parts of it—are re-published by a YouTube user (a ‘new sender’, so to speak) who was not part of the original production or distribution context. The original sender, in this case the film company, is thus embedded as a first sender within the re-publication on YouTube by a new sender.

Communication on the sender-recipient-level is established through uploading and watching videos. It does not require any comments to be sufficiently established. However, comments are an additional option for recipients to engage further. This brings us to the third level, the level of comments.

2.3 Level of Comments

In contrast to mass media, YouTube allows an alternation of sender and recipient roles. Everyone can release videos and comment on videos. In this way, YouTubers can interact with one another publicly but, of course, without being able to change the content of released videos on the first level. Though meant as a video sharing website originally (Vonderau, 2016), YouTube enables interactional exchanges of several different kinds: only between video releasers through video responses (Adami, 2009b), between releasers and commenters, or only between commenters.

Therefore, comments are often specifically addressed either by technical options (such as a reply-button), or by using specific signs (like the @-symbol). However, as YouTube communication is persistent, everybody can witness the whole interaction and join in at any time (indexed by the time stamps).¹¹ In this way, communication on YouTube creates an endless ‘open state of talk’ (Goffman, 1981) leading to asynchronous communication typical for social media platforms. In our case, *Gronkh*’s video was released on June 10, 2016 and has received 1,143 comments as of August 2018. Most of them, in turn, received answers by other commentators. The oldest comment is from June 2016, while the newest one (at the time of writing this chapter) was posted in July 2018. Thus, the communication that was generated by this video of *Gronkh* currently bridges a time span of approximately two years (and is still ongoing).

The model from Dynel (2014) that we have discussed so far should be extended by a fourth level, which we call website-user interaction.

2.4 Level of Website-User Interaction

Besides interaction between a video sender and recipients, there is interaction between the website as a communicator, in this case the platform YouTube as a part of Google, and a YouTuber as a user, whether as a producer and/or as a recipient (Eisenlauer, 2014). YouTube provides a designed space, including for

¹¹ The persistence of YouTube comments is limited, however, because video releasers have the ability to moderate comment-level interaction by deleting comments and blocking users.

example templates, basic functionalities, indexes, and metrics as well as a basic broadcast structure like channels and multichannel networks (Vonderau, 2016). The platform's basic structures can be seen as affordances (Gibson, 1979; Arminen et al., 2016) creating exchange in the form of a human-machine interaction on a separate level. This means that, while watching videos and posting comments, users also interact with software designed for specific purposes. This level also has to be taken into consideration when building a corpus.

In our case, for example, watching the video of *Gronkh* happens in a predefined template provided by the website. At the same time, watching the video triggers the display of several automatically generated features like indexes, playlists, and ads. Both are 'communicative acts' which are not attributable to the video author/releaser but to the hosting/distributing website.

In addition, on this fourth level the question of ratified/unratified participation reappears as the difference between (un-)registered and (not) logged-in participants as it is the platform which regulates the formal dimension of participation via technical implementations (Boyd, 2014).

2.5 Interim Conclusion: Participation on YouTube

The basic structure of communication on YouTube looks like this: A (an active user) is doing C (posting a video or posting a comment) for B (indeterminate group of viewers), who, in turn, have the chance to answer either by writing a comment or by releasing a video response. The result of this is an interlacing structure that includes an interaction within the video to be seen and commented on by the audience, creating a second (sender-receiver), a third (comments/response video), and a fourth (website-user) level of interaction.

In this model, communicative acts (or turns, or moves) can occur at all levels in any modal form (verbal or non-verbal, spoken or written). Every logged-in person is able to either just watch or to get engaged at any level (e.g., video producing and uploading, writing comments, and so on). Comments may or may not address video releaser(s)/author(s), or other commenter(s). 'Participation' means any activity at the production or reception end. However, identifying participants drawing only on online data are limited to its active users, as passive ones are not visible on the surface of the website.¹²

The basic participation framework underlying communication on YouTube outlined above is summarized in Figure 6.

¹² Although passive users are not interacting, they, as imagined audiences, are still part of the participation framework.

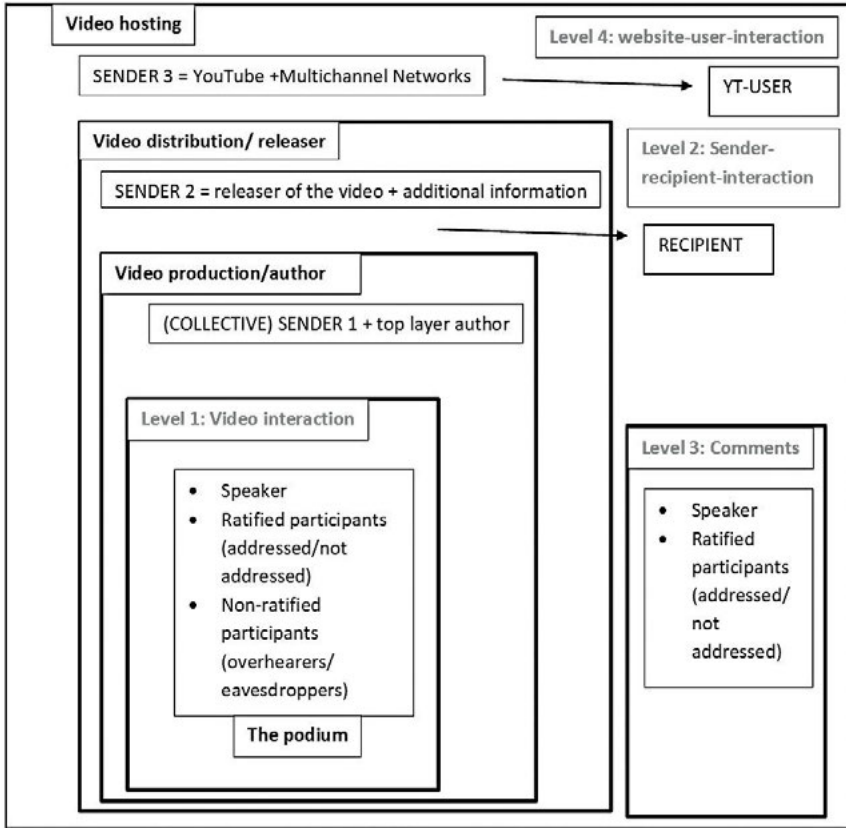


Fig. 6: Basic participation framework of communication on YouTube.

3 YouTube Corpora

In the second part of this chapter, we use the model developed above to formulate requirements for a multimodal corpus of YouTube data. As a multimodal form of communication, YouTube draws heavily on multimodal resources like audio-visibility, images, and design. In order to understand communication on YouTube, it is necessary to include everything that is relevant for meaning making. The decision what is relevant follows from the conceptual schema we introduced above, which is, in turn, rooted in our theoretical framework of interactional pragmatics we outlined above. In other words: we delineate which elements of YouTube's communication should be part of a corpus in order to be able to do research in the

framework of our approach. Thus, this delineation has to be as broad as possible. The corpus structure should be designed in such a way that it allows the study of many different research questions.

Recent studies using YouTube corpora are concerned mainly with data sampling and data mining strategies (cf. Abidi et al., 2017; Androutsopoulos, 2014; Bou-Franch & Garcés-Conejos Blitvich, 2014; Frobenius, 2013, 2014; Ivkovic, 2013; Maybury, 2012; Siersdorfer et al., 2014; Tereick, 2013, 2016; Uryupina et al., 2014; Zhang & Kramarae, 2014). In contrast, our main concern is not how to gather or store data but to outline in a systematic way which components should be included in a YouTube corpus. This chapter represents a first step towards these goals by defining such requirements, which remain purely conceptual at this stage.¹³

We distinguish between components and relations. *Components* are single elements which can be identified visually on the webpage by recipients using a browser (for example the videos or the comment section). Components should be captured individually. *Relations* refer to the overall structure of the webpage, in particular relations between components resulting in a ‘designed space’ of the webpage (for example, how a video is embedded within a webpage). Those relations have to be captured as a whole, preserving the semiotic structure. Components and relations are different forms of capturing communication on YouTube and, of course, overlap in its representation within the corpus.

The next sections deal first with the components and then with the relations. As the construction of a corpus cannot start with everything at once, we outline a stepwise strategy, starting with the most important parts.

3.1 Components

With respect to *components* we differentiate between *video data*, *interactional data*, and *metadata*.

Video data are the core element as all communication on YouTube is about single clips: They are searchable, titled, ranked, and commented on, and they are unambiguously locatable by a URL. Therefore, the basic communicative unit on YouTube is the single video accessible by its URL.

¹³ That is, we are not dealing with the question of how and with which tools the data are gathered and stored at this stage. This also means that we are not dealing with questions of implementation of a corpus and what a corpus should look like exactly on a description level in this paper. These aspects will be addressed in a next step, after we have specified what kind of data are needed to do research according to our approach.

Videos should be transcribed to allow a search for linguistic and multimodal elements (like words, phrases, gazes, visual elements, etc.). As transcription is a complex and time-consuming process, it is impossible to transcribe all aspects of a video right at the beginning. For practical research reasons, the transcription should therefore be successively refined, depending on the research question at hand. In a first step, a rather rough transcription should record the verbal exchange according to the transcription conventions of GAT2 (cf. Selting et al., 2011). Verbal transcriptions allow capturing the entire verbal exchange and thus have a clear cut outcome. Multimodal transcriptions, on the other hand, need to be neatly adjusted to the research question at hand. Not everything that is visually accessible can be part of the transcript. Multimodal transcriptions are, thus, in contrast to verbal transcriptions, highly selective (cf. Mondada, 2018; Stukenbrock, 2009).¹⁴ In addition, multimodal transcriptions are even more complex than verbal transcripts and should only be included if they are relevant for the research question at hand.

Therefore, in further steps, only selected sections are transcribed using a multimodal extended GAT2-system (Mondada, 2014, 2018). Multimodal transcripts of this kind represent different modal resources, e.g., talk, embodied conduct like gaze, gestures, posture shifts, etc., or game events on a screen in their temporal unfolding interplay. The transcripts can also include screenshots whose location is indicated within the transcript (Stukenbrock, 2009). The transcripts are only auxiliary means to produce a working document and to represent the exact temporal relations between different resources which are otherwise not accessible. However, audio and video data remains an indispensable basis and should not be replaced by the transcript at any stage in the process of analyzing. In our example above, as Figure 7 illustrates, besides the spoken material (here after the capital A), we also note game events (GE), game sounds (GS), status displays (StaD), and physical activities (FC) in case there is a visual representation of the gamer(s), a so-called facecam (FC).

As one can see in the transcript, when player A (*Gronkh*) is announcing that he is going to press a button for starting the game (okay i'm going to press ready now), the action of pressing has already been conducted as the 'ready-button' on his screen changes before A actually announces his action. This is conveyed by the status display line abbreviated with StaD. In addition, we see that in this stage, other modalities are either not yet available (like the facecam which is inserted later) or are in auto-play mode (like the music and the movements of the avatar). This

14 One reason for this is that in contrast to the verbal mode, the visual mode is not necessarily based on action units. Thus, with respect to the visual mode it is often unclear whether events are accountable actions or not.

Multimodal extended GAT2-transcript		
1	A	*(.) also ich drück jetzt ma auf FERTIG, *(.) okay* i_m going to press ready now
	GS	>>threatening music--->&
	StAD	*button READY changes
	GE	*avatar is moving automatically
	FC	{{no Facecam yet}}

Fig. 7: Example of a multimodal transcript.

observation can, additionally, be represented by a screenshot. Having access to the temporal unfolding of both talk and the effects of controlling actions represented on the screen allows us to see that A actually produces a retrospective comment in an announcement-like form.¹⁵ This sheds light on how gamers are dealing with time in coordinating different temporalities (of talking, gaming, and presenting). To add a relation to absolute time, the transcript may be extended by a timeline or time stamps (cf. Stukenbrock, 2009). Multimodal extended transcripts are an essential basis to pursue questions of temporal relations.¹⁶

In addition, YouTube videos are posted by a *sender* who communicates under a name, usually a pseudonym. The sender represents, together with the video, the basic pragmatic structure of YouTube communication. In our case, a sender, calling himself *Gronkh*, is releasing a video accessible via a unique address, i.e., a URL. The structure of YouTube suggests an understanding of the video as a communicative act of its releaser.

¹⁵ This, of course, only holds for the audio-visual presentation of the Let's Play, i.e., what viewers of the video get to see. Often there is a slight delay between facecam/multiplayer audio communication and the video feed. It would be interesting to integrate such questions of delay caused by processes of technical transmission. In principle, this would require an approach that also includes the integration of production data. At the current time, however, we focus only on the product.

¹⁶ There are, of course, other proposals for transcribing multimodal data (cf. Baldry & Thibault, 2006; Norris, 2004; Flewitt et al., 2009). Our approach proposes to use talk and pauses as a scaffold for the temporal alignment of other events. Transcripts of this kind focus on interaction and are useful to investigate dynamic, talk-based processes (Mondada, 2018). For other kinds of material, like static websites, they are less useful.

Another crucial element to be integrated in a YouTube corpus is *interactional data*. This concerns first of all the *comment section*, which we categorize as ‘first stage interactional data’ since comments are responsive actions either to the posted video or to other comments. The comments should be, at first, gathered as a coherent block. They need to be annotated following TEI standards¹⁷ for computer-mediated communication, especially regarding so-called interaction signs such as emojis, interaction words, and addressing terms (as proposed by Beißwenger et al., 2012).

A YouTube corpus should also integrate *metadata*. We distinguish three types of metadata: First, *automatically generated metadata* consisting mainly of the release date and the URL, which also contains the video ID. Secondly, *semi-automatically generated metadata*, which is generated by clicks and involves metrics like the number of views, likes and dislikes, or comments. Finally, *self-generated metadata* refers, for example, to the name of the releaser’s channel, the selected pseudonym or the short content description below the video.

However, the components outlined above—videos, interactional data, and metadata—are not isolated elements. Rather, they are embedded within the given structure of a webpage. This means that not only components but also their *relations* need to be considered when generating corpora of multimodal website data. The next section is devoted to this aspect of relations between elements.

3.2 Relations

Besides the elements listed und discussed above, their relations also have to be taken into account. Consequently, the visible structure of the webpage is important, as predefined slots have a meaning-making function for the text elements.¹⁸ Each webpage accessed on YouTube has a similar structure, thus entailing a certain recognition value. This means that before we discover the specific content of particular sections and its semiotic features, we already know the basic frame of meaning and its relation to other sections. This partly stems from the specific

¹⁷ TEI (Text Encoding Initiative) is a non-profit membership consortium. The goal of TEI is to develop a ‘set of high-quality guidelines for the encoding of humanities texts, and to support their use by a wide community of projects, institutions, and individuals’, which is based on ongoing research due to the dynamic textual domains that are still being explored (see <http://www.tei-c.org> for further information).

¹⁸ We are not going into analysis or interpretation at this stage. How relations between elements on a website are to be interpreted, and what particular meaning they may have, is not within the scope of this paper. The only thing which is relevant here is that they have a meaning-making function and that they have to be captured for this reason.

arrangement of elements. Figure 8 illustrates the typical structure of YouTube sites, their sections, and typical relations.¹⁹

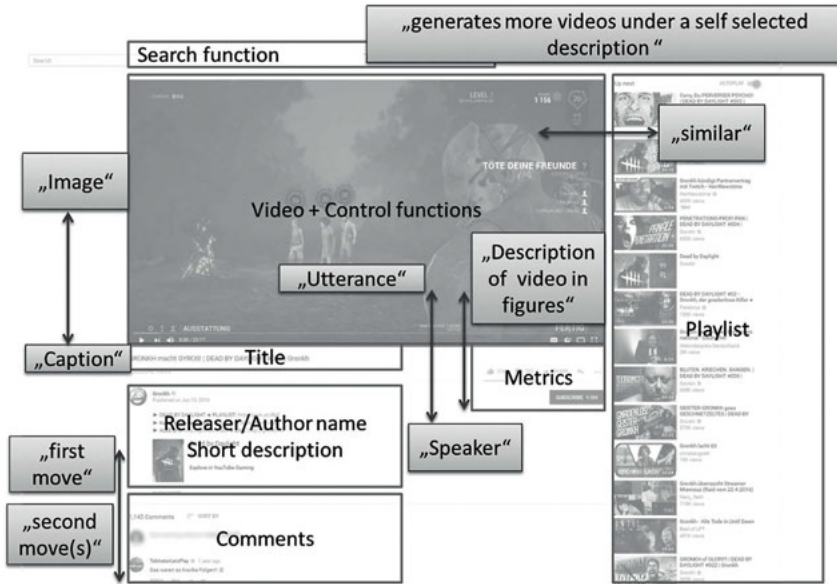


Fig. 8: Typical structure of YouTube pages, their sections, and typical relations.

As the diagram shows, all YouTube pages are composed of the same sections (indicated by the white boxes). Thus, we know the basic meaning of corresponding content, like ‘the title’ of the video, which is always placed below the video, ‘the comments’, which are placed at the bottom end of the webpage and may require scrolling and/or clicking to view all of them. In addition, sections are not only prefiguring a certain meaning, but also implicate meaningful relations between certain sections (indicated by the grey boxes and corresponding arrows). For example, the ‘video section’ and ‘the title section’ below are understood as an ‘image’ and its ‘caption’. Similarly, the metrics are taken as ‘describing the present video in figures’, the ‘play list section’ as ‘video suggestions similar to the present one’ and the ‘search section’ as a function to generate more video search results by using

¹⁹ The nature of those relations (logical, rhetorical, etc.) is of no relevance for our argumentation at this stage and only becomes relevant in the analysis. The examples only intend to illustrate that relations of this kind exist and that they should be represented in a corpus.

own search queries. The name in the ‘releaser section’ is read as ‘the speaker’ (the one to whom the video is ultimately attributed to as a communicative act) and the ‘comment section’ in relation to ‘the speaker section’ and its ‘video utterance’ as a ‘second move’ in relation to the video releaser’s ‘first move’.

Furthermore, a YouTube page contains several *active elements*, most importantly hyperlinks, which are understood as enabling specific kinds of actions, i.e., selecting and, by that, moving on to new content (Huber, 2002; Storrer, 2000). All in all, hyperlinks create a relational network of further potential contents in the background, which may or may not be activated by recipient’s selection. The hypertextual structure of websites, as Storrer (2000) has argued, amongst others, mean to be confronted with a user-generated ‘text’ as the linearity of traditional texts are replaced by respective individual ways of reading. Moreover, there are further active elements like buttons (such as, for instance, a subscribe-button), which can be understood as prompts to act in a certain way.

With respect to corpus building purposes, the hypertextual structure has to be preserved in order to represent basic interactional affordances of the website. This should be done in two ways. First the surface structure of the website should be captured in a way that allows identifying hyperlinks, e.g., as screenshots. Secondly, the covert deep structure of interconnected contents via hyperlinks should also be represented. Thus, for corpus generating purposes both screenshots of relevant websites and the underlying hypertextual structure should be integrated. How the latter can be represented is a technical problem which is not in the scope of this paper.

Finally, websites are not only hypertextual but furthermore they change due to individual reception (as for instance the viewing modes of the player) and due to user interfaces and devices (the website looks different on a computer compared to a smartphone, see Zichel, 2016). In addition, webpages are adapted to individual user habits and search histories based on website-specific algorithms. This variation is difficult to represent in a corpus. The only feasible solution is to define a default standard for representation (and, if relevant, possible variations).

Figure 9 provides a final overview of which elements should be included in a YouTube corpus.

4 Conclusion and Further Problems

What we have presented so far is work in progress. The background of our chapter is a larger project based at the Institute for German Language and the University of

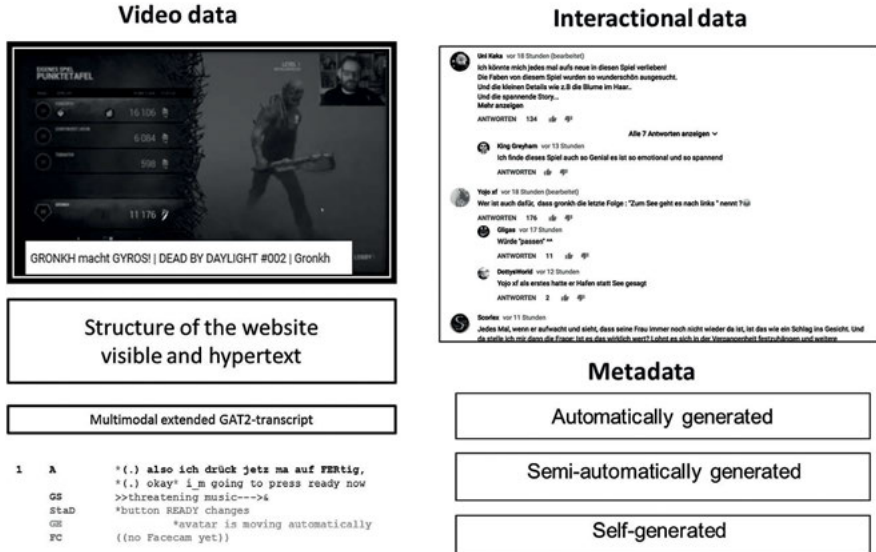


Fig. 9: Elements of a YouTube corpus.

Mannheim.²⁰ One aim of the project is, in a first step, to consider what a corpus of audio-visual, multimodal social media data should look like. Further steps concern questions around technical implementation, issues of data protection and privacy, issues of hosting and retrieval, and finally the actual construction of such a corpus. The present paper is a contribution to the first step: What kind of data are required in the corpus given our theoretical background? Our outline of a possible corpus construction was rooted in conceptual considerations about how communication and participation on YouTube are organized. In Part 1 of our chapter, we delineated levels of interaction: level of video interaction, sender-recipient-level, level of comments, and level of website-user interaction. As we are interested in interaction processes, corpus construction should entail data which are suitable for investigating these processes. Accordingly, in Part 2 we sketched out which elements and relations should make up such a corpus.

As YouTube communication is multimodal in nature, it should be represented as such on the data level. This means, as we outlined above, that all potentially relevant meaning-making processes should be captured both as single components

²⁰ The project envisages setting up a center of multimodality bringing together different researchers dealing with questions of multimodality and/or multimodal data.

(e.g., the video) and in their relatedness to surrounding components (e.g., the video embedded in a webpage). This is the only way to create a basis for reconstructing communication processes exhaustively.

There are still remaining problems. One of the main and most notorious problems in the area of corpus compilation is the dynamic nature of social media websites like YouTube. First, the *videos* as time-based media are dynamic. This problem can be partially solved by not only archiving the videos as films but by transcribing the videos including screenshots as mentioned above. In doing so, parts of the videos' content can be provided in a fixed and thus searchable way.

Secondly, *websites* change over time. Data collections are, therefore, always only snapshots of a certain moment in time. This can only be solved by multiple data acquisitions representing different stages in the 'life' of a website and its content.

Finally, websites are often *customized versions* adapted to individual users. However, since customization usually does not concern the core elements of the platform like the video or the comments, customized elements like individually adapted playlists can be disregarded initially. Moreover, they only appear in the screenshot versions of the website.

Further central problems are questions of *data protection* and, related to this, procedures of *anonymization*, as well as the *technical implementation* of data gathering and data archiving. This applies in particular to technical solutions for hosting, retrieving, mining, and processing YouTube data.

Appendix: Transcriptions Conventions

Speaker's signs/Display Screens/Facecams

- A/a **Gronkh**
- B/b **Pan**(dorya)
- C/c (Herr) **Curry**(wurst)
- D/d **Tobi**(nator)

Abbreviations for other events

- GE = Game Event
- GS = Game Sound
- StaD = Status Display

Conventions for the notation of physical activities (cf. Mondada, 2014)

Nonlinguistic events and activities

- appear after the abbreviations GE, GS, StaD and FC
- in lines following pauses or conversation activities (without own number)
- are aligned with conversation/pauses with the help of special characters (like \$, + etc.) indicating the beginning (simple sign) and the end (double sign) of events
- are assigned to the players (A/B/C/D; e.g. 'GEa' means 'game event on the screen of Gronkh', etc.)

Further conventions for the notation of physical movement

- > movement continues
- >\$ movement continues after the line until reaching
\$ the same sign
- >> continues after transcript ends
- >> starts before transcript

Special conventions for the notation of screen events

Abbreviations

- CS Counting Sign
- C(1-4) Characters/Avatars
- FC Facecam

Bibliography

- Abidi, Karima, Mohamed-Amine Menacer & Kamel Smaili. 2017. CALYOU: A Comparable Spoken Algerian Corpus Harvested from YouTube. In *18th Annual Conference of the International Communication Association (Interspeech)*, 3742–3746.
- Ackermann, Judith. 2016. *Phänomen Let's play-Video: Entstehung, Ästhetik, Aneignung und Faszination aufgezeichneten Computerhandelns*. Wiesbaden: VS Verlag.
- Adami, Elisabetta. 2009a. Do You Tube? When Communication Turns into Video e-Interaction. In Domenico Torretta, Marina Dossena & Annamaria Sportelli (eds.), *Migration of Forms, Forms of Migration: Atti de XXIII Convegno Nazionale AIA*, Bari: Progedit.
- Adami, Elisabetta. 2009b. 'We/YouTube': Exploring Sign-Making in Video-Interaction. *Visual Communication* 8. 379–399.
- Adami, Elisabetta. 2015. What I Can (Re)Make out of it: Incoherence, Non-Cohesion, and Re-Interpretation of YouTube Video Responses. In Marta Dynel & Jan Chovanec (eds.), *Participation in Public and Social Media Interactions*, 233–257. Philadelphia: John Benjamins.
- Androutsopoulos, Jannis. 2014. Linguaging when Contexts Collapse: Audience Design in Social Networking. *Discourse, Context & Media* 49(4). 290–309.

- Arminen, Ilkka, Christian Licoppe & Anna Spagnolli. 2016. Respecifying Mediated Interaction. *Research on Language and Social Interaction* 49(4). 290–309.
- Baldry, Anthony & Paul J. Thibault. 2006. *Multimodal Transcription and Text Analysis: A Multimedia Toolkit and Coursebook with Associated On-Line Course*. London and New York: Equinox.
- Bateman, John A., Janina Wildfeuer & Tuomo Hiippala. 2017. *Multimodality – Foundations, Research and Analysis. A Problem-Oriented Introduction*. Berlin: Mouton de Gruyter.
- Beißwenger, Michael. 2009. Multimodale Analyse von Chat-Kommunikation. In Karin Birkner & Anja Stukenbrock (eds.), *Die Arbeit mit Transkripten in Fortbildung, Lehre und Forschung*, 117–143. Mannheim: Verlag für Gesprächsforschung.
- Beißwenger, Michael, Maria Ermakowa, Alexander Geyken, Lothar Lemnitzer & Angelika Storrer. 2012. A TEI Schema for the Representation of Computer-Mediated Communication. *Journal of the Text Encoding Initiative* 3. 1–36.
- Bou-Franch, Patricia & Pilar Garcés-Conejos Blitvich. 2014. Conflict Management in Massive Polylogues: A Case Study from YouTube. *Journal of Pragmatics* 73. 19–36.
- Bou-Franch, Patricia, Nuria Lorenzo-Dus & Pilar Garcés-Conejos Blitvich. 2012. Social Interaction in YouTube Text-Based Polylogues: A Study of Coherence. *Journal of Computer-Mediated Communication* 17. 501–521.
- Boyd, Michael S. 2014. (New) Participatory Framework on YouTube? Commenter Interaction in US Political Speeches. *Journal of Pragmatics* 72. 46–58.
- Burgess, Jean & Joshua R. Greenberg. 2014. *YouTube: Online Video and Participatory Culture*. Cambridge: Polity.
- Clayman, Steven E. 2006. Understanding News Media: The Relevance of Interaction. In Paul Drew, Geoffrey Raymond & Geoffrey Darin (eds.), *Talk and Interaction in Social Research Methods*, 135–154. London: Sage.
- D'hondt, Sigurd, Jan-Ola Östman & Jef Verschueren (eds.). 2009. *The Pragmatics of Interaction. Handbook of Pragmatics*. Philadelphia: John Benjamins.
- Dürscheid, Christa. 2005. Medien, Kommunikationsformen, kommunikative Gattungen. *Linguistik online* 22(1). 3–16.
- Dynel, Marta. 2014. Participation Framework underlying YouTube Interaction. *Journal of Pragmatics* 73. 37–52.
- Dynel, Marta & Jan Chovanec. 2015. *Participation in Public and Social Media Interactions*. Philadelphia: Benjamins.
- Eisenlauer, Volker. 2014. Facebook as a Third Author – (Semi-)Automated Participation Framework in Social Network Sites. *Journal of Pragmatics* 72. 73–85.
- Flewitt, Rosie, Regine Hampel, Mirjam Hauck & Lesley Lancaster. 2009. What are Multimodal Data and Transcription? In Carey Jewitt (ed.), *The Routledge Handbook of Multimodal Analysis*, 40–53. London: Routledge.
- Frobenius, Maximiliane. 2013. Pointing Gestures in Video Blogs. *Text & Talk* 33. 1–23.
- Frobenius, Maximiliane. 2014. Audience Design in Monologues: How Vloggers Involve their Viewers. *Journal of Pragmatics* 72. 59–72.
- Frobenius, Maximiliane, Volker Eisenlauer & Cornelia Gerhardt. 2014. Special Edition: Participation Framework Revisited: (New) Media and their Audiences/Users. *Journal of Pragmatics* 72. 1–90.
- Gibson, James J. 1979. *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.
- Goffman, Erving. 1981. Footing. In Erving Goffman (ed.), *Forms of Talk*, 124–159. Philadelphia: University of Pennsylvania Press.

- Goffman, Erving. 1983. The Interaction Order. *American Sociological Review* 48. 1–17.
- Goodwin, Charles. 1986. Audience Diversity, Participation and Interpretation. *Text* 6(3). 283–316.
- Goodwin, Charles & Marjorie Harness Goodwin. 2004. Participation. In Alessandro Duranti (ed.), *A Companion to Linguistic Anthropology*, 222–244. Malden, Mass: Blackwell.
- Gronkh. 2016. *Multiplayer Let's Play "GRONKH macht GYROS– DEAD BY DAYLIGHT #002 – Gronkh*. YouTube. https://www.youtube.com/watch?v=ikDqiW_DRqc. Last accessed: 17 September 2019.
- Habermas, Jürgen. 1967. *Zur Logik der Sozialwissenschaften*. Tübingen: Mohr.
- Habscheid, Stephan. 2000. 'Medium' in der Pragmatik. Eine kritische Bestandsaufnahme. *Deutsche Sprache* 28. 126–143.
- Hale, Thomas. 2013. From Jackasses to Superstars: A Case for the Study of 'Let's Play'. Last accessed: 17 September 2019. http://www.academia.edu/5260639/From_Jackasses_to_Superstars_A_Case_for_the_Study_of_Let_s_Play_September_2013.
- Heritage, John. 1985. Analyzing News Interviews. Aspects of the Production of Talk for an Overhearing Audience. In Teun Adrianus van Dijk (ed.), *Handbook of discourse analysis*, 95–117. London: Academic Press.
- Holly, Werner. 1997. Zur Rolle von Sprache in Medien. Semiotische und kommunikationsstrukturelle Grundlagen. *Muttersprache* 19. 215–229.
- Horton, Donald & Richard R. Wohl. 1956. Mass Communication and Para-Social Interaction: Observations on Intimacy at a Distance. *Psychiatry* 19. 215–229.
- Huber, Oliver. 2002. *Hyper-Text-Linguistik. TAH: Ein textlinguistisches Analysemodell für Hypertexte. Theoretisch und praktisch exemplifiziert am Problemfeld der typisierten Links von Hypertexten im WWW*: Universität München dissertation.
- Hutchby, Ian. 2006. *Media Talk. Conversation Analysis and the Study of Broadcasting*. Maidenhead: Open University Press.
- Ivkovic, Dejan. 2013. The Eurovision Song Contest on YouTube: A Corpus-Based Analysis of Language Attitudes. *Language & Internet* 10(1). 1–25.
- Jones, Graham & Bambi Schieffelin. 2009. Talking Text and Talking Back: 'My BFF Jill' from Boob Tube to YouTube. *Journal of Computer-Mediated Communication* 1050–1079.
- Knorr-Cetina, Karin. 2009. The synthetic situation: Interactionism for a global world. *Symbolic Interaction* 32. 61–87.
- Konijn, Elly A., Sonja Utz, Martin Tanis & Susan B. Barnes. 2008. *Mediated Interpersonal Communication*. London and New York: Routledge.
- Kress, Gunther. 2010. *Multimodality: A Social Semiotic Approach to Contemporary Communication*. London: Routledge.
- Kress, Gunther & Theo van Leeuwen. 2006 [1996]. *Reading Images: The Grammar of Visual Design*. London and New York: Routledge.
- Levinson, Stephen C. 1988. Putting Linguistics on a Proper Footing: Explorations in Goffman's Concepts of Participation. In Paul Drew & Anthony Wootton (eds.), *Erving Goffman. Exploring the Interactional Order*, 161–227. Cambridge: Polity Press.
- Marx, Konstanze & Axel Schmidt. 2018. Let's Play (together) oder schau mal, wie ich spiele – (Interaktive) Praktiken der Attraktionssteigerung auf YouTube. In Konstanze Marx & Axel Schmidt (eds.), *Interaktion und Medien – Interaktionsanalytische Zugänge zu medienvermittelter Kommunikation*, 319–352. Mannheim: OraLingua.
- Marx, Konstanze & Axel Schmidt. forthcoming. Making Let's Plays Watchable: An Interactional Approach to Multimodality. In Crispin Thurlow, Christa Dürscheid & Federica Diémoz (eds.), *Visualizing (in) the New Media*, London: John Benjamins.

- Maybury, Mark T. 2012. *Multimedia Information Extraction: Advances in Video, Audio and Imagery Analysis for Search, Data Mining, Surveillance, and Authoring*. Hoboken and New Jersey: Wiley.
- Meyrowitz, Joshua. 1990. Redefining the Situation: Extending Dramaturgy into a Theory of Social Change and Media Effects. In Stephen H. Riggins (ed.), *Beyond Goffman: Studies on Communication, Institution, and Social Interaction*, Berlin: Mouton De Gruyter.
- Mondada, Lorenza. 2014. Conventions for Multimodal Transcription. Last accessed: 17 September 2019. https://franzoesistik.philhist.unibas.ch/fileadmin/user_upload/franzoesistik/mondada_multimodal_conventions.pdf.
- Mondada, Lorenza. 2018. Multiple Temporalities of Language and Body in Interaction: Challenges for Transcribing Multimodality. *Research on Language and Society* 51. 85–106.
- Norris, Sigrid. 2004. *Analyzing Multimodal Interaction: A Methodological Framework*. London and New York: Routledge.
- O’Keeffe, Anne. 2007. *Investigating Media Discourse*. London: Routledge.
- Recktenwald, Daniel. 2017. Toward a Transcription and Analysis of Live Streaming on Twitch. *Journal of Pragmatics* 115. 68–81.
- Scannell, Paddy. 1991. Introduction: The Relevance of Talk. In Paddy Scannell (ed.), *Broadcast Talk*, London: Sage.
- Schmidt, Axel. 2011. How to Deal Methodologically with Entertaining Hatred and Aggressive Humor on the Web (and Television). *Studies in Communication Sciences* 11(2). 133–166.
- Schmitz, Ulrich. 2015. *Einführung in die Medienlinguistik*. Darmstadt: WBG.
- Selting, Margret, Peter Auer, Dagmar Barth-Weingarten, Jörg Bergmann, Pia Bergmann, Karin Birkner, Elizabeth Couper-Kuhlen, Arnulf Deppermann, Peter Gilles, Susanne Günthner, Martin Hartung, Friederike Kern, Christine Mertzlufft, Christian Meyer, Miriam Morek, Frank Oberzaucher, Jörg Peters, Uta Quasthoff, Wilfried Schütte, Anja Stukenbrock & Susanne Uhmann. 2011. A System for Transcribing Talk-in-Interaction: GAT 2, translated and adapted for English by Elizabeth Couper-Kuhlen and Dagmar Barth-Weingarten. *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion* 12. 1–15.
- Siersdorfer, Stefan, Sergiu Chelaru, Jose San Pedro, Ismail Altinogvde & Wolfgang Nejdl. 2014. Analyzing and Mining Comments and Comment Ratings on the Social Web. *ACM Transactions on the Web (TWEB)* 8. 17.
- Starbreeze Studios. 2017. *Dead by Daylight*. Starbreeze Studios.
- Stephan, Heike. 2014. *Let’s Plays: Kommentierte Spielevideos und ihre Auswirkungen auf die Spielemagazine*. Hamburg: Diplomica-Verlag.
- Storrer, Angelika. 2000. Was ist “hyper” am Hypertext? In Werner Kallmeyer (ed.), *Sprache und neue Medien*, 222–249. Berlin and Boston: De Gruyter.
- Stukenbrock, Anja. 2009. Herausforderungen der multimodalen Transkription: Methodische und theoretische Überlegungen aus der wissenschaftlichen Praxis. In Karin Birkner & Anja Stukenbrock (eds.), *Die Arbeit mit Transkripten in Fortbildung, Lehre und Forschung*, 145–169. Mannheim: Verlag für Gesprächsforschung.
- Tereick, Jana. 2013. Die Klimalüge auf YouTube: Eine korpusgesteuerte Diskursanalyse der Aushandlung subversiver Positionen in der partizipatorischen Kultur. In Claudia Fraas, Stefan Meier & Christian Pentzold (eds.), *Online Diskurse. Theorien und Methoden transmedialer Online-Diskursforschung*, 226–257. Köln: Halem.
- Tereick, Jana. 2016. *Klimawandel im Diskurs: Multimodale Diskursanalyse crossmedialer Korpora*. Berlin and Boston: De Gruyter.

- Thompson, John. 1995. *The Media and Modernity: A Social Theory of the Media*. Cambridge: Polity Press.
- Tolson, Andrew. 2010. A New Authenticity? Communicative Practices on YouTube. *Critical Discourse Studies* 7(4). 277–289.
- Uryupina, Olga, Barbara Plank, Aliaksei Severyn, Agata Rotondi & Alessandro Moschitti. 2014. SenTube: A Corpus for Sentiment Analysis on YouTube Social Media. *LREC* 4244–4249.
- Vonderau, Patrick. 2016. The Video Bubble: Multichannel Networks and the Transformation of YouTube. *Convergence: The International Journal of Research into New Media Technologies* 22(4). 361–375.
- Vorderer, Peter (ed.). 1996. *Fernsehen als 'Beziehungskiste': Parasoziale Beziehungen und Interaktionen mit TV-Personen*. Opladen: Westdeutscher Verlag.
- Zhang, Wei & Cheri Kramarae. 2014. "SlutWalk" on Connected Screens: Multiple Framings of a Social Media Discussion. *Journal of Pragmatics* 73. 66–81.
- Zichel, Jana. 2016. *Sprache im Layout. Analyse von kohäsionsstiftenden Mitteln auf Websites aus internetlinguistischer Perspektive*. Technische Universität Berlin MA thesis.

Lauren O'Hagan

Class, Culture, and Conflict in the Edwardian Book Inscription: A Multimodal Ethnohistorical Approach

Abstract: This study uses three examples of Edwardian (1901–1914) book inscriptions—a prize inscription, gift inscription, and bookplate—to demonstrate how the adoption of an ethnohistorical approach, in which choices of image, color, typography, and materiality are grounded in archival research, can strengthen multimodal analysis. Furthermore, it argues that, while book inscriptions may seem insignificant markers of ownership, they, in fact, act as a material microcosm of many of the social tensions that existed between class groups in early twentieth-century Britain. The analysis reveals that inscriptions were primarily used to objectify their owners' economic means and cultural necessities, and assert themselves in a social space, whether to uphold their rank or keep their distance from other groups. These findings demonstrate the importance of embedding hypotheses concerning the function and form of artifacts in concrete historical documents.

Keywords: multimodality, ethnohistory, book inscriptions, class conflict, culture

1 Introduction

In the past thirty years, new literacy studies (NLS) has brought attention to the importance of writing as a social practice that is embedded in power relations (e.g., Street, 1984; Barton & Hamilton, 1998). A substantial component of NLS has been its focus on 'ordinary writing', that is, "writing that is typically unseen or ignored and is primarily defined by its status as discardable" (Sinor, 2002, 5). When viewed from a social perspective, ordinary writing has the ability to reveal meaningful information about how individuals and social groups organize their lives and make sense of their experiences, and how culture and knowledge is produced and reproduced. Book inscriptions, defined as "ownership marks or annotations present on the front endpapers or title pages of a book" (O'Hagan, 2018, 44), are a type of 'ordinary writing' that have been largely underexplored due to their appearance as seemingly insignificant markers of ownership. When investigated within the context of the Edwardian era (1901–1914), however, book inscriptions act as a microcosm that reflect with an unusual intensity, the social conflicts and tensions that existed between class groups in early twentieth-century Britain.

<https://doi.org/10.1515/9783110608694-006>

Typically, ordinary writing has been examined using ethnographic or ethno-historical approaches. These approaches involve living with a community and observing their habits and customs or using archival material to explore said community and its practices, respectively. Ethnohistory as a field focuses on groups of people whose perspectives are underrepresented in official narratives of history backed by national institutions of power (Faudree & Pharaoh Hansen, 2013, 240). By emphasizing these historically disenfranchised groups, ethnohistory provides an alternative to the dominant perspective, filling in what patriarchal accounts of history have excluded and creating a broader view of social history with a better understanding of the role of 'forgotten people' in social change. While other disciplines may draw upon historical data when attempting to understand the ideologies, cultures, and traditions of a particular social practice (i.e., cultural studies, aesthetics), the advantage of ethnohistory is the precedence that it gives to the oppressed, thus making them subjects of formal historical analysis in their own right.

In the context of multimodal studies, despite Kress's well-established claim that "all texts have always been multimodal" (Kress, 1996, 20), there still remains a heavy focus on digital media and other forms of modern technology (i.e., Ravelli & van Leeuwen, 2018; Benson, 2017) as opposed to historical texts. This is particularly the case with the growing field of multimodal digital humanities (Berry, 2012; O'Halloran, 2015, e.g.), which has tended only to "leverage the potential of the visual and aural media that are part of contemporary life" (Svensson, 2010, 57). Furthermore, many of the studies that use historical data (e.g., O'Toole, 1994; Kress & van Leeuwen, 2006 [1996]; Granelli & Martinez-Hinarejos, 2016) seem to have neglected the valuable role that archival documents can play when exploring power dynamics in text creation and production. The introduction of an ethnohistorical approach to multimodality can not only bring about a greater recognition of the importance of grounding choices of image, color, typography, and materiality in archival research, but it can also reframe power relations, particularly favoring the perspective of marginalized groups over the controlling classes. This can, in Samuel's words, "reconstitute the vanished components of the world we have lost" (Samuel, 1988, 43). In addition, this approach also has the potential to lay the groundwork for digital archives that consist of historical examples of multimodal texts, thus offering new ways to collect, explore, and interpret data.

Despite the fact that the two approaches have their origins in different research domains, ethnohistorical methodologies have many objectives similar to the social semiotic paradigm within multimodality: both focus on the range of social and cultural resources that are available to a person in a specific context; both draw attention to the motivations that influence a person's selection from these resources; and both accentuate the social effects that these resources may have.

What ethnohistory adds, however, is a focus on “real life rather than abstractions, with ordinary people rather than privileged elites, with everyday things rather than sensational events” (Samuel, 1988, 42) when carrying out analysis.

This means that it allows more flexibility in interpretation, as it acknowledges that texts do not have fixed meanings and are often influenced by the subtleties of power, ideology, and cultural distinction. Moreover, when exploring notions of class conflict, ethnohistory incorporates a broad range of theory from sociology, philosophy, and cultural studies, which can provide multimodality with working hypotheses, avenues of approach, and problem areas that can be refined and tested. Multimodality, on the other hand, can strengthen ethnohistory through the provision of robust methods of analyses with their own established terminology, thus downplaying the possibility of merely anecdotal findings.

In this study, three examples of Edwardian (1901–1914) book inscriptions are used to demonstrate how the co-application of a social semiotic approach to multimodality and ethnohistory can enrich our understanding of historical artifacts, particularly in terms of social class. While the Edwardian era was not the first time that people used book inscriptions, it is perhaps the most interesting period in which to explore this practice, as both increasing literacy and the dramatic decrease in the price of books enabled all classes of society to own them for the first time. This means that these examples of ‘ordinary writing’ can be explored from the perspectives of all class groups rather than from an upper-class bias. With this in mind, all examples have been collected from second-hand bookshops, given the claim that most archives, libraries, and personal collections remain heavily biased towards the writing of the elite or distinctive educated individuals (see Gillen & Hall, 2010, 170).

Supporting multimodal analysis with archival documents, such as censuses, vital certificates, and military records, enables inscriptions to be made sense of within a larger and broader context of patterned practices and sociopolitical forces. This will allow social semiotic approaches to multimodality to move beyond text-centered analyses, as hypotheses concerning the function and form of artifacts can be derived and explored from concrete historical documents.

Section 2 outlines the benefits of adopting an ethnohistorical approach to multimodality. Then, Section 3 introduces the case study of Edwardian book inscriptions, providing background information on their uses and meanings. In Section 4, the ethnohistorical methodology employed for this study will be described, while Section 5 involves an analysis of three examples from the dataset to demonstrate how this multimodal ethnohistorical framework can be employed. Finally, Section 6 describes the theoretical conclusions of this study.

2 Towards an Ethnohistorical Approach to Multimodality

Kress and van Leeuwen's *Reading Images: The Grammar of Visual Design* (2006 [1996]) is one of the most widely-received works in multimodality. They conceptualize semiotic resources as interrelated systems of meaning which together constitute and manifest culture (O'Halloran et al., 2019, 7). These systems are organized according to three metafunctions: the representational, the interpersonal, and the compositional, each of which roughly corresponds to the three Hallidayan metafunctions (ideational, interpersonal, textual) in systemic-functional linguistics.

While Kress and van Leeuwen's (2006 [1996]) social semiotic framework is useful for examining the ways in which images communicate meaning, it has received criticism from a range of scholars for its reliance on small datasets that offer limited empirical evidence (Bezemer & Jewitt, 2010, 194) and its neglect of genre conventions (Bateman, 2008, 46), sociocultural context (Durie, 1997, 92), and comparison of modes (Hiippala, 2015, 25). Furthermore, due to its text-centeredness, it can often result in subjective analyses that give "a post hoc rationalisation of design decisions" (Bateman et al., 2004, 67) that may have occurred for other reasons independent of the image itself. When exploring historical artifacts, such as book inscriptions, these issues are particularly relevant, as they risk obscuring the people involved in their production and downplaying the complexities of the Edwardian sociopolitical landscape. Adopting an ethnohistorical approach offers one solution to improve multimodal analysis, as it provides a model that is built on primary evidence and foregrounds social practices as being anchored in the systems and the institutions of the social world (O'Hagan, 2019).

Ethnohistorical methodologies were first used in the 1930s by Fritz Röck to explore African culture through historical artifacts, but they became widely employed in the United States in the 1950s as a result of the Indian Claims Act of 1946, which sought to give voice to the claims of Native American tribes over land. In the field of linguistics, ethnohistorical approaches were largely pioneered by Dell Hymes (1962) under the umbrella term 'linguistic anthropology' through which he proposed an 'ethnography of communication' as an approach towards analyzing patterns of language use within speech communities. Despite the fact that many of Hymes's research methods are ethnohistorical in nature, the term has not gained widespread usage amongst linguists. Instead, the method is generally referred to as an "ethnography of documents" (Laurier & Whyte, 2001, 4), the "anthropology of writing" (Barton & Papen, 2010, 3), or simply, a "historical approach to ethnography" (Gillen, 2013, 491). Within the context of this study, I have chosen to reinstate

the anthropological term ‘ethnohistory’, as the definition provided by Faudree & Pharaoh Hansen (2013) seems to best encompass its aims and research methods:

Ethnohistory—understood as the histories of indigenous people, ethnic minorities and marginalized genders or classes—is a field where attention to language has been employed successfully to construct complex pictures of past sociality. The field differentially integrates methods and theories from a diverse set of disciplines, including social history, historical linguistics, linguistic anthropology and critical theory (Faudree & Pharaoh Hansen, 2013, 240).

In recent years, there has been a growing interest in semiotic perspectives on ethnography. Katriel (2015, 458), for example, has suggested that ethnographic methodologies must begin to consider interconnections between the temporality, performativity, and materiality of communication. However, the most fervent supporters of the co-application of multimodal and ethnographic methodologies remain those working in NLS, such as Pahl & Rowsell (2006). They have argued that, despite often being considered two separate research domains, multimodality and ethnography should be viewed as complementary frameworks for investigating the social practices of reading and writing, as both share a view of texts as material and situated, and both use specific research tools to trace social practices and contexts.

This study argues that the co-application of multimodal and ethnohistorical approaches brings further advantages. First, ethnohistory provides multimodality with concrete evidence to support analysis and explore texts within a specific time period. This means that arguments can be rooted in historical concreteness and the context of wider sociopolitical forces, thus ensuring that generalizations are not made until sufficient evidence is provided (Axtell, 1979; Rowsell & Chen, 2011). Furthermore, as Axtell (1979, 5) claims, ethnohistory also has the advantage of being able to move both forwards and backwards in time. This enables cultural patterns to be explored in their original historical context of use, while also informing current and future practices.

In the case of social class (the focus of this study), this is especially helpful, given the continued disparities between certain economic and social groups in British society. Another advantage of ethnohistory is that its focus on the under-represented allows explorations of power, ideology, and cultural distinction to be carried out from a perspective that empowers such groups to claim their history back as their own. In this way, it provides a new panorama of their lives and struggles that is not clouded by bias judgments made by “privileged white people” (Sheehan, 1969, 269). Overall, through blending synchronic analysis with diachronic evidence, a multimodal ethnohistorical approach can facilitate the reconstruction of cultural practices, therefore demonstrating that signs do not

exist in vacuity, but instead are “shaped by the histories and values of societies and their cultures” (Kress & van Leeuwen, 2006 [1996], 34).

3 Edwardian Book Inscriptions as a Case Study

Book inscriptions have a long history in Britain, dating back to the Medieval age when scribes would freight precious manuscripts with curses to discourage thievery. However, they became popular in the nineteenth century as a result of Victorian commodity culture and Britain's growing obsession with portable property, defined by Plotz (2008, 2) as “everyday culture-bearing objects”. By the beginning of the twentieth century, the book occupied a chief position in the households of all classes of society and “hardly a family [...] was without its little shelf of books and its sheaf of current periodicals” (Altick, 1957, 5). Book owners used inscriptions as social artifacts loaded with symbolic capital that indicated the amount of prestige that they held (or wished to hold) within Edwardian society (Hammond, 2006, 194). These inscriptions ranged from the most rudimentary signature or initials of the owner to an elaborate, custom-designed bookplate or Sunday school prize sticker.

Unlike other ownership marks, book inscriptions should not be confined to the status of a primary impulse or proprietary instinct of claiming an object as one's own; instead, they should be viewed as registers of the cultural and social situation in which the owner and the book met. Book inscriptions act as forms of ‘disembodied language’ that transform past events into artefacts, make former presences known, and produce speech acts that invite readerly projection (Crain, 2016, 145). When examined in detail, the power dynamics involved in their creation also comes to the forefront, both from the perspective of the owner and the ideological constructs that shaped the Edwardian world view and system of ideas. These dynamics often disclose information about the social status of the inscriber and the “face” (Goffman, 1959) they wish to present to those who come into contact with the inscription.

Despite their possibility of revealing vast information about a particular community's literacy, cultural and social practices, book inscriptions remain largely underresearched. Thus far, it is only within the field of provenance studies that they have been explored in any depth, with a particular focus on the ownership inscriptions of wealthy or famous historical figures (e.g., North-Lee, 1979; Pearson, 1998). Furthermore, most literature available on book inscriptions predates the Edwardian era (e.g., Castle, 1892; Hamilton, 1895), as the late nineteenth century was a time in which great public interest first arose in the topic. No prior attempts

have been made to investigate the semiotic features of Edwardian book inscriptions, nor to consider their role as indicators of social class. Thus, it is worthwhile exploring the potential of inscriptions as new primary resources to explore class struggles in early twentieth-century Britain.

As a case study, three book inscriptions have been selected for multimodal ethnohistorical analysis. These inscriptions come from a wider dataset of 3,000 Edwardian book inscriptions that were collected from books in the second-hand bookshop Bookbarn International, in Somerset, England. As most of these book inscriptions are rare, only exist in a single instance, and are not protected in official archives or libraries, they have been collected and preserved through digital remediation. They are now part of an ongoing project that aims to establish a permanent digital archive that can safeguard them for the future. All of the collected inscriptions were written between 1901 and 1914. Furthermore, all 3,000 inscribers have been investigated using census records and assigned to a class group (underclass, working-class, lower-middle-class, upper-middle-class, upper-class) based on five criteria: occupation, father's occupation, address, family size, and number of infant mortalities. This has enabled class-based patterns to be established in terms of inscription types and their semiotic features. The three book inscriptions chosen for analysis in this study have been selected because they represent the three most commonly occurring inscription types in the dataset (i.e., prize inscription, gift inscription, bookplate). Furthermore, their semiotic features make them prototypical examples of a working-class, lower-middle-class, and upper-class Edwardian inscription. Thus, their meanings and functions can be said to be representative of other similar inscription types in the larger dataset.

4 Ethnohistorical Methodology

Figure 1 shows an outline of the ethnohistorical methodology that was adopted for this study. As is typical of studies that are rooted in ethnohistory, this methodology incorporates the data collection and analysis processes that were undertaken.

Stage 1 involved the collection of the 3,000 inscriptions, which took place over a nine-month period in Bookbarn International. All Edwardian book inscriptions, as well as the books in which they were found and any other interesting features, such as booksellers' labels, later inscriptive marks and advertisements, were photographed and field notes were taken on each inscription and the bibliographical details of each book. In Stage 2, all the documentary photographs were uploaded to a computer, edited with Adobe Photoshop and coded with a unique identifying

label according to their inscription type. The field notes were then stored digitally in an Access database.

In Stage 3, the genealogical website www.ancestry.com was used to access original historical records for each book owner, including censuses, birth, marriage and death certificates, street directories, military lists, and immigration/travel documents. This information helped to locate the inscriptions within a clear context and account for the influence of social conventions and norms on design choices. For the Edwardian period, the 1911 census was particularly useful, as it provides detailed information on the owner's age, gender, address, profession, marital status, siblings, children, place of birth, nationality, and infirmities. On the basis of this information, each inscriber was assigned to a class group based on the five criteria outlined in the previous section. This assignment was also supported by Charles Booth's 1903 Poverty Maps¹, which classify streets into seven colors from black to yellow based on class and wealth. *The Times*, *Illustrated London News*, *The Boys' Own Journal* and *The Girls' Own Journal* were also used to collect institutional information on the companies in which each owner was employed and the social clubs that they attended, while the *Archive of British Publishing and Printing* at the University of Reading was used to obtain data on the artists, engravers, printers, stationers, or booksellers involved in the book production and inscription process.

Stage 4 involved gathering broader information on the sociopolitical context of Edwardian Britain. Using information provided by newspapers on key events that were happening in the world at the same time as the inscription was being produced, for example, may affect the interpretation of the inscription (then and now), as well as the original intentions of the owner. Furthermore, researching the various components of the inscription process (i.e., design, engrave, print, etc.) may help to establish the composition of relationships between inscribers, as well as how acts were ordered within the process of inscription as books moved from initial purchase to frequent usage to afterlife.

At this stage, the linguistic form of the inscriptions was also taken into consideration, particularly the type of language, spelling, and register used by the owner and what this may reveal about their social status and education. Despite being a written genre, inscriptions can also make use of paralinguistic and prosodic features through choices of typography, color, and picture. Form may also reveal selection rules that govern the use of particular message forms when a choice is made between possible alternatives (i.e., whether the owner uses their full name, a nickname, an honorific, etc.). Acknowledging the performative function of inscriptions as forms of cultural capital gives additional meaning to the semiotic and

¹ <https://booth.lse.ac.uk>, last accessed: 29 August 2019

material choices of book owners and suggests that particular design choices may have been influenced more by owners' social goals than strict rules of composition.

Stage 5 involved a preliminary multimodal analysis of each inscription. In all inscriptions, the style and cultural connotation of typeface (van Leeuwen, 2006) was considered, as well as the value and modulation categories of the semiotics of color (van Leeuwen, 2011). The distinctive physical qualities of typeface and its specific intentions, inherent associations, and cultural references were often used by owners to reflect particular aspects of their personal identity, while choices of colors were strongly influenced by culturally symbolic meanings or, in the case of some bookplates, by the norms of heraldry. The inscriptions are also investigated in reference to their material features. The semiotics of materiality was developed in my MA dissertation on Edwardian bookplates (O'Hagan, 2015), and is made up of writing implement, printing technique, paper, and physical setting. Here, it also encompasses the separate category of texture (Djonov & van Leeuwen, 2011), as texture in book inscriptions can be a tangible or metaphorical property (i.e., they could be printed on paper, leather and velvet, or shading and tones could be used to reflect a particular sociocultural connotation). Any images were explored using the representational, interpersonal, and compositional metafunctions (Kress & van Leeuwen, 2006 [1996]) to determine the use of participants, salience, modality, visual framing, and distribution of information value.

The physical copies of the books in which the inscriptions were found were also considered because, by the Edwardian era, publishers were producing books in a range of formats, bindings, and paper types to appeal to all class groups in society (Lerer, 2012, 127). Therefore, a book's format, paper type, typeface, covers, and spine, for example, can offer valuable clues into the social status and wealth of the book owner. Furthermore, booksellers' stamps can indicate where books were purchased, while publisher's advertisements and promotional dossiers can highlight a book's intended audiences. This data can provide primary evidence to reduce the potential subjectivity of multimodal analysis.

The final stage combined the previous three stages to carry out a detailed multimodal ethnohistorical analysis of a selection of inscriptions from the dataset. For the present analysis, one prototypical inscription from the three most frequently occurring inscription categories—a prize inscription, gift inscription, and bookplate—was chosen for analysis. Prototypicality was based on the definition of items in a category that contain the most central features (Rosch, 1975).

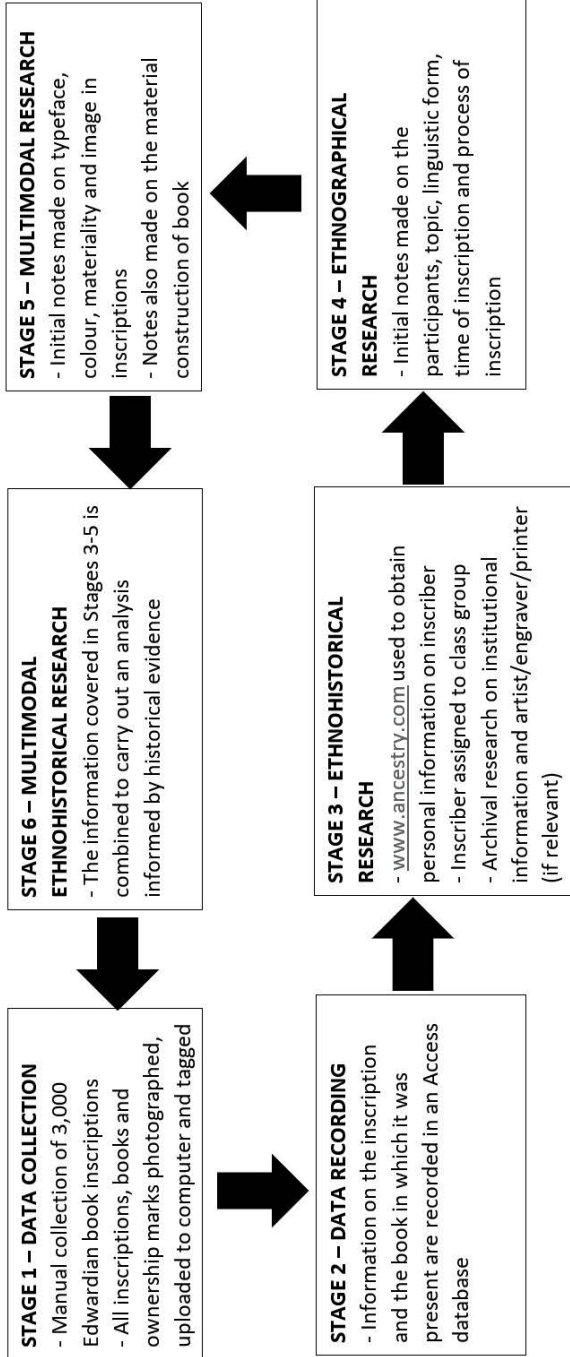


Fig. 1: Ethnohistorical methodology.

5 Book Inscription Analysis

In this section, three inscriptions will be studied using multimodal ethnohistorical analysis. The analysis brings together an exploration of the semiotic features of each inscription, as well as the sociocultural context and personal backgrounds of all participants involved in their creation.

5.1 Prize Inscription

A prize inscription is the name given to any mark in a book which indicates that the book was awarded as a school or Sunday school prize. In Edwardian Britain, awarding books as prizes had become standard practice for most schools, Sunday schools, and other institutions. While prize books were typically awarded to a person in recognition of an outstanding achievement or contribution, they also served a secondary function of moral education and they were often used by educational and religious institutions as tools to disseminate approved fiction to working-class children. Bodmer (1999, 137) claims that the prize book reinforced the power structure of old and young, while Grenby (2011, 174) argues that it advocated the idea that books were to be bestowed on recipients as something that had to be earned. For this reason, I consider prize inscriptions to be a form of ‘imposed ownership’, a type of ritual communication with a strong control function that was enforced by authority figures.

The prize inscription in Figure 2 was awarded to Katie Cowell, a member of the working-class, by Ballaugh Primitive Methodist Sunday School, and it featured inside the religious novel *Broken Barriers* by Bessie Marchant.

Primitive Methodism was one of the religious denominations in Edwardian Britain that was most closely associated with the working classes (Calder, 2016, ix). It played an important role in the formative phase of the Trade Union movement in England, and was particularly widespread on the Isle of Man, where Ballaugh is located. At the time of inscription, Cowell was 19 years old and working as a servant for a Church of England clergyman. The fact that Cowell still attended Sunday school at the age of 19 is not unusual. As Lacquer (1977, 85) notes, although most attendees were children, the age of scholars ranged from five to thirty years old.

Broken Barriers was probably awarded to Cowell to provide her with a suitable model of behavior. Reynolds (2008, 206) claims that, when awarding books to children, institutions often struck a balance between the eradication of working-class culture and the reinforcement of class divisions and social inequality. Analysis

Presented by the Managers
of the
Ballaugh Primitive Methodist
Sunday School
To
Leticia Cowell
Jan'y 22/08

Fig. 2: Prototypical prize inscription (from *Broken Barriers*, 1889, personal photograph taken by the author, 2016).

of the book's contents confirms this aim, as cleaning and nursing are the two main activities of the book's protagonist, Ruth Maplesden, who claims that all girls' knowledge "must be gained from books" (Marchant, 1889, 144).

The prize inscription is written in black fountain pen and plain indelible pencil. The indelible pencil served as a convenient substitute for the fountain pen, as it could be carried on one's person without need for ink or fear of leaks, and provided firm pressure and permanent markings (Dube, 1998). It is likely that the managers of the Sunday school opted for indelible pencil when writing the recipient's name and date so that the inscription did not fade or smudge and would serve as an enduring emblem of Cowell's good behavior. The fountain pen, on the other hand, was reserved for writing the name of the Sunday school. The differences in writing implement also suggest that the two acts of inscription were carried out at different times. It is possible that the Sunday school name was written in all the copies when

they were originally purchased, whereas the prize winners' names were added after they had been selected to receive a prize.

In this example, the type of handwriting is a form of Vere Foster looped cursive. The Vere Foster handwriting style was introduced into schools in the late-nineteenth century. It was deemed a single, general purpose style that blended the needs of elegance and speed, thus making it suitable for both middle- and working-class boys and girls (Smith, 1977, 27). This style can be seen clearly in the prize inscription through the letters that are slightly slanted to the right and characterized by rounded ascenders and descenders. The clarity of letterforms in this handwriting would have ensured that the message of the inscription was interpreted clearly by the recipient. The inscriber has also chosen to center particular elements of the inscription (e.g., 'of the' in line 2; 'to' in line 5). This not only makes those words more salient, but it also provides symmetry between the first and third lines and the fourth and sixth lines respectively. As Kress & van Leeuwen (2006 [1996], 93) note, symmetrical composition and equidistant spatial arrangement of different elements connote their equivalent importance. The inscriber has also chosen to underline the date, which marks the end of the reading path and signals the completion of the inscription.

The prize inscription is an example of the continuous conflict between the book as an object of social control and the book as a source of intellectual emancipation. Across the working-class groups in the larger dataset, it is the most common form of inscription. While this suggests that Edwardian institutions were keen to educate the working classes, it is clear that the content of books was chosen by teachers acting *in loco parentis* with the aim of conveying moral messages as a form of protection against undesirable attitudes and behaviors in working-class children's lives. Reynolds (2008, 205) claims that prize books were specifically aimed at working-class boys and girls because they had not yet developed their own coherent world view. Thus, it was easier to curb their awareness of inequality and maintain class divisions.

However, the broader findings of this study suggest that Reynold's generalization cannot be applied across all members of the working classes. First, the fact that many prize books in the dataset, including that of Cowell's, survive in excellent condition suggests that many owners adapted the intended purpose of the books and drew meaning more from their aesthetic appeal than their actual content. This was particularly important for unskilled working-class children, who may have had very few personal possessions and tangible indicators of achievement in their lifetime. Second, the presence of defaced and damaged prize inscriptions in the dataset suggests that while working-class children may have been the intended objects of control, they developed their own responses by "accepting, rejecting, absorbing, adapting, distorting or countering" (Entwistle, 1990, 36–37), rather than

blindly accepting middle-class messages. Such defaced inscriptions demonstrate that, even as children, members of the working classes with few prospects of social mobility knew how to carry out acts of symbolic resistance.

5.2 Gift Inscription

A gift inscription is an annotation that proclaims a relationship between two (or more) people and is often exchanged on a particular occasion, such as Christmas or a birthday. By giving a book as a present, the gift inscription mediates between the category of an object and a relationship, and is an example of what I call 'constrained ownership', given that ownership is granted to the recipient by the giver. When writing a gift inscription, although the book is used as a medium to express social relationships, the inscriber has the ability to shape the medium to their own purposes. This makes the book's status as a commodity become ambiguous and endows it with a fetish-like social power that is unrelated to its true worth (Kopytoff, 1986, 83).

Figure 3 shows a gift inscription written by Herbert A. Prince and given to Mrs Ellen Holman. It comes from the 1911 edition of *A Journalist in the Holy Land*, a travel book about Egypt and Palestine by Arthur E. Copping.

At the time of inscription, Prince was a 44-year-old insurance clerk, living in Sutton, Surrey; Holman was a 52-year-old housewife, also living in Sutton. Holman's son, Lennox, was a work colleague of Prince's [1911 census]. Prince was part of the lower-middle class of Edwardian Britain, while Holman's social background suggests an upper-middle-class woman. Their differences in class may explain Prince's use of atypically ornate calligraphy when writing this otherwise prototypical inscription. Having obtained social mobility through their employment as clerks, the newly emergent lower-middle class was emphatically not working class and felt stridently aware of the fact. Conscious of Holman's own status, Prince may have used calligraphy to index his aspiring education and culture. Prince's choice of book is also interesting, as Egypt and Palestine were popular destinations for middle-class Edwardian tourists. In purchasing this book, Prince was perhaps surreptitiously signaling his desire to be accepted into Holman's social circle. This is supported by the fact that Prince's book cost 6 shillings² [Religious Tract Society Catalogue, 1911]—a considerable amount of money for a clerk whose weekly wage was £3³.

² This equates to roughly £33.60 in modern money.

³ This equates to roughly £336 in modern money.

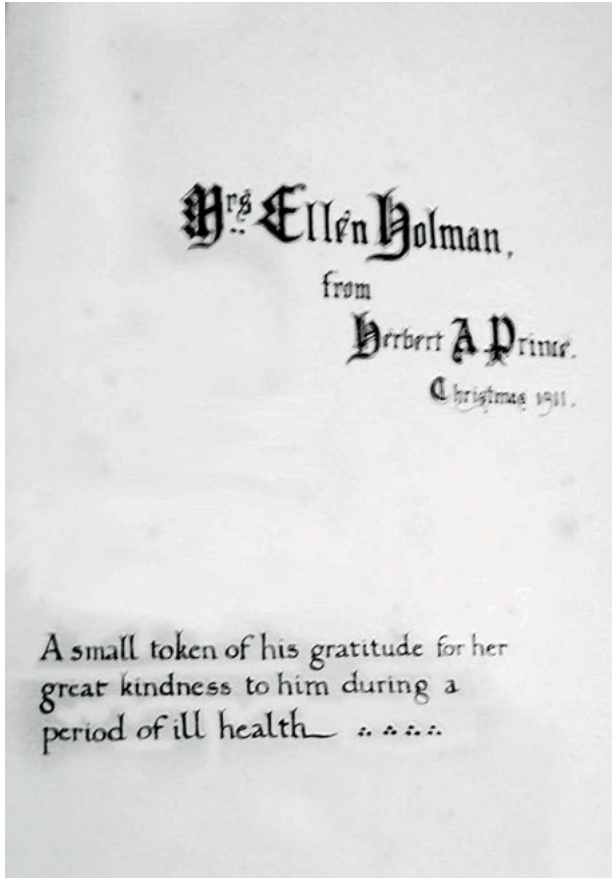


Fig. 3: Prototypical gift inscription (from a journalist in Holy Land, 1911, personal photograph taken by the author, 2016).

The first part of the inscription is located on the center right of the front free endpaper and is written in ornamental gothic script. From an inspection of Prince's handwriting in the 1911 census, it is clear that the handwriting used in the inscription does not resemble his everyday writing practices. Writing in 1906, the craftsman Edward Johnston stated that, "Gothic lettering is one of the most picturesque forms of lettering and therefore of ornament—and besides its ornamental value, there is still in the popular fancy a halo of romance about 'black letter', which may fairly be taken into account" (Johnston, 1906, 331).

Thus, it is likely that, presenting the book as a gift, Prince deliberately chose this lettering for its aesthetic appeal and positive reflection on himself. Traditional

gothic lettering used glossy black for the body text, vermilion for capital letters, and gold for decoration. Here, despite the monochrome black ink, Prince has used shading to give the impression of different hues. The first letters of both names and the word 'Christmas' are emboldened, thus accentuating their appearance on the page. The writing style also adheres to other traditional characteristics that gothic script possessed, such as straight vertical orientation and lines produced by small controlled nib movements.

The second part of the inscription, which is separated from the first in a new frame at the bottom left corner of the front free endpaper, indicates a change in writing style. Here, the letters are based on humanist minuscule—a style of script that was invented in secular circles in Italy at the beginning of the fifteenth century. This writing was typically associated with intelligence, the revival of antiquity and beauty (Meiss, 1960, 109), and it may have been chosen by Prince to promote an image of himself as someone who is well-educated (and hence, justifiably part of Edwardian middle-class society). This part of the inscription is followed by four three-dotted triangles. Within mathematics, this symbol signifies 'therefore', and may have been employed in this context to signal the physical act of giving the book as a token of gratitude. Alternatively, this symbol can be considered to represent an asterism, which was often used to indicate a break in a text. Although asterisms are used nowadays by typographers as end marks, this was not the case in Edwardian Britain. Therefore, this use of the asterism violates our traditional understanding of its use, as no more text follows. This highlights the importance of considering semiotic choices within their original socio-historical period of production. Furthermore, the fact that the two parts of the inscription are located in separate frames is also significant, as this separates the act of thanking from the representation of the participants involved in the speech act. This disassociation is further exemplified by the use of the personal pronouns 'his' and 'her' that cannot be linked back to the referents as easily when they appear in a separate frame.

According to Carrier (1990, 581), the exchange of gifts is not neutral; rather, it is deeply embedded in cultural meaning. This is particularly apparent in Prince's inscription, whose chief aim is to obtain social respectability over any other factor. In Edwardian Britain, giving a gift was strongly bound up with notions of a "gift-debt" (Mauss, 2011, 42) that had to be repaid, thereby forging a mutual interdependence between giver and receiver. Thus, Prince uses gift-giving to establish a personal link between him and Holman, and aspires to a similar social status to his upper-middle-class recipient. It is well-established that the lower-middle-class were far more susceptible to the gift economy of Edwardian Britain (Bailey, 1999) and used gift inscriptions more than any other group as a means of inspiring respect and esteem, while also advancing their interests in upward aspiration. This is a trend that can be seen across the broader dataset.

5.3 Bookplate

A bookplate can be defined as a label usually affixed in the front cover of a book, identifying the person or institution to which it belongs. Bookplates emerged from a stage of practical utility to become an object of artistic value, which embodied the individual characteristics of their owners. Traditionally, bookplates were the stronghold of the upper classes of society who commissioned artists to custom design armorial bookplates with heraldic symbols. During the mid-nineteenth century, as the Victorian concern for ‘keeping up appearances’ grew, middle-class owners began to recognize the potential of bookplates as identity markers. Consequently, stationers and booksellers started to offer bookplate design as an in-house service. By the beginning of the Edwardian era, the application of mass-production newspaper print methods and machinery led to the emergence of cheaper mass-produced bookplates that could be bought in bulk from booksellers. This drastically changed the bookplate market, enabling the lower classes to afford them for the first time. Outraged at the commercialization of this once bespoke practice, upper-class owners began using particular semiotic and material choices to set themselves apart from others.

This can be clearly seen in the bookplate in Figure 4, an example of ‘voluntary ownership’, which shows a pictorial library interior design that belonged to the upper-class Edwardian newspaper editor, Ralph D. Blumenfeld. Blumenfeld was American-born but became a naturalized British citizen in 1907 [1911 census]. His bookplate featured in a 1903 travel guide to Oxford. According to the 1903 advertising catalogue of A & C Black, *Oxford* was the most expensive book they sold at 20 shillings⁴. Its high price was due to the fact that the book was a limited-edition print run with sixty hand-drawn colored plates by the artist John Fulleylove.

Blumenfeld’s bookplate was printed on copperplate paper and custom-designed by Elizabeth W. Diamond, an avid American bookplate artist and collector in the early twentieth century. Her initials, E.W.D, can be seen in the bottom right-hand corner. A bookplate by Diamond would have cost roughly £20 and demonstrates the disposable income that Blumenfeld had at this time.

Blumenfeld’s bookplate captures a typical Edwardian upper-class lady in her drawing room. Although it is uncertain who the lady in the picture is, images show a similarity to Ralph’s wife, Theresa. The picture shows one participant—the lady—who is looking out of the window at a row of thatched cottages. The houses bear a resemblance to Blumenfeld’s residence at Hill Farm, Great Eastern in Essex [1911 census], and act as a material sign of his high social status.

⁴ This roughly equates to £112 in modern money.



Fig. 4: Prototypical Bookplate (from *Oxford*, 1903, personal photograph taken by the author, 2016).

The way the lady is presented also provides a lot of information about upper-class life in Edwardian Britain. The bookplate features a wooden lectern, reading chair, double hung windows, stacks of books, flowers and ink and quill—all characteristic features of an Edwardian drawing room (Musson, 2014). These elements act as circumstances that serve as deliberate displays of wealth and high social status. The use of shading gives the furniture the look of solid wood, and perhaps was chosen to reflect the dependable social status of the lady herself. The representation of a wood-like material also grants the border the properties of a physical frame.

The lady's appearance is characteristic of a pre-Raphaelite woman, known for her long curly hair, thick neck, solid jawline and low-necked dress. Here, she is

engaged in an act of “offer” (Kress & van Leeuwen, 2006 [1996], 119): she is not interested in the viewer and, instead, is totally immersed in looking out of the window. This is also characteristic of pre-Raphaelite images in which women are often shown looking away from the viewer. Sawhney (2006) argues that this pose implies the fetishization of the female as a result of male fantasy. This is supported by Chartier (2002, 173), who argues that female reading was often associated with sensual pleasure and secret intimacy. If we consider that the image is of Blumenfeld’s wife, the concept of the male viewer as a ‘privileged voyeur’ (Sawhney, 2006) gains additional meaning. However, the fact that her head is angled away from the plane of the viewer suggests that there is no reciprocity between the two entities.

The lady’s left-facing position is also noteworthy when compared with the other library interior bookplates in the complete dataset: all of them show the owner facing left. Johnson (2000) claims that direction in portraits was determined by a set of unwritten laws that indicated that a person facing left was looking to past accomplishments, while a person facing right was looking to the future. Given that most owners of library interior bookplates were upper-class Edwardians who feared the collapse of a hierarchical society, it is significant that their bookplates may have been used to foreground the ‘glorious past’ in a bid to hold onto it. The oblique angle of the lady is also meaningful, as according to Tagg (1988, 37), in Edwardian Britain, frontality was considered “a code of social inferiority”. Thus, her side positioning serves as an implicit message that could be verbalized as “I am not part of your world and I do not want to make contact with you. However, feel free to marvel at my wealth and splendor.”

The presence of Ralph D. Blumenfeld’s name etched on the banderole next to the lady may suggest joint ownership of the book between husband and wife: by means of a verbal representation for Ralph and a visual representation for Teresa, which was not uncommon in upper-class bookplates (Stimpson, 2009, 60). This joint ownership is also strengthened by the presence of the Blumenfeld coat of arms resting on the floor in the foreground, the three bees signaling industry, creativity, and eloquence (Velde, 2000). The proximity of the German writing below, ‘bedächtig, beständig, bescheiden’ (thoughtful, steady, humble), encourages the reader to attribute these qualities to the three bees (Zakia, 2007, 28) and, by extension, to the family itself.

Thus, it would seem that Blumenfeld uses the bookplate as a symbolic form of domination to perpetuate his high social status and set himself apart from the lower classes who could only obtain status symbolically. For Blumenfeld, the bookplate is an item of cultural capital that enables cultural consumption “to fulfil a social function of legitimating social difference” (Bourdieu, 2010, xxx). This is a pattern that can be observed in custom-designed bookplates throughout the dataset, the majority of which rely on self-portraits, furniture, or coats of arms to represent the

owner's wealth and high social status. Some are also printed on silk or velvet or use silkscreen, aquatint, and woodcut print methods.

6 Theoretical Conclusions

In this paper, I have argued that multimodality could benefit from the adoption of an ethnohistorical approach. An ethnohistorical perspective to multimodality has the advantage of moving multimodal, especially social semiotic, analysis beyond a text-centered focus by grounding analysis in archival evidence on particular ideologies, cultures, and traditions. The findings gathered from this small-scale case study of Edwardian book inscriptions suggest that carrying out similar analyses on more examples from the larger dataset would offer a valuable extension to current text-centered analyses and provide a greater understanding of artefacts through resources that may not have been considered before. This is particularly important for exploring 'ordinary writing' and capturing the voices of unrepresented people who are often forgotten in history.

The example of Cowell's prize inscription shows how the working-class book often oscillated between an object of social control and a source of intellectual emancipation. While the topic of Cowell's book suggests an imposition of the awarding institution's views on the role of women, it is, nonetheless, significant that Cowell owned a book, given that women's personal ownership had only just been achieved thanks to the 1882 Women's Property Act. Furthermore, it shows an increasing acceptance and will towards women's education and literacy following the 1870 Education Act.

Prince's gift inscription, on the other hand, shows his middle-class aspirations and highlights how marks of ownership could be used as performative constructs of social mobility. This was achieved through his elaborate choice of typography and color. However, it is clear that Prince's self-constitution of status meant that he was deprived of any actual profits associated with recognized status.

Blumenfeld's bookplate indicates the full repertoire of semiotic resources available to the Edwardian upper class. This enabled them to commission artists to design custom-made bookplates whose elements were chosen from a range of materials, colors and fonts. Through his bookplate, Blumenfeld transformed into an inventor of the "stylisation of life" (Bourdieu, 2010, 50), while the classes below became actively involved in this stylization as they searched for individuality and self-expression.

Bringing ethnohistory and multimodality together has provided a way of uncovering socioculturally induced meanings and functions specific to Edwardian

society. For example, choices of writing implements were largely motivated by the fact that black ink was most widely available in shops, as well as long-established social conventions, which dictated that black ink was the most appropriate for writing. Similarly, when considering handwritten inscriptions, it is important to bear in mind that most handwriting can be directly linked to the style that was taught in schools at that time. In cases in which the handwriting style does not match with that which was taught in Victorian or Edwardian schools (i.e., the gift inscription above), the historical connotations of particular styles must be considered.

Understanding the traditions of Edwardian society has also made it clear how, although book owners had a certain freedom in their choices of image, color, typography, and materiality, as inscriptions began to take on more standardized forms, owners found themselves constrained by the need to adhere to traditional standards of composition. This meant that when creating inscriptions, they kept three factors in mind: an awareness of unspoken historical rules, a recognition of the boundaries of social acceptability, and the constraints and possibilities of the meaning resources available. As a result, all inscriptions were written or pasted on the center of the front endpaper—the same place that they had been inscribed for more than three hundred years prior. These findings shed light on some particularly interesting sociocultural variables that affect composition and perhaps have not been considered previously.

Combining multimodal tools with archival records has also helped to emphasize the ways in which meaning potentials can shift over time. For example, the types of furnishings displayed in the upper-class bookplate and the use of copperplate paper were strongly bound up with wealth and social status. Nowadays, in a society in which most items are mass-produced, we may take these choices for granted and fail to acknowledge their symbolic importance for Edwardians. Furthermore, this approach indicates a need to recognize the subtle rules of Edwardian society when attributing meaning to inscriptions in order to recognize examples of deviation and the potential significance of this nonconformity.

The methodology used in this study has also enabled a greater understanding of the communicative roles that institutions, such as schools or churches, may have had in influencing and dictating semiotic choices. For example, the prize inscription example demonstrates that many of the semiotic choices were made with an awareness of the fact that the book would be presented at a public prize ceremony attended by parents and children. Consequently, great attention was paid to the choice of book and wording of the inscription to present a positive image of the awarding institution. Creating a good impression of their supposed generosity could bring other benefits, such as increased membership or monetary donations.

Currently, there is much debate about whether multimodality should be viewed as a framework within semiotics or a discipline in its own right. This study has demonstrated that, while a social semiotic approach to multimodality is useful, on its own, it is too text-centric and does not give enough attention to external motivations that may have influenced a text's design. Indeed, van Leeuwen & Jewitt (2000, 138) and Bezemer & Jewitt (2010, 194) have both argued that multimodality can only ever be one element of an interdisciplinary equation which must also encompass other theories and methodologies.

The ethnohistorical methodology used in the current study provides one such way of enhancing multimodal analysis. It suggests the importance of cross-checking and triangulating multimodal analyses with historical awareness of institutions and social structures. Furthermore, by its focus on 'ordinary writing', it highlights the possibility of revolutionizing how multimodal artefacts are understood, particularly in terms of their importance in the lives of marginalized groups as symbolic forms of power. Without an archival investigation into the book owners, nor a detailed exploration of the social norms and conventions of inscriptions, many of these underlying meanings would not have come to light.

Therefore, this study provides support for the assumption that multimodality would benefit greatly from the introduction of other research methods to achieve more in-depth analyses that are imbedded within the cultural codes of a particular group. In recent years, there has been a growing awareness of the need for interdisciplinary collaboration between multimodal scholars (i.e., O'Halloran & Smith, 2011; Bateman et al., 2017). Nonetheless, it is clear that there still remains a requirement to develop a growing sense of what other disciplines and approaches can offer in order to take this further. Within the context of this study, collaborating with IT technicians, social historians, and library and information scientists to develop a digital archive of the collected book inscriptions may offer one important way to rescue large numbers of endangered examples of ordinary writing and revolutionize how they are understood, particularly in terms of their semiotic and material components. It may also provide a unique learning resource for schools, universities, and the general public, and open up discussions between institutions regarding the high cultural value of historical material artefacts. This potentiality makes it clear that rather than considering whether multimodality should be its own discipline, continued efforts must be made to anchor it in interdisciplinarity and recognize the benefits of achieving an integrated view that goes beyond the viewpoints offered by any one discipline alone.

Bibliography

- Altick, Richard Daniel. 1957. *The English Common Reader*. Chicago: University of Chicago Press.
- Axtell, James. 1979. Ethnohistory: An Historian's Viewpoint. *Ethnohistory* 26(1). 1–13.
- Bailey, Peter. 1999. White Collars, Gray Lives? The Lower Middle Class Revisited. *Journal of British Studies* 38. 273–290.
- Barton, Davis & Walter Hamilton. 1998. *Situated literacies*. London: Routledge.
- Barton, Davis & Uta Papen. 2010. *The Anthropology of Writing*. London: Continuum.
- Bateman, John A. 2008. *Multimodality and Genre: A Foundation for the Systematic Analysis of Multimodal Documents*. Basingstoke: Palgrave Macmillan.
- Bateman, John A., Judy L. Delin & Renate Henschel. 2004. Multimodality and Empiricism: Preparing for a Corpus-based Approach to the Study of Multimodal Meaning-Making. In Eija Ventola, Cassily Charles & Martin Kaltenbacher (eds.), *Perspectives on Multimodality*, 65–87. Amsterdam: John Benjamins.
- Bateman, John A., Janina Wildfeuer & Tuomo Hiippala. 2017. *Multimodality – Foundations, Research and Analysis. A Problem-Oriented Introduction*. Berlin: Mouton de Gruyter.
- Benson, Paul. 2017. *The Discourse of YouTube*. London: Routledge.
- Berry, David M. (ed.). 2012. *Understanding Digital Humanities*. Palgrave Macmillan.
- Bezemer, Jeff & Carey Jewitt. 2010. Multimodal Analysis: Key Issues. In L. Litosseliti (ed.), *Research Methods in Linguistics*, 180–197. London: Continuum.
- Bodmer, G. 1999. Workbooks and Toybooks: The Task and Gift of a Child's Book. *Children's Literature Association Quarterly* 24(3). 136–140.
- Bourdieu, P. 2010. *Distinction: A Social Critique of the Judgement of Taste*. London: Routledge.
- Calder, Sandy. 2016. *The Origins of Primitive Methodism*. Woodbridge: Boydell Press.
- Carrier, J. 1990. Reconciling Commodities and Personal Relations in Industrial Society. *Theory and Society* 19. 579–598.
- Castle, Egerton. 1892. *English Book-plates: An Illustrated Handbook*. London: A & C Black.
- Chartier, R. 2002. The Practical Impact of Writing. In D. Finklestein & A. McCleery (eds.), *The Book History Reader*, 111–159. London: Routledge.
- Crain, Patricia. 2016. Children in the Margins. In Patricia Crain (ed.), *Reading Children: Literacy, Property, and the Dilemmas of Childhood in Nineteenth-Century America*, 109–143. University of Pennsylvania Press.
- Djonov, Emilia & Theo van Leeuwen. 2011. The Semiotics of Texture: From Tactile to Visual. *Visual Communication* 10(4). 541–564.
- Dube, Liz. 1998. The Copying Pencil: Composition, History, and Conservation Implications. *The Book and Paper Group Annual*. Washington DC: *The American Institute for Conservation* 17. <https://cool.conservation-us.org/coolaic/sg/bpg/annual/v17/bp17-05.html>. Last accessed: 17 September 2019.
- Durie, Mark. 1997. Book Review: Reading Images: The Grammar of Visual Design, by Gunther Kress and Theo van Leeuwen. *English in Australia* 119. 92.
- Entwistle, Dorothy. 1990. *Children's Reward Books in Non-Conformist Sunday Schools, 1870-1914: Occurrence, Nature and Purpose*: University of Lancaster dissertation.
- Faudree, Paja & Magnus Pharaoh Hansen. 2013. Language, Society and History: Towards a Unified Approach? In Nicholas J. Enfield, Paul Kockelman & Jack Sidnell (eds.), *Cambridge Handbook of Linguistic Anthropology*, 227–249. Cambridge: Cambridge University Press.
- Gillen, Julia. 2013. Writing Edwardian Postcards. *Journal of Sociolinguistics* 17(4). 488–521.

- Gillen, Julia & Nigel Hall. 2010. Edwardian Postcards: Illuminating Ordinary Writing. In David Barton & Uta Papen (eds.), *The Anthropology of Writing*, 169–189. London: Continuum.
- Goffman, Erving. 1959. *The Presentation of Self in Everyday Life*. New York: Anchor Books.
- Granelli, Emilio & Carlos Martinez-Hinarejos. 2016. A Multimodal Crowdsourcing Framework for Transcribing Historical Handwritten Documents. In *Proceedings of the 2016 ACM Symposium on Document Engineering*, 157–163. Vienna, Austria.
- Grenby, Matthew. 2011. *The Child Reader, 1700-1840*. Cambridge: Cambridge University Press.
- Hamilton, Walter. 1895. *Dated Book-Plates*. London: A & C Black.
- Hammond, Mary. 2006. *Reading, Publishing and the Formation of Literary Taste in England, 1880-1914*. Aldershot: Ashgate Publishing Company.
- Hiippala, Tuomo. 2015. *The Structure of Multimodal Documents: An Empirical Approach*. London: Routledge.
- Hymes, Dell. 1962. The Ethnography of Speaking. In Thomas Gladwin & William C. Sturtevant (eds.), *Anthropology and Human Behavior*, Washington: Anthropology Society of Washington.
- Johnson, Dale. 2000. Left vs. Right, The E-Sylum. Last accessed: 17 September 2019. http://www.coinbooks.org/esylum_v03n15a05.html.
- Johnston, Edward. 1906. *Writing and Illuminating and Lettering*. New York: Dover Publications.
- Katriel, Tamar. 2015. Expanding Ethnography of Communication Research. *Communication Theory* 25. 454–459.
- Kopytoff, Igor. 1986. The Cultural Biography of Things: Commoditization as Process. In Arjun Appadurai (ed.), *The Social Life of Things*, 64–91. Cambridge: Cambridge University Press.
- Kress, Gunther. 1996. Representational Resources and the Production of Subjectivity: Questions for the Theoretical Development of Critical Discourse Analysis in a Multicultural Society. In C. Caldas-Coulthard & M. Coulthard (eds.), *Text and Practices: Readings in Critical Discourse Analysis*, 15–31. London: Routledge.
- Kress, Gunther & Theo van Leeuwen. 2006 [1996]. *Reading Images: The Grammar of Visual Design*. London and New York: Routledge.
- Lacquer, Thomas W. 1977. *Religion and Respectability: Sunday Schools and Working-Class Culture 1780–1850*. New Haven: Yale University Press.
- Laurier, Eric & Angus Whyte. 2001. 'I saw you': Searching for Lost Love via Practices of Reading, Writing and Responding. *Sociological Research Online* 6. 1–10.
- van Leeuwen, Theo. 2006. Towards a Semiotic of Typography. *Information Design Journal* 14(1). 139–155.
- van Leeuwen, Theo. 2011. *The Language of Colour. An Introduction*. London and New York: Routledge.
- van Leeuwen, Theo & Carey Jewitt (eds.). 2000. *Handbook of Visual Analysis*. London: SAGE Publications.
- Lerer, Seth. 2012. Devotion and Defacement: Reading Children's Marginalia. *Representations* 118(2). 126–153.
- Marchant, Bessie. 1889. *Broken Barriers*. London: Joseph Toulson.
- Mauss, Marcel. 2011. *The Gift: Forms and Functions of Exchange in Archaic Societies*. Eastford: Martino Fine Books.
- Meiss, Millard. 1960. Towards a More Comprehensive Renaissance Palaeography. *The Art Bulletin* 42(2). 97–112.

- Musson, Jeremy. 2014. The History of the Country House Drawing Room. Last accessed: 17 September 2019. <http://www.countrylife.co.uk/articles/history-country-house-drawing-room-63348>.
- North-Lee, Brian. 1979. *British Bookplates – A Pictorial History*. London: David and Charles.
- O'Hagan, Lauren. 2015. Multimodality in Edwardian Bookplates: The Rediscovery of a Lost Art. Unpublished MA Thesis. Cardiff University.
- O'Hagan, Lauren. 2018. *Class, Culture and Conflict in the Edwardian Book Inscription: A Multimodal Ethnohistorical Approach*. Cardiff, Wales: Cardiff University dissertation.
- O'Hagan, Lauren. 2019. Towards a Multimodal Ethnohistorical Approach: A Case Study of Bookplates. *Social Semiotics* 29(5). 565–583. <https://doi.org/10.1080/10350330.2018.1497646>.
- O'Halloran, Kay L. 2015. Multimodal Digital Humanities. In Peter P. Trifonas (ed.), *International Handbook of Semiotics*, 389–416. Dordrecht: Springer.
- O'Halloran, Kay L. & Bradley Smith (eds.). 2011. *Multimodality Studies – Exploring Issues and Domains*. London: Routledge.
- O'Halloran, Kay L., Sabine Tan, Duc-Son Pham, John A. Bateman & Andrew Vande Moere. 2019. Interpreting Text and Image Relations in Violent Extremist Discourse: A Mixed Methods Approach for Big Data Analytics. *Terrorism and Political Violence* 31(3). 454–474. <https://doi.org/10.1080/09546553.2016.1233871>.
- O'Toole, Michael. 1994. *The Language of Displayed Art*. London: Leicester University Press (Pinter).
- Pahl, Kate & Jennifer Rowsell. 2006. *Travel Notes from the New Literacy Studies*. Multilingual Matters.
- Pearson, David. 1998. *Provenance Research in Book History*. London: The British Library.
- Plotz, John. 2008. *Portable Property*. Princeton: Princeton University Press.
- Ravelli, Louise J. & Theo van Leeuwen. 2018. Modality in the Digital Age. *Visual Communication* 17(3). 277–297.
- Reynolds, K. 2008. Rewarding Reads? Giving, Receiving and Resisting Evangelical Reward and Prize Books. In Julia Briggs, David Butts & Matthew Grenby (eds.), *Popular Children's Literature in Britain*, 189–208. Aldershot: Ashgate.
- Rosch, Eleanor. 1975. Cognitive Reference Points. *Cognitive Psychology* 7(4). 532–547.
- Rowsell, Jennifer & Lisha Chen. 2011. English Studies Through a New Literacy Studies-Multimodal Lens. In Brian V. Street & Constant Leung (eds.), *The Routledge Companion to English Studies*, 464–475. London: Routledge.
- Samuel, Raphael. 1988. What is Social History...? In Juliet Gardiner (ed.), *What is History Today...?*, 42–57. London: Macmillan.
- Sawhney, Paramvir. 2006. The Pre-Raphaelite Body. Last accessed: 17 September 2019. <http://www.victorianweb.org/painting/prb/sawhney.html>.
- Sheehan, Bernard W. 1969. Indian-White Relations in Early America: A Review Essay. *The William and Mary Quarterly* 26. 267–286.
- Sinor, Jennifer. 2002. *The Extraordinary Work of Ordinary Writing*. Iowa City: University of Iowa Press.
- Smith, Peter. 1977. *Developing Handwriting*. London: Macmillan.
- Stimpson, Felicity. 2009. The Library at Wallington: National Trust Libraries 3. *Book Collector* 23(3). 45–71.
- Street, Brian. 1984. *Literacy in Theory and Practice*. Cambridge: Cambridge University Press.

- Svensson, Patrik. 2010. The Landscape of Digital Humanities. *Digital Humanities Quarterly* 4(1). <http://digitalhumanities.org/dhq/vol/4/1/000080/000080.html>. Last accessed: 17 September 2019.
- Tagg, John. 1988. *The Burden of Representation: Essays on Photographies and Histories*. London: Palgrave Macmillan.
- Velde, Francois. 2000. Heraldic Atlas. Last accessed: 17 September 2019. <http://www.heraldica.org/topics/glossary/atlas.htm>.
- Zakia, R. 2007. *Perception and Imaging*. Waltham: Focal Press.

John Harnett

Cognitive Pathfinders: Highlighting Cross-Modal Interaction and the Orchestration of Memory in Comics

Abstract: What does it mean to navigate a graphic narrative? Is it just an elaborate analogy for reading or does it typify the level of multimodal negotiation elicited by the graphic novel's orchestration of its image-text domain? This chapter will attempt to address such enquiry and subsequently aim it towards the furtherance of multimodality as a new discipline by utilizing three of the medium's more complex graphic novels: David Mazzuchelli's *Asterios Polyp*, Chris Ware's *Building Stories*, and Alan Moore and Eddie Campbell's *From Hell*. The concept of navigation-as-reading will be promoted by addressing the level of plurivectoral scansion and tactile manipulation that each novel mandates. The aim is to demonstrate how the integration of image and text, and the orchestration of visual layout, facilitates a level of cross-modal realization that grants the reader more than one way to navigate the momentary frailty of human relations in *Asterios Polyp*, account for the fragmented persistence of memory in *Building Stories*, or identify the diffracted resonance of a protagonist's stream of consciousness in *From Hell*.

Keywords: multimodality, graphic novel, comics, navigation, narrative, orchestration

1 Introduction: Framing the Multimodal Agency of Sequential Narrative

There is an array of disciplinary access points to the field of comics studies, from approaches which attempt to gauge the pedagogical application of comics to classroom and/or university syllabi, to the historical/archival reappraisal of some of the medium's earliest practitioners, to gender representation, to increasing trends in the digitization of the medium, and to the collation of empirical data regarding the demographics of reading preferences and strategies-of-engagement of both experienced and inexperienced age groups and readerships (Jacobs, 2013; Kunzle, 2007; Stuller, 2012; Groensteen, 2013 [2011]; Cohn, 2013).

Other scholars and theorists dedicated to the exploration of the medium forward semiotic and narratological analyses which are often motivated by the theoretical potential for comics media to manifest a combinatory visuo-verbal, or

<https://doi.org/10.1515/9783110608694-007>

multimodal, lexicon—a pursuit that is of relevance to this book as a whole, and of specific interest to this chapter in particular. Notably, a common factor linking all of these approaches is the necessity for, or establishment of, an acutely agentive and multimodal reader and it is in this spirit that the field of comics studies can be applied to the concerted aim of this book to argue for the status of multimodality as a distinct discipline.

However, while this chapter will attempt to contribute to the focused mission statement of the overall book to validate such disciplinary status, it is important to note that in variance to many of the other investigations embodied by this project, this chapter will not attempt to provide a framework or methodology from which to categorize the comics medium in its entirety. This is due in no small part to the complexity of suitably partitioning and evaluating the widely fluctuating variance between each illustrator's command of artistic style and imagistic expression. Rather, this chapter will attempt to premise a specifically discourse-oriented and interpretive critique of three graphic novels that have been purposely chosen for their complex and uniquely-tailored articulations of their respective image-text narratives and, as such, their deployment of a considerable degree of multimodal techniques to serve the motivations of their respective story worlds and challenge the reader's visual/textual orientation and modal competency.

To that effect, this chapter will be divided into three sections with each one tending to a different graphic novel and, as such, a different array of diegetically motivated multimodal strategies. Section one will examine David Mazzuchelli's exploration of existential anxiety in *Asterios Polyp* and attempt to gauge the semiotic impact of aligning image, text, color, and font arrangement to serve the interest of portraying the dynamic of sacrificing companionship and human intimacy for the grandeur of narcissistic pride. Section two will attempt to outline one strategy of approach to the highly complex temporal orchestration of Alan Moore and Eddie Campbell's *From Hell* and in so doing highlight the close reading and degree of tactile modality required to successfully align the novel's recurring confluence of narratological and diegetic timelines. Finally, section three will attempt to map two of the many narratological trajectories embedded within Chris Ware's *Building Stories* in order to distinguish the involved level of tactile and ergodic interaction, underpinned by the high degree of multimodal fluency, required on the reader's part to navigate one of the medium's most uniquely designed, yet, paradoxically, thematically quotidian narratives.

As each section will attempt to demonstrate, the graphic novels chosen for this analysis elicit an attuned degree of multimodal fluency given the involved level of readerly orientation and engagement required to successfully appreciate each aspect of their complex layouts. Indeed, for all of these texts the reader is tasked with visually navigating the spatial arrangement of narrative fragments that exploit

the sequential/simultaneous aspects of page layout while at the same time s/he is required to conceptualize each visual narrative as a matrix, or braid, of inter-related nodes where an acquired level of visual retention holds the key to apprehending each story's underlying premise. It is the intention of this book as a whole to advance a framework in which all means for making meaning become recognizable, and ultimately classifiable, under the banner of multimodality. To that end, this chapter will endeavor to stress the necessity to multimodally discipline the readers' mind to make transparent the configuration and deployment of the affordances and constraints of a medium whose *modus operandi* depends on the combined articulation of two modes of communication—image and text.

As explicitly stated above, while this chapter will not attempt to locate its premise on the basis of a proposed methodology in its own right, it will nevertheless prove helpful to highlight some of the methodological framework that has already been forwarded in the field of comics studies in order to house this specific analysis within the parameters of multimodal discourse. According to Neil Cohn, a prominent comics theorist who has published extensively to promote discourse on the establishment of a visual lexicon for the medium,

multimodality is considered to be the normal and predisposed state of human communication, and all expressive modalities are assumed to connect to the same underlying conceptual structure of meaning. Each modality thus has unique schematic 'lexical' and grammatical structures which serve as 'handles' for commonly shared conceptual structures. However, each modality affords different functional advantages and disadvantages for the way in which they package conceptual information (Cohn, 2016, 317).

The underlying conceptual structure of meaning that Cohn signals here can be understood as a corollary approach to the argument forwarded by Scott McCloud in *Understanding Comics*, who suggested that because comics panels fracture time and space a heightened level of deductive reasoning is sometimes required to bridge the conceptual gap between them in order to construct what he terms a "continuous, unified reality" (McCloud, 1994, 67). However, McCloud did not just generate fertile soil for the furtherance of comics studies he did so in comic book form, suitably demonstrating that even discourse on the medium was quite capable of self-referential and multimodal articulation.

Nick Sousanis furthered this self-referential approach with the publication of *Unflattening*, a term which he also employed as a verb to describe "a simultaneous engagement of multiple vantage points from which to engender new ways of seeing" (Sousanis, 2015, 32). By creatively visualizing the way we hold different ways of knowing in relationship he documented how diverse narrative perspectives can be unified to create what he terms "literary parallax" (Sousanis, 2015, 45). Indeed, this concept admits consideration of the multimodal property known as "cross-modal

realization” which describes situations where “a character, an object or a particular setting has the potential to be represented or reappear simultaneously or successively in different modes which allows one to treat any comic as a multimodal text and examine how certain discourse elements are cross-modally realized” (Tseng et al., 2018, 7). Ultimately, whether one refers to the dynamic enacted in these texts as literary parallax or cross-modal realization, one is still addressing that sense of awareness of the unification of affordances offered by two distinct modes of meaning making and this chapter will attempt to suggest a number of approaches intended to highlight such awareness/fluency.

2 Cross-Modal Interaction in David Mazzuchelli’s *Asterios Polyp*

David Mazzuchelli’s *Asterios Polyp* is a densely coded narrative that consistently challenges the reader’s modal orientation and comprehension of the use of visual metaphor. It depicts a character who abandons the security and accolades of a successful academic life that, for him, has become a stifling facade and an uncomfortable reminder of his own narcissism. Along the path of his personal odyssey, he reflects on the consequences of his emotional detachment and the lack of empathy he showed his wife by reliving the most intimate missed opportunities of his broken marriage. In much the same light as Tseng et al.’s assessment of Mazzuchelli’s graphic novel adaptation of Paul Auster’s *City of Glass*, Mazzuchelli’s *Asterios Polyp* is

inherently a narratively complex work, employing varieties of metaphorical visual images to delineate issues of identity, time, social and family relationships depicted in verbal texts, etc. It could, thus, be expected to be cognitively challenging for the readers to track the recurring characters, objects, and places, as well as their inter-relations (Tseng et al., 2018, 5).

At its diegetic heart the plot corresponds to, while being a sardonic deflation of, the Campbellian monomyth as documented in *The Hero with a Thousand Faces*. In other words, it subverts such archetypal heroic tropes in the development of the protagonist’s journey as: the initiation of a state of crisis brought on by an unchecked state of hubris; the undertaking of a perilous journey in an effort to restore moral/psychological balance; the subsequent endurance of great suffering; the acquisition/bestowal of a powerful talisman or boon; and the return of the enlightened hero to his estranged love or community. However, on a deeper level it is realized as a quest for—and a quest orchestrated by—aesthetic symmetry and

psychological harmony. And in that regard it exploits a range of visual and textual affordances which consistently task, and reward, the reader's multimodal fluency.

Firstly, the novel's narrator is Ignazio Polyp, Asterios's twin brother who died in childbirth. This is important because it lets Mazzuchelli source the narrative from an omniscient perspective, which allows him to project innovative stylistic emphasis onto particular character traits and emotive states. It also establishes the novel's embedded emphasis on symmetry by establishing Asterios's lifelong quest for a sense of wholeness based on unsatisfactory substitutions for a twin he never knew. He tries to articulate the sense of unease this instilled in him throughout his life when he offers that it is, "like searching for your reflection in the mirror" (Mazzuchelli, 2009, 116). Commenting on the impact of this symmetrical upheaval, Ignazio explains that "[a]bstractions have always appealed to my brother—especially systems and sequences that are governed by their own internal logic. In addition, he's always been fond of analogues and metaphors" (Mazzuchelli, 2009, 105).

This conceptual process is initially represented on a page at the beginning of the story by depicting the womb as an approximation of a yin-yang symbol, which visually describes a sense of harmony and counterpoise. However, the visual and textual interaction on the lower half of the same page form a combined, symmetrical aesthetic which emphasizes Asterios's isolation and the psychological weight of the metaphorical burden he carries on his shoulders (Mazzuchelli, 2009, 21). This is just one example, amongst a highly imaginative plethora in the novel as a whole, of image-text correspondence or enhancement and it represents a solid point of consideration for the adoption and promotion of multimodal awareness. As such, it typifies an instance of cross-modal realization, or what Jeff Bezemer and Gunther Kress call 'transduction', where a shift in semiotic material from one mode to another occurs and it is done in this case to emphasize an emotive state of mind (Bezemer & Kress, 2016, 53). However, the symmetrical harmony of the aforementioned yin-yang womb depiction is deconstructed by way of Ignazio's death and the crux of Asterios's existence thereafter is a subconscious desire to find a meaningful substitute.

Tracing his efforts to do so requires attuning oneself to the modal range of visual representation that Mazzuchelli employs with variances in both style and color used to signal a range of emotional states in the two main characters and chart important thematic developments in the story world. Asterios is depicted in blue and often in exaggerated geometric shapes whereas his wife, Hana, is depicted in pink or red hues and outlined with curved or wavy lines as can be seen in their first meeting. These modal affordances are either emphasized or restrained to suit the intensity of emotion that Mazzuchelli wishes to depict.

Representing intensity is a familiar entrance point into multimodal discourse (Bezemer & Kress, 2016, 7) and in this case the power dynamic that defines Asterios and Hana's relationship and the intensity of emotion that sometimes erupts in the heat of argument between them are represented by a disproportionate ratio of color distribution that emanates from the characters themselves and often extends outwards to color or shape the environments they occupy. For example, at one point Asterios reflects on an argument between himself and Hana that is the culmination point of a series of incidents where he had either taken control of conversations by speaking over her or casually dismissed her opinion on a variety of issues and can be seen in Figure 1.

As the argument/sequence continues, it gradually employs a saturation of hermeneutic imagery and colors to convey Hana's frustration and resignation as she is ultimately housed and isolated within the detached psychological and metaphoric framework that Asterios imposes upon her (Mazzuchelli, 2009, 217). For the reader, processing the multimodal layout of this argument evokes a sense of imbalance and a loss of symmetry that ideally captures the narcissistic and defeating sense of self-gratification that Asterios initially takes from such confrontational victories.

There are many more examples of this dynamic in the narrative and importantly they are summarized in an asymmetrical page layout at one point as Asterios fully reckons with the impact that a life of self-promotion had on his marriage (Mazzuchelli, 2009, 265). Thus, when symmetry is offset, it signals imbalance. Such imbalance is multimodally channelled to signify breakdowns in communication and the resulting unhappiness it causes. On the other hand, symmetrical representations signify epiphany, reconciliation, and harmony. For Asterios, restoring symmetrical/psychological harmony to his life firstly requires displacing the fulcrum upon which his universe revolves, in other words: his ego. This reorientation takes place mid-conversation in a bar in the aptly named town of Apogee when a supporting character uses three beer mats to explain the key to a successful relationship, see Figure 2.

There is a point in the narrative where Ignazio explains that "in the certitude of symmetry, the consonance of counterpoise, Asterios found a measure of solace" (Mazzuchelli, 2009, 108). This simplistic beer mat construct embodies such solace and is the moment of Asterios's epiphany in the story. However, it is also a useful avatar for the nature of multimodal discourse itself. As Carey Jewitt explains:

When several modes are involved in a communicative event all of the modes combine to represent a message's meaning [...]. Any one mode in that ensemble is carrying a part of the message only: each mode is therefore partial in relation to the whole of the meaning [...]. An image may be used to link two layers within a text, and the character of what the image links

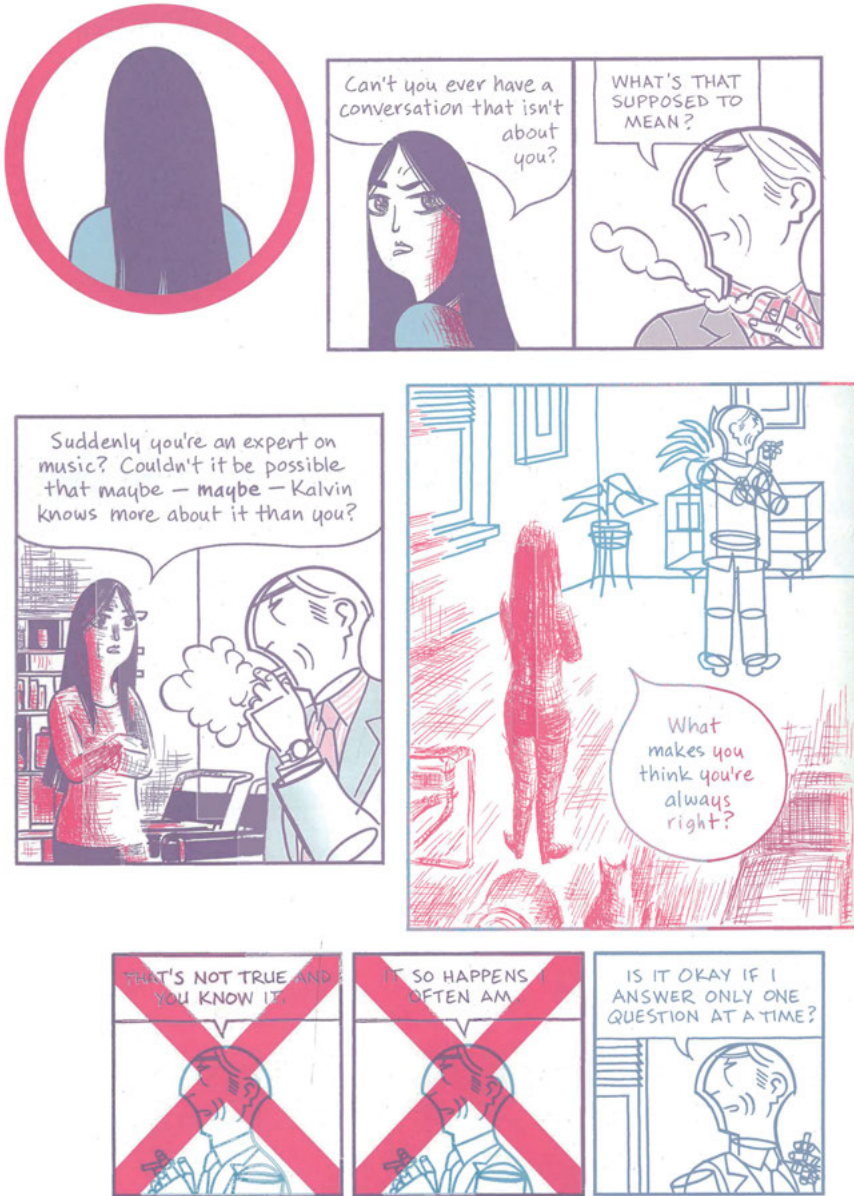


Fig. 1: Cross-modal communication/interaction in Mazzuchelli (2009, 214). Reprinted by permission of Pantheon Books.



Fig. 2: Multimodal counterpoise in Mazzuchelli (2009, 258). Reprinted by permission of Pantheon Books.

to may elaborate the meaning of the image itself [...], analysis of the moment-by-moment processes of constructing multimodal ensembles can enable the analyst to unpack how meanings are brought together (Jewitt, 2016, 73).

Apart from the insight that Asterios takes from this composed symmetry, Mazzuchelli also uses the moment to introduce a new color in the narrative to signal the development of Asterios' emotional maturity and to prepare the reader for the novel's closing act. As Randy Duncan points out, "Green, formed by the combination of the cool cyan of Asterios the architect and the warm yellow of Apogee, appears for the first time (and) Mazzuchelli uses it to signal that duality has been displaced by synthesis" (Duncan, 2012, 52). In so far as readerly orientation and navigation is concerned, this also marks the end of Ignazio's narrative voice as Asterios finally relinquishes the burden of impossible comparison to an illusory twin. Subconscious duality is thus replaced by a singular objective as Asterios acknowledges his shortcomings as a husband and returns to his estranged wife with the boon of acquired wisdom in an effort to reconcile their differences. This reunification also tasks the reader's multimodal fluency with the stylistic affordances that the novel has deployed thus far. For example, there is no further variance in the visual depiction of either character and a uniform style is conferred on both. Similarly, their speech balloons no longer overlap to signify dominance and submission. Rather, the tails of the balloons interweave to depict a shared conciliatory perspective between husband and wife (Mazzuchelli, 2009, 310–311).

For the multimodal reader, engaging with such cross-modal interaction evokes considerations of Lakoff and Johnson's argument that the development of metaphorical imagination is seen as a crucial skill in creating rapport. Notably, they argued that their conceptual structure, "involves all the natural dimensions

of our experience, including aspects of sense experiences: color, shape, texture, sound, etc. (to the effect that) artworks can provide new ways of structuring experience” (Lakoff & Johnson, 1980, 235). Thus, one of the metaphors they construct is CLOSENES IS STRENGTH OF EFFECT, which they base on their premise that “the spatial metaphors inherent in our conceptual system [...] automatically structure relationships between form and content” (Lakoff & Johnson, 1980, 136). If one applies this relational concept to the manner in which Asterios and Hana’s efforts to reconcile their grievances are depicted then one can apprehend the stylistic decision to depict their word balloons as wrapping themselves round each other until eventually they share the same space on the page.

Notably, there is a moment at an earlier point in the story when Asterios and Hana are introduced to a musical composer, who discusses his creative process with them and states that “simultaneity—the awareness of so much happening at once—is now the most salient aspect of contemporary life” (Mazzuchelli, 2009, 209). By presenting a narrative that challenges the visual reader to assign distinct thematic values to techniques including visual metaphor, cross-modal interaction, and the variation between symmetrical and asymmetrical page layouts, Mazzuchelli effectively promotes the impact that such salient awareness can have on storytelling as a whole. As such, *Asterios Polyp* harmonizes the modal affordances of two channels of meaning making and articulates a sense of equivalence between them which empowers the story’s simple message—that we cannot live in isolation, nor are we constrained to a singular mode of empathetic expression.

3 Temporal Orchestration as a Modal Affordance in Alan Moore and Eddie Campbell’s *From Hell*

Alan Moore and Eddie Campbell’s *From Hell* represents a work of such complexity that it could open up a number of pathways into multimodal discourse, but it has been chosen here to emphasize the level of temporal/tactile navigation that the reader must be prepared to negotiate in a comic or graphic novel. The novel is set in Victorian-era East End London and details the decaying mindset of its protagonist, William Gull, as he succumbs to the influence of Masonic symbolism and carries out a series of gruesome murders. Navigating the location and recurring frequency of seemingly discordant panels that depict his stream-of-consciousness throughout the novel’s vast temporal structure requires a detailed level of readerly engagement.

The novel is a specifically applicable example of the layering of varying time frames that the medium has at its disposal due to the frequency with which it

regularly draws attention to the fluctuating relationship between diegetic and narratological time. At some points it accomplishes this by exploiting the temporal relations that are orchestrated between panels that are located on the same page and at other stages by drawing attention to a thematically-networked orientation of panels throughout the work as a whole, which is used to highlight the story's suggested connection between ancient mythology, Victorian-era socio-cultural traditions and prejudices, and an apocalyptic fear of the future. To anticipate one strategy for applying a coherent reading to the level of temporal confluence that the novel exhibits, it will prove useful to cite a deceptively simple three-panel exhibit of temporal confluence by the Irish comics artist Barry Hughes (see Figure 3, which in spite of its minimalist design suitably highlights how sequential narrative can destabilize the narratological distinctions between past, present and future).

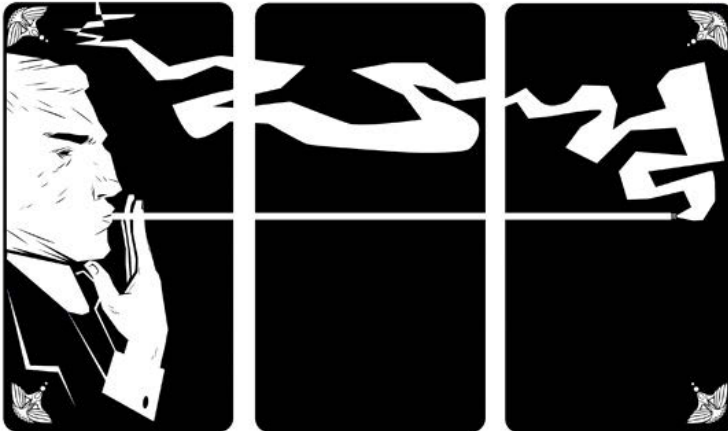


Fig. 3: Temporal confluence (Barry Hughes, 2001). Reprinted by permission of the artist.

This sequence effectively realizes the visual conflation of multiple time frames in the same panel and is introduced here to act as a modal primer for the degree of temporal orchestration that the reader must negotiate/apprehend in order to expose the underlying meaning behind *From Hell*'s complex design. In particular, it will be demonstrated in due course, by way of reference to Figure 4, that the deceptively simple confluence of time frames in the first panel of Hughes's sequence can be used to decode one of the most thematically/chronologically discordant panels in the novel as a whole. In addition, the use of a top-down visual perspective in key panels will be analyzed as a recurring modal affordance, which signifies key points of confluence in *From Hell*'s overall temporal matrix.

However, the first phase in initiating such a multimodal venture is to propose a theoretical basis for the techniques used in the novel by Moore and Campbell that allow the reader to resynchronize his or her comprehension of its temporal topography. Accordingly, by aligning a temporal/narratological hypothesis devised by the French structuralist Gerard Genette alongside comics theorist Thierry Groensteen's analysis of the graphic novel as a network of privileged relations a constructive sense of multimodal orientation is prepared.

According to Christian Metz, "one of the functions of narrative is to invent one time scheme in terms of another" (Metz, 1974, 18). As students of narratology are aware, this function permits the author of a text—or indeed the author/illustrator collaboration in a graphic novel—to actively distinguish story time from narrative time. Acknowledging the premise that a narrative can be orchestrated as a confluence of temporal orders that can be realized at arbitrary points of intersection is especially relevant when applying it to an artifact as multimodally valent as a comic/graphic novel. Indeed, the realization that such temporal confluence can be used to enhance certain elements of the story world evokes considerations of the work as a branched network of connections. Put simply, if a narrative purposely deploys the intersection of two (or more) time frames with considerable frequency then it begins to form a temporal matrix and within that matrix each point of temporal intersection—or panel—resonates with the potential to address any other, regardless of its physical location within the story.

According to Genette, comprehension of such a dynamic establishes in traditional literature what he refers to as an "omnitemporal reader" (Genette, 1980, 78). In a graphic novel, such a reader is capable of establishing a multitude of telescopic relationships between randomly dispersed panels and thus negating the linear constraints of conventional readerly engagement. This is forwarded as a semantic and cognitive process that allows the reader to gradually perceive an overall ubiquity to a story-world that is at once spatial and temporal. From a medium-specific vantage point, Genette's proposal can be made contemporarily relevant when conceived in terms of Groensteen's conceptual braid analogy, which describes a comic in terms of a network of privileged relations based on the repetition of thematically inter-related panels or sequences. According to Groensteen,

braiding over-determines the panel by equipping coordinates that we can qualify as hyper-topical, indicating their belonging to one or several notable series, and the place that it occupies. Therefore, the panel is enriched with resonances that have an effect of transcending the functionality of the site that it occupies [...] a space that we can cross, visit, invest in [...] an activated and over-determined site where a series crosses (or is superimposed on) a sequence (Groensteen, 2007 [1999], 147).

Given that space and time are regularly equated to be interchangeable properties in the field of comics discourse (McCloud, 1994; Cohn, 2013; Ball & Kuhlman, 2010), then Groensteen's identification of hyper-topical coordinates are equally substantiated by referring to them as hyper-temporal. To extend Groensteen's analogy, such junction points not only represent spaces that can be crossed, visited, or invested in; they also represent time frames that can be similarly traversed.

From Hell exemplifies such temporal orchestration and a viable entrance point into its matrix-like design is presented early in the story when William Gull is initiated into the secret order of the Freemasons (Moore & Campbell, 2013, chap. 2, 9). The manner in which this occasion is structured is important because it signifies one of the earliest points in the novel's design where a seemingly incongruous panel disrupts the coherence of a sequence. In this case the panel in question depicts the rotting corpse of a seagull and the reader must try to decipher its meaning and contextualize its placement. This sets in motion a steady accumulation of similarly realized anachronistic junctures where both Gull and the reader are thrust into a state of confusion and for the sake of expediency one such example shall be dealt with here.

When Gull carries out his final and most intensely ritualistic mutilation in chapter ten he is in the full grip of insanity and as the relentless influx of his visions overpowers him he is briefly exposed to an important event in his imminent future—a clandestine trial before his Masonic peers (Moore & Campbell, 2013, chap. 10, 24). The sequential and temporal order in which he foresees this event is depicted in one order and yet the order in which the trial actually transpires two chapters later is laid out with deliberate variance (Moore & Campbell, 2013, chap. 12, 22). Having at this stage in the narrative already been exposed to jarring premonitions of a future that terrifies him the temporal manipulation that takes place between these sequences skillfully aligns the semantic obstruction experienced by the reader with the psychological turmoil experienced by the protagonist.

Genette points out that in an omnitemporal narrative it is entirely at an author's discretion to group together, in defiance of all chronology, events connected by spatial proximity (Genette, 1980, 85). Additionally, much can be gleaned in this situation from Groensteen's point that when sequential format is disrupted by rearranging panels, such as occurs in this instance, then a sequence's structuring function begins to share a unique bond with its expressive function (Groensteen, 2007 [1999], 48).

In *From Hell*, this unique bond between structure and expression is based on temporal relativity and it quite effectively disorients the reader, challenging his or her perception of narrative time in order for it to be subsequently enhanced. Gull's mind becomes a locus point for more than one time frame at once and he loses his mind. As Eddie Campbell, the novel's illustrator, explains, “[a]t any

moment his subconscious is capable of occupying, a place where past, present and future co-exist as his impressions are assembled into a frighteningly suggestive pattern” (Moore & Campbell, 2013, 265). This pattern becomes fully realized in one panel in particular and is presented here in Figure 4.

In chapter 14, Gull dies but his disembodied consciousness takes flight from his body and begins to navigate time on two levels. On one level his consciousness begins to drift through time as it relates to the scope of all that has come to pass in the novel thus far and his transcendent journey is portrayed as a retrospective amalgamation of key events in the narrative. However, on a second level, it also moves beyond events as laid out in the story-world revealing to the reader glimpses of a future that Gull could not possibly have experienced in his own lifetime.

In defiance of temporal logic, at one point Gull’s seemingly omniscient spirit begins to descend from the sky and move in closer to better familiarize itself with two men having a discussion on the steps of Christ Church in Spitalfields, one of whom is himself from an earlier point in the story. Representing a deeply contextual amalgam of the kind of temporal layering presented in Barry Hughes’ three-panel sequence in Figure 3, this occurrence in the novel demonstrates a masterstroke in the conflation of temporal orders as both linear story time and non-linear narrative time are framed within the same location on the page. This subsequently generates the realization that both Gull and the reader are revisiting a conversation from their respective pasts, a discovery which loads the event with the retroactive conclusion that within the temporal duality that Gull’s subconscious eventually experiences the reader is not—and was not—the only one observing this conversation on either occasion.

Thus, Gull’s timeline is viewed from the simultaneous perspectives of past and future, and by way of repetition the reader gradually unveils a suggestive pattern in the narrative that traverses moments of premonition and recollection, often resulting in an amalgamation of both within the same panel. The panel itself functions no differently than the first panel in Hughes smoking man sequence, one simply conceives of William Gull’s spirit in place of the ubiquitous smoke from the cigarette, with the notable difference being that in the case of *From Hell* this climactic temporal intersection represents the culmination point of an appreciably vast and complexly woven narrative. Notably, this concept mirrors one of Nick Sousanis’s conclusions on the medium as a whole. In *Unflattening*, he proposes that when a reader regards a multimodal construct as a de-centered, laterally branching structure where each node is connected to any other, then its multiplicity of approaches for constituting experience realigns his or her vision from a one-dimensional to a multidimensional vantage point (Sousanis, 2015, 39).

Therefore, in *From Hell* you have a temporal folding effect that perfectly evokes the need for fluency with the kind of omnitemporal narrative design outlined by



Fig. 4: Temporal and diegetic collocation (Moore and Campbell 2013, chap. 14, p. 12). Reprinted by permission of Knockabout Comics.

Genette and advances the theory that the novel as a whole can be compared to a network of telescopic relations. And it is arguably no coincidence that the symbolic avatar for Gull's liberated stream of consciousness is a bird as one can argue that Campbell's recurring use of the same visual perspective as that contained in Figure 4 signifies instances of temporal confluence.

As the reader encounters, and subsequently re-encounters, any instance of this bird's-eye, top-down viewpoint (of which there are over 40 panels, which are disseminated throughout the narrative) s/he is encouraged to entertain the diegetic probability that William Gull's ubiquitous presence is doubly invested in the narrative's anachronistic orchestration. To that effect it can be offered that

these panels represent the temporally autonomous nodes in *From Hell*'s overall matrix.

Accordingly, this strategy has been forwarded here to suggest that in order for the reader to familiarize him or herself with the temporal variance afforded to the more complex offerings within the medium of sequential narrative s/he must be willing to unsubscribe from conventional reading paths and ensure that repetition links throughout the work are identified in order to apprehend any comic or graphic novel as a potential network of privileged relations. Only in deference to the necessary tactile modality involved in this process, by literally flicking back and forth through the novel to trace and confirm such temporal dispersion, can s/he tend to those properties which are grouped together in matrices formed on the basis of firstly reading, and then re-reading/interrogating, narrative strategies that unlock more than one way in to the story world.

This subsequently raises interesting medium-specific discourse centered in formalistic approaches to the overall medium. However, it goes further than that and indeed places center-stage such multimodal considerations of reading a comic as tactile interaction and temporal reappraisal as the agentive reader/investigator gradually unifies the deliberately fragmented/dispersed components of stories that may often offer multiple entrance and exit points. As such, *From Hell* strengthens the urgency for a multimodal disciplinary correlation within the humanities alongside the advancement of cognitive narratology and the reappraisal of creative world-building as a whole.

4 Tactile Assemblage and Mnemonic Orchestration in Chris Ware's *Building Stories*

While *From Hell* requires a consistent degree of tactile vigilance through the necessity of flicking back and forth through the text in order to make sense of initially anachronistic and discordant panel placement it nevertheless exists as a single artifact. In contrast to this, and in an effort to magnify the scope of multimodal fluency the reader needs to have at his/her disposal when engaging with comics, this analysis will conclude its attempt to platform the benefit of multimodally disciplining the reader's mind by considering the implications and requirements of a decidedly unique, deliberately fragmented, and narratologically frustrating story world.

To that effect the chapter will conclude with a perspective on Chris Ware's *Building Stories* that emphasizes the innovative degree of visual and tactile coordination that the medium is capable of articulating. This unconventional novel-in-a-box

gives a highly fragmented account of three women living in an apartment block in Chicago but it primarily focuses on one, a young artist who is trying to cope with anxiety and a lack of faith in her creative talent as well as physically navigate daily life with the support of a prosthetic leg. Ware uses 14 (and subsequently 15 when you come to realize that the perimeter of the box itself displays panels that also depict the story's main protagonist) separate components to tell the story and from the moment one lifts the lid, this disorientating approach to the act of reading initiates a high degree of visual navigation and tactile interaction. This interaction varies depending on the particular component in the collection that the reader engages with. In some fragments the story is conveyed in a conventional comic book format while elsewhere it is articulated in extendable horizontal strips, expansive spread-sheets, or foldout board-game layouts.

There is no identifiable hierarchical order to the way these fragments should be read, the box does not contain a table of contents nor are the fragments numbered/labelled to cue readerly engagement. As such, the most apparent obstacle facing the reader is choosing an entrance point into this rhizomatic and multimodal work and subsequently organizing its fragments into a coherent order. According to the semiotician Daniel Chandler, “[t]urning experience into narratives seems to be a fundamental feature of the human drive to make meaning. We are “storytellers” with “a readiness or predisposition to organize experience into a narrative form” (Chandler, 2007, 115). Adding a comics-oriented thread to this argument, theorist Jan Baetens has observed that

the impact of our sequential reading habits is so strongly narrative that those habits help us make narrative sense of panels and drawings that seem to defy any direct figuration [...] not only are we capable of reading non-figurative material in a narrative manner, we are also very keen to do so, since narrative is such an efficient and satisfying strategy for handling problems and difficulties in any material we may be reading (Baetens, 2011, 100).

Indeed, when engaging with a narrative that so explicitly draws attention to its disordered structure one might also benefit from the structural insights of Roland Barthes. In his essay, *Structural Analysis of Narratives*, he appropriated the medical term ‘dystaxia’ (which originally refers to a breakdown in muscular coordination) to explain what happens when a narrative’s linearity is offset and the relationship between consecution and consequence is deliberately confused (Barthes, 1977, 118). With reference to *Building Stories*, no matter which one of the story fragments that the reader engages with s/he is confronted with a variety of diagrammatic and/or multi-panel layouts, one such example of which can be seen in Figure 5.

Thus is generated a visually simultaneous impact to the reader’s eye due to the fact that the absence of an implicit linear reading order in such layouts delays the diegetic momentum of the story itself but conversely may reward random gaze

patterns with the revelation of liminal visual cues, thus describing the presentation of a multimodal narrative compelled by what Thierry Groensteen refers to as, “simultaneity and panopticism” (Groensteen, 2007 [1999], 7).

From a multimodal standpoint this has the effect of visually enhancing Barthes concept of dystaxia as it is arguably easier for the reader’s eye to abandon conventional linearity when presented with a simultaneous layout of visuals than it is for the human eye to wander arbitrarily from word-to-word in a strictly textual medium. According to Groensteen, “every comics reader knows from experience that, in practice, even when the gaze functions like an “irremovable beam”, the eye’s movements on the surface of the page are relatively erratic and do not respect any precise protocol” (Groensteen, 2007 [1999], 47/113). However, the recent application of eye-tracking technology to the medium would seem to contest this presumed lack of directional protocol and has started to yield results that maintain that while the visual aspects of sequential narrative “demand far less cognitive effort in people’s reading process, verbal cues are possibly more prominent in the reader’s navigation of the specific identification process” (Tseng et al., 2018, 17).

Fittingly, the first strategy that may be adopted to attenuate *Building Stories*’ panoptic and narratological influx and bring a sense of diegetic coordination to the reading experience is activated by tending to the inclusion of Ware’s directional arrows, and concurrently, to the placement and appearance of directional lexias on any given strip, page, or poster in the combined work. In other words, Ware’s frequent creative placement and redirection of words, phrases, and even entire sentences, that often need to be read vertically or on other occasions to the effect that they actually wind around certain images, or move from the outside to the inside of a building, as can be seen in Figure 5 for example.

The resulting effect of following the path laid out by such meandering text can often signal the importance of a particularly proximate visual component to such sentence arrangement within the overall layout. Indeed, Ware manages to conflate the conventional modal distinctions between text and image in the comics medium by way of the fact that his verbal enunciations demonstrate a multitude of visual dimensions also. Joseph Witek explains that when navigating complex page layouts directional arrows can be used to temporarily suspend the normal reading process in order to foreground the spatial relations of unconventional page structures (Witek, 2009, 152).

However, he offers that such layouts often platform the fundamental navigational impedance brought about by attempting to assess which panel is meant to be read in which order and he qualifies the resulting obstruction by observing that “readers who are trying to figure out the proper way to read the page are readers who are not immersed in the story” (Witek, 2009, 152). Thus, in order to re-establish such immersion, Ware occasionally includes directional arrows which

act as vectors to redirect readers away from the standard ‘Z’ type reading pattern. In addition, as can be seen in Figure 5, Ware locates circuitous and directional sentence fragments (directional lexias) around the structural focal point of the story world (i.e., the building itself) and fully exploits the architectural use of space as it pertains to layout and page design. Abandoning conventional reading vectors in this manner sustains an undercurrent of readerly anxiety by showcasing the inhibitive sense of disorientation that *Building Stories* commands. It consistently invokes a sense of visual dystaxia by mandating erratic reading strategies.

Notably, these reading strategies also have the effect of mirroring the meandering stream-of-consciousness of the story’s narrator. According to Gary Spencer Millidge, “normally, the use of arrows is considered bad practice—a substitute for poor storytelling where the reading order isn’t obvious. But when conceived as part of the intended design as Ware does they can be extremely effective” (Millidge, 2009, 82). Indeed, this directional design strategy is only furthered by Ware’s motivated location of text fragments within the ‘architectural multi-linearity’ of his layouts, and as Gene Kannenberg suggests:

Text becomes diagrammatically directive in both *narrative* and *meta-narrative* fashions. By *narrative* I here refer to the way in which the placement of lexias act to guide the reader’s gaze across the page in a specific direction in order to read various elements in a particular order. By *meta-narrative*, however, I refer to the ways in which the appearance or placement of lexias on the page serve to reflect thematically on characters or events in the narrative (Kannenberg, 2009, 311).

For example, in Figure 5, although the text begins in the conventional top left corner of the layout, it arbitrarily changes direction on the page in such a way as to locate some lexias outside the building and to house others inside it and even embed others within the foundation upon which it stands. In this way it becomes a part of the infrastructure of the building, literally moving around it much like its fictional occupants.

In an interview Ware addressed this technique and explained that “the idea behind that is supposed to be that the main character is doing these stories for a creative writing class and she’s used the building as a character itself and its sort of this self-conscious way for her to get inside of it” (Reid, 2012, 4). Thus, arrows and circuitous sentence fragments are employed as directional vectors, which impose coherence onto panoptic layouts and ultimately serve as mechanisms to guide the disorientated reader’s eye.

The second strategy to help orient the reader-as-assembler/visual navigator relies on the use of visual repetition to emphasize the point that, even in a fragmented story world that initially seems to exhibit no explicit master narrative, the recurring placement of specific visual motifs and sequential patterns can be used

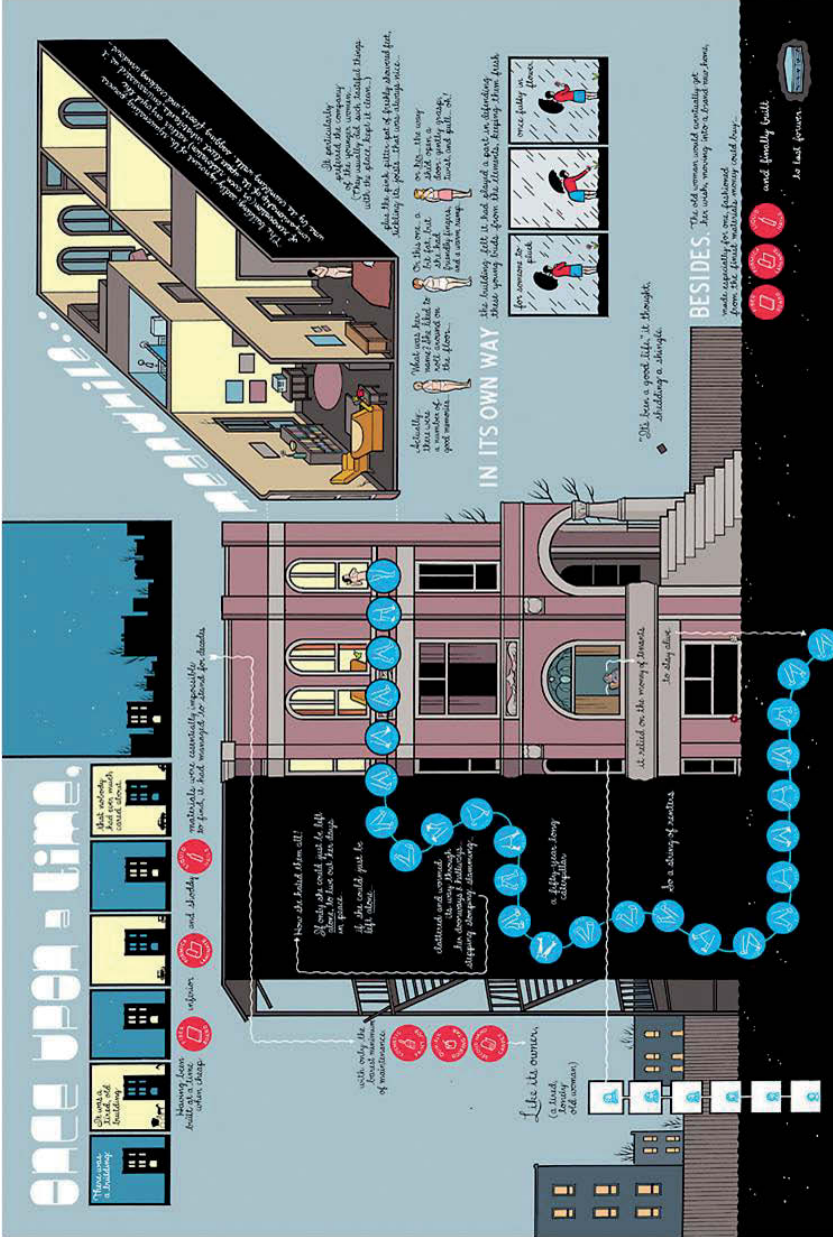


Fig. 5: Directional arrows and circuitous sentences (Ware 2012). Reprinted by permission of Jonathan Cape.

to formulate a thematic basis. In other words, the reader's ability to conceptually assemble a coherent narrative is facilitated by both repetition and change from panel to panel.

As encountered with *Asterios Polyp* and *From Hell*, this is a key factor in what are known as braided narratives, where certain images and panels share a narratologically distanced but diegetically pertinent relationship with one another. There are many instances of visual repetition scattered throughout *Building Stories*' varied fragments but given the gradual realization as to who the story's main protagonist is one of the most significant is the repeated depiction of a quotidian journey to the grocery store that the three occupants of the building embark upon throughout their lives, a voyage that Ware deliberately frames and paces as a despondent pilgrimage of self-reflection.

For the unnamed blonde girl occupying the second storey of the building, who endures a loveless and verbally abusive relationship, a mundane journey to the store, in which she determines to put an end to "the stupid pattern" of her life, is located in a twenty-five panel, center-spread sequence in one of the more conventionally sized comics in the collection (Ware, 2012). Elsewhere, an identically rendered, twenty five panel sequence/journey occurs in the centerspread of the only other similarly sized comic and it depicts the building's landlady walking to the store whilst expressing her exasperation with her overbearing and controlling mother.

Notably, in this iteration the sequence depicts two different periods in her life, contrasting the timid nature of her younger self against the dispirited burden of old age. As such, the reader bears witness to a lifelong pattern of repetitive frustration, which is based on the simultaneous representation of multiple time frames, which display and confirm her confining lack of self-confidence (Ware, 2012).

However, as can be seen in Figure 6, for the main protagonist the same journey is depicted in a double-sided horizontal strip that literally extends the narrative on a tactile level as the reader fully unfolds it and encounters her confession that "the horrible emptiness inside me just won't go away [...] everything makes me feel awful [...] alone" (Ware, 2012). This gradually recognizable pattern qualifies as just one of the novel's strategic uses of repetition to establish thematic and narrative cohesion. In spite of the variation in production design between this narrative fragment and the previous two examples it contains precisely the same sequential panel ordering as those comics.

However, Ware would seem to be staging the main protagonist's torment as a point of emphasis in this case by emulating the orientable properties of what is referred to in mathematics as a moebius strip. This has the effect of enhancing the reader's multimodal engagement as visual and tactile interaction with a looped narrative combine to highlight the enduring anxiety that haunts the novel's



Fig. 6: The Landlady and the main protagonist: Confined order versus durative repetition (Ware 2012). Reprinted by permission of Johnathan Cape.

protagonist. This level of engagement becomes an essential dimension of the narrative project by inciting what Groensteen refers to as translinear and plurivectoral reading strategies (Groensteen, 2007 [1999], 155). As mentioned earlier, there is no mandated reading order to *Building Stories*.

However, Ware is clear that the primary organizing principle of the text is architecture; only in this case the architectonic design that he cites as inspiration can relate both to the visually realized characteristics of the building within the story world itself (which often narrates events that transpire in and around it with an omniscient voice; see Figure 5 above) or it can be used to describe the building

process that the reader must conduct as s/he physically opens, manipulates and aligns/realigns individual story fragments in the pursuit of structural coherence and diegetic resolution (Wolk, 2012, 2).

As such, *Building Stories* stands as a cogent and engaging avatar of the medium as a whole and confirms such thinking within the larger community of multimodality scholarship that, “every act of creating meaning from a multimodal text, happening as it does at the intersection of structure and agency, thus contributes to the ongoing process of becoming a multimodally literate person” (Jacobs, 2013, 17).

5 Conclusion: Multimodality and the Discipline of Reading Comics

In *Graphic Encounters: Comics and the Sponsorship of Multimodal Literacy*, Dale Jacobs proposes that, by examining comics as multimodal texts and reading comics as an exercise of multi-literacy, “we can shed light not only on the literate practices that surround comics in particular, but on the literate practices that surround all multimodal texts and the ways in which engagement with such texts can and should affect our thinking about them” (Jacobs, 2013, 9).

Through the selection of three particularly diverse texts, this chapter has attempted to highlight a selection of such practices by demonstrating the structural variance and multimodal literacy that the medium of sequential narrative is capable of articulating—and by extension the degree of consistent and adaptive agency it requires from its readers. However, one may well ask why such multimodal fluency is important in the first place, or to hone even further enquiry on the purpose of this book in general—why studying comics emphasizes the need for multimodality to be recognized as a discipline in its own right.

The answer to this question takes root in the basis of basic problem-solving and stands as an important reminder of Daniel Chandler’s above-noted observation that turning experience into narratives seems to be a fundamental feature of the human drive to make meaning. Conversely, so too does turning narrative into experience corroborate and strengthen this fundamental human feature. To that end, consistent engagement with such a multimodally valent medium equips the modern reader with an adaptable skill set that unlocks the entire cognitive and narratological potential of texts that at any point are capable of revealing an inherently orchestrated, though not always apparent, relationship between image and text.

In other words, such engagement disciplines the reader's mind to the medium's full potential. In turn, an established, practiced, and consistently referenced discipline of reading is formed, one that belongs within the larger multimodal community, offering its own unique perspectives to it and learning from the accommodating expanse of the multimodal community as a whole new strategies of approach and innovative ways of rearticulating the human condition, and the ways we continue to engage with it.

Bibliography

- Baetens, Jan. 2011. Abstraction in Comics. *Substance* 40(1). 94–113.
- Ball, David M. & Martha B. Kuhlman (eds.). 2010. *The Comics of Chris Ware: Drawing is a Way of Thinking*. Jackson: University Press of Mississippi.
- Barthes, Roland. 1977. *Image – Music – Text*. London: Fontana Press. Edited and translated by Stephen Heath.
- Bezemer, Jeff & Gunther Kress. 2016. *Multimodality, Learning and Communication: A Social Semiotic Frame*. New York and London: Routledge.
- Chandler, Daniel. 2007. *Semiotics: The Basics. Second Edition*. London and New York: Routledge.
- Cohn, Neil. 2013. *The Visual Language of Comics: Introduction to the Structure and Cognition of Sequential Images*. London: Bloomsbury.
- Cohn, Neil. 2016. Linguistic Relativity and Conceptual Permeability in Visual Narratives: New Distinctions in the Relationship between Language(s) and Thought. In Neil Cohn (ed.), *The Visual Narrative Reader*, 316–340. London and New York: Bloomsbury Academic.
- Duncan, Randy. 2012. Image Functions. Shape and Color as Hermeneutic Images in Asterios Polyp. In Matthew J. Smith & Randy Duncan (eds.), *Critical Approaches to Comics. Theories and Methods*, 43–54. London and New York: Routledge.
- Genette, Gérard. 1980. *Narrative Discourse*. Ithaca, NY: Cornell University Press. Translated by Jane E. Lewin.
- Groensteen, Thierry. 2007 [1999]. *The System of Comics*. Jackson, Miss.: University Press of Mississippi. Translated by Bart Beaty and Nick Nguyen, from the original French *Système de la bande dessinée* (1999).
- Groensteen, Thierry. 2013 [2011]. *Comics and Narration*. Jackson, Miss.: University Press of Mississippi. Translated by Ann Miller, from the original French *Bande dessinée et narration: Système de la bande dessinée 2* (2011).
- Jacobs, Dale. 2013. *Graphic Encounters: Comics and the Sponsorship of Multimodal Literacy*. London: Bloomsbury Academic.
- Jewitt, Carey. 2016. Multimodal Analysis. In Alexandra Georgakopoulou & Tereza Spilioti (eds.), *The Routledge Handbook of Language and Digital Communication*, 69–84. London and New York: Routledge.
- Kannenberg, Gene. 2009. The Comics of Chris Ware. In Jeet Heer & Kent Worcester (eds.), *A Comics Studies Reader*, 306–324. University Press of Mississippi.
- Kunzle, David. 2007. *Father of the Comic Strip: Rodolphe Topffer*. Jackson: University Press of Mississippi.

- Lakoff, George & Mark Johnson. 1980. *Metaphors We Live By*. Chicago: Chicago University Press.
- Mazzuchelli, David. 2009. *Asterios Polyp*. New York: Pantheon Books.
- McCloud, Scott. 1994. *Understanding Comics: The Invisible Art*. New York: HarperPerennial.
- Metz, Christian. 1974. *Film Language: A Semiotics of the Cinema* chap. Problems of Denotation in the Fiction Film, 108–146. Oxford and Chicago: Oxford University Press and Chicago University Press. Translated by Michael Taylor. Originally published in French in 1968.
- Millidge, Gary Spencer. 2009. *Comic Book Design*. Sussex: Ilex Press.
- Moore, Alan & Alan Campbell. 2013. *From Hell*. London: Knockabout Comics.
- Reid, Calvin. 2012. A Life in A Box: Invention, Clarity and Meaning in Chris Ware's Building Stories. Last accessed: 17 September 2019. <https://www.publishersweekly.com/pw/by-topic/industry-news/comics/article/54154-a-life-in-a-box-invention-clarity-and-meaning-in-chris-ware-s-building-stories.html>.
- Sousanis, Nick. 2015. *Unflattening*. Massachusetts.: Harvard University Press.
- Stuller, Jennifer K. 2012. Feminism: Second-Wave Feminism in the Pages of Lois Lane. In Matthew J. Smith & Randy Duncan (eds.), *Critical Approaches to Comics: Theories and Methods*, 235–251. New York and London: Routledge.
- Tseng, Chiao-I, Jochen Laubrock & Jana Pflaeging. 2018. Character Developments in Comics and Graphic Novels: A Systematic Analytical Scheme. In Jochen Laubrock Alexander Dunst & Janina Wildfeuer (eds.), *Empirical Comics Research: Digital, Multimodal, and Cognitive Methods*, 154–175. London: Routledge.
- Ware, Chris. 2012. *Building Stories*. London: Jonathan Cape.
- Witek, Joseph. 2009. The Arrow and the Grid. In Jeet Heer & Kent Worcester (eds.), *A Comics Studies Reader*, 149–156. Jackson: Mississippi University Press.
- Wolk, Douglas. 2012. Inside the Box: Building Stories, by Chris Ware. Last accessed: 17 September 2019. <https://www.nytimes.com/2012/10/21/books/review/building-stories-by-chris-ware.html>.

Christopher Taylor

Audio Description: A Multimodal Practice in Expansion

Abstract: The European ADLAB project, coordinated by the University of Trieste, sought to promote the practice of audio description (AD), a method designed to provide access to blind and sight-impaired persons, particularly in the field of cinema and television, by creating a set of pan-European ‘strategic’ guidelines for the profession. A new European project (ADLAB PRO) is now in progress, again coordinated in Trieste, aimed firstly at creating the profile of the professional audio describer, but also including an extension of audio description research into museums, art galleries, churches, important landmarks, and so on. This new development weds AD to the early multimodal work of O’Toole, Kress and van Leeuwen, Baldry and Thibault, etc. As regards the descriptions of museum exhibits, the most effective approach to satisfying the needs of the blind and sight-impaired public needs to be found. Empirical analyses of the linguistic components of descriptions for sighted and non-sighted visitors, as they are presented in audioguides for example, need to be compared, in order to explore the variations in textuality required for audio descriptions. Cognitive linguistics, systemic-functional linguistics, discourse analysis, and other sources will be activated in the search for the most user-friendly yet informative format. But over and above these considerations, there is an ever clearer need to go beyond the image/word symbiosis and bring in other senses in the multimodal approach to AD. One of the project’s aims is to study this approach in depth and extend it to various areas of audio description research for museums. This intersensorial approach will then be expanded further to embrace the use of music, sounds, smell, and taste wherever they can enhance the multimodal experience.

Keywords: multimodality, audio description, museum, senses, intersensorial

1 Introduction

The European ADLAB project (www.adlabproject.eu), coordinated by the University of Trieste, sought to promote the practice of audio description (AD), a method designed to provide access to blind and sight-impaired persons, succinctly described by Snyder (2008) as ‘the visual made verbal’, particularly in the field of cinema and television, by creating a set of pan-European *strategic* guidelines for the profession. The term *strategic* was used to distinguish the approach adopted

<https://doi.org/10.1515/9783110608694-008>

from more descriptive and prescriptive guidelines previously published (Independent Television Commission, 2000; Vercauteren, 2007; Mazur & Chmiel, 2012) in that the guidance provided is more in the form of suggestions for dealing with particular problematic situations. It should also be pointed out that the AD service could prove useful for other users such as language learners, people with learning difficulties, newly arrived immigrants, even normally sighted people looking for a deeper descriptive experience.

A new European project (ADLAB PRO), now in its second year and again coordinated in Trieste, is aiming firstly to create the profile of the professional audio describer and to design a curriculum for post-graduate courses or in-service training. It also includes an extension of audio description research into, e.g., museums, art galleries, churches, and important landmarks. AD in museums is not new, but as many museums are now modernizing, diversifying, and catering for new types of visitors, new multimodal approaches to AD are being sought. In particular it has been recognized that the oral description of objects of a patently visual nature (one goes to a museum essentially to see things) can be augmented effectively through the integration of other senses, particularly touch but also hearing and even smell and taste (Secchi, 2004, 2014; De Coster & Mühleis, 2007; Eardley et al., 2016).

This new development weds AD to the early multimodal work of O'Toole (1994); Kress & van Leeuwen (1996) and Thibault (2000). As regards the descriptions of museum exhibits, be they three-dimensional sculptures or two-dimensional paintings, the most effective approach to satisfying the needs of the blind and sight-impaired public is constantly being sought. Empirical analyses of the linguistic components of descriptions for sighted and non-sighted visitors, as they are presented in audioguides for example (available in situ or downloaded) have been compared, in order to explore the variations in textuality required for audio descriptions as they attempt to compensate for the semiotic modalities that the blind have no access to.

In particular, there is an ever clearer need to go beyond the image/word symbiosis and bring in other senses in the multimodal approach to AD. The introduction of 'touch tours' in many museums has offered blind patrons the opportunity to appreciate works of art, particularly sculptures, through the tactile sense, either feeling real exhibits or replica models. It has been shown how the sense of touch can bring an artifact alive, though necessarily accompanied by an oral input. "In the absence of sight, speech is essential to the integration of sensory input and therefore to perception" (Fryer, 2016, 5). Thus, the key to optimum comprehension would seem to lie in an efficacious blending of the senses. For example, the Anteros Museum in Bologna is quite unique in that it caters specifically for blind and sight-impaired patrons. Visitors are invited to gain a first impression of a sculpture

or painting in relief through touch. They are then ‘talked through’ the exhibit by an individual ‘voice talent’ using a tried and tested formula.

One of the ADLAB PRO project’s aims is to study this approach in depth and extend it to all those areas referred to above. This multimedial, and what might be termed intersensorial, approach will then be expanded further to embrace the use of music, sounds, smell, and taste wherever they can enhance the multimodal experience. The aim of this research is then to explore how diverse multimodal properties, through new approaches to AD, can benefit the growing community of persons with sight loss. How can verbal description for the visually impaired, still the cornerstone of any attempt to recreate images in the mind, be usefully supported and enriched by other senses substituting vision. A number of case studies will be outlined in order to demonstrate what is possible and what works.

2 Audio Description and Multimodality

“Audio description is the technique used for making theatre, movies and TV programs accessible to blind and visually impaired people” (Benecke, 2004, 78). This statement from some years ago referred to a new genre ‘translating’ the visual into words, at the time limited to stage and screen. While these sectors remain the principal focus of audio description studies, the field has more recently been extended to other areas such as traditional and modern museums, art galleries, and the other sectors mentioned above. Film is a perfectly contextualized audiovisual text, in the sense that the audience, usually seated in a cinema or at home, follows a pre-fixed sequence with a clear beginning and end. It is a time-based materiality. In the case of museums and similar environments, the spatial, temporal, and textual confines are more porous and changing (and the user is usually standing or moving). Whereas film AD is intended to follow a script inserted between the dialogues, museum AD needs to be more flexible. Although museum AD is also a kind of meta-text, substituting for a visible object, when a human describer is involved, he or she may digress from any ‘script’ as they see fit, as in any case of ekphrasis. There is no dialogue, though other sounds are possible, nay desirable, e.g., music, sound effects. Udo & Fels (2009, 1) wrote of a Fashion Show AD accompanied only by music “to allow emotion and excitement, as well as description of the important visual elements”.

But in all its forms AD is inextricably linked to the concept of multimodality. The blind and sight-impaired community, while unable to access the crucial semiotic modality of image, can still participate in a multimodal social process involving all the other senses. What is required is a clear knowledge of how to integrate these

senses in the most meaningful way, creating a multimodal literacy empowering a large, and ever growing, section of the population in terms of both production and reception. Helen Keller’s comment that “the only thing worse than being blind is having sight but no vision”¹ leads to the axiom that the blind have no sight but they may have vision, while the seeing have sight but sometimes no vision, in the Collins English Dictionary sense of “the ability [...] of great perception”. Thus, one of the main requirements of the audio describer is the ability to visualize, and to activate vision.

By way of example, we shall stay for the moment with screen AD and the short film *Across Still Water* (Grimberg, 2015), which features a boy, John, who is going blind but refuses help, for example the use of a white cane. He wishes to avoid discrimination, but in a pub he meets a blind friend, Ishmael, with his guide dog. As Ishmael attempts to persuade John to seek assistance of some kind, a white cane (folded) is seen on a table. In the next shot the cane has disappeared. The cane is not part of the narrative—John does not want it, and Ishmael does not need it. But it is a visual symbol of the difference between independent Ishmael and reluctant John. However, the AD does not mention it, in fact giving credence to John’s fears. The AD therefore is not just film narrative but must carry other kinds of interpretation intended by the film director. The second interesting point about the production of this film is that it was shown experimentally, with AD, to a blind and a sighted audience together. In this way it was not the original film that was regarded as the end-product plus a supplement, but the version with AD was to be considered the end-product, thereby removing any sense of discrimination. For this reason, it is also recommended that blind persons be involved in the production of audio description, and that AD be considered an essential part of the filming process. There is a long way to go towards putting this recommendation into standard practice, but the mere formulating of the idea is an indication of the current growing awareness of the need for accessibility.

3 Approaches to Audio Description

Cognitive linguistics, systemic-functional linguistics, discourse analysis, multi-modal semiotics, and other disciplines have all been activated in the search for the most user-friendly yet informative format for AD (Holsanova, 2016; Taylor, 2012, 2017; Braun, 2011; Vercauteren & Remael, 2014; Vandaele, 2012; Holland, 2009).

¹ Helen Keller Quotes. (n.d.). BrainyQuote.com. Retrieved June 16, 2018 from BrainyQuote.com: https://www.brainyquote.com/quotes/helen_keller_383771.

The organization of information in narrative structure, information flow as represented in theme progression, and the complex intertextual connections between image, location, and printed word when faced, for example, with a painting hung in a specific place in a gallery with explanatory captions, have all been brought into play. Another very useful concept in the search for optimum solutions is to be seen in the augmenting of the sense of presence (Fryer & Freeman, 2012). The idea of presence is explained in relation to the fact that there is more to enjoying a film than simply understanding the story. Media form is known to impact on this idea of presence in sighted people, for example the way a film is shot influences the audience's emotional engagement. In this way film studies can also be shown to impact on AD. The sense of presence can be replicated for the blind through describing camera shots and identifying strange sounds. The purpose is to create the perceptual illusion of non-mediation—of 'being there'. A striking example is that of the sense of fear produced in the film *The Shining* (Kubrick, 1980). The description of Jack Nicholson's face in the famous scene where he appears for the first time as a madman needs to reflect the 'madness in the eyes' and the 'curled lip rage' and thus requires considerable attention to detail. The seeing audience could be heard to gasp at this scene and for people with sight loss to feel the same sense of presence; the description should aim to cause just such a gasp. This sense of 'presence' can be generated also in static art AD. The questions posed by Dobbin et al. (2016) show this clearly:

- What would it be like to step into a painting?
- Do you know what a painting feels like?
- Does it make a sound?
- How does it smell? (Dobbin et al., 2016, 11).

Such considerations bring us to the controversial question of appraisal, or the level of subjectivity that can be allowed in AD, the acceptable level of personal interpretation or judgment. There is a broad divide across the Atlantic on this issue. Generally speaking, the American school, particularly the disciples of Joel Snyder of the American Audio Description Project, insists on complete objectivity. The ITC², in its AD guidelines also says "the best audiodescribers objectively recount the visual aspects of an image. Subjective or qualitative judgments or comments get in the way" (Independent Television Commission, 2000).

² The Independent Television Commission (ITC) licensed and regulated commercial television services in the United Kingdom (except S4C in Wales) between 1 January 1991 and 28 December 2003.

This would include ‘madness in the eyes’ and the ‘curled lip rage’. However, some practitioners and researchers, particularly in Europe, are a little more flexible. Take the example in Figure 1 based on Vercauteren & Orero (2013, 192).



Fig. 1: Sadness (Personal photograph).

The most objective approach (op. cit. Vercauteren & Orero, 2013) might be to give the description ‘the eyelids droop as the inner corners of the brows rise, the corners of the lips pull down, and the lower lip pushes up in a pout’.

Or we could say ‘the girl looks sad’ or ‘has a sad look on her face’. Or, in the case of *The Shining*, ‘his frightening face’. The choice lies with the describer and his or her adherence to the more objective or more subjective approach.

Now let us move on to museum AD where the use of appraisal in the description of art works and other artifacts, though also controversial, can quite often be seen to lean towards the more subjective end of the spectrum. Secchi (2014, 197) refers to “words as the aesthetic equivalents of artistic images”. Spoken AD can then be blended with the integration of other senses, particularly the tactile sense, in a process of synesthesia. However, it must also be remembered that although subjectivity may be more common, even in the museum context people with sight loss are averse to receiving too much help and like to make their own judgments based on the description they receive. For this reason, the weaving of words, touch, smell, and taste needs to create a multimodal impact that engages receivers through the sensory channels available to them, providing the input required to understand and to interpret the object in question, with the addition of a judicious amount of assistance if deemed necessary.

4 Touch Tours

The introduction of ‘touch tours’ in many museums has offered blind patrons the opportunity to appreciate works of art, particularly sculptures, through the tactile sense, either feeling real exhibits or replica models. Gunther Kress (2003, 140) pointed out that communication is moving from telling to showing, but perhaps telling and showing works for the blind, where the ‘showing’ is to be understood through the sense of touch. “Words can give a great deal of information (dimensions, structure, ‘meanings’ of ambivalent signs) but if visitors are to have a vivid mental image of the work of art, it is necessary to explore the field of intersensorial possibilities” (De Coster & Mühleis, 2007, 198). As mentioned above, these concepts mark a great step forward for blind patrons although the touching still needs to be accompanied by a verbal description and instruction. The recreation of a picture in the mind of a blind person, through the use of other senses, is based on the notion that the senses that substitute vision can act directly on the intellect. As pointed out by Secchi (2014, 198) “the analytical power of touch also enables the experience of seeing more deeply”. Eardley et al. (2016, 263) refer to “physical and intellectual access”.

Thus, the key to optimum comprehension would seem to lie in an efficacious blending of the senses. At the afore-mentioned Anteros Museum in Bologna, visitors are invited to gain a first impression of a sculpture or painting in relief through touch (a three-dimensional translation of works by Botticelli, Raphael, and others). They are then ‘talked through’ the exhibit by an experienced ‘voice talent’, a professional audio description speaker. The visitor gains a hands-on experience of such elements as texture, shape, gesture, vector ... and ultimately meaning. Figure 2 shows Guido Reni’s painting *Atalanta and Hippomenes* (1620–1625), reproduced in relief form for precisely this purpose.

One of the ADLAB PRO project’s aims has been to study this approach in depth and extend it to all those areas mentioned above. This intersensorial approach has been expanded further to embrace the use of music, sounds, movement, smell and taste wherever they can enhance the multimodal experience, for example the use of surround sound in headphones or re-enactments of the subject matter.

The National Museum of Ethnology in Kyoto, Japan, invites patrons, including sighted patrons, to “have a conversation with an artefact” (Hirose, 2013) and investigate the difference between the senses of touch, sight, and hearing. One even sighted visitor claimed “I was able to savour elaborate handworks and various textures that I couldn’t (sic) appreciate just by looking at them” (Hirose, 2013).

In order to probe the question of *how* to describe for a blind and sight-impaired public, a questionnaire was issued to actual museum audio describers asking



Fig. 2: Atalanta and Hippomenes (1620-1625), Guido Reni, Museo Nazionale Capodimonte, Naples. Public Domain.

them what elements they rated most highly. Although describers work with various forms of museum artifacts (paintings, sculptures, installations), and the questionnaire was rather generic in nature, the results perhaps being more indicative than conclusive, those who responded opted for the following:

- use of non-visual imagery (shape, texture, ...)
- use of color
- factual and contextual information (what the exhibit tells you)
- any additional input (see Hutchinson, 2016).

The role of AD was seen as a verbal substitute for the visual, a way to explore meaning and aid understanding, but also a means of creating an emotional experience.

5 Museums in the 21st Century

The modern museum institution is, in many places, becoming a hybrid enterprise (see Samis & Michaelson, 2017, on the visitor-centered approach). It remains a store of artifacts and collections and a site for tourism. It also remains a place of research, especially in the more important places but museums are becoming ever more places of entertainment, often twinned with an educational component in what is now known as edutainment. Promotion now plays an important role as museums compete with other leisure activities to attract visitors. Practically all medium-sized and large museums now contain a shop for the merchandising of museum-related wares and other cultural (and non-cultural) items. Museums everywhere are moving “away from a collections-centered identity and towards more visitor-centered experiences” (Wood & Latham, 2014, 13). As Samis and Michaelson conclude: As we succeed in welcoming new audiences, internal organizational change is not the only form of change we see. Rather, the culture of our museums themselves will inevitably change. (Samis & Michaelson, 2017, 175)

Last but not least, museums are a place of service providing access for the disabled in the form of lifts and ramps for the physically handicapped, but also sign-language assistance and subtitles for the deaf, and braille and audio description for the blind. But how well does this latter service work? What are the guidelines for an effective AD?

A study (Piazza, 2017) was undertaken at the University of Trieste to test the efficacy of existing guidelines, by examining the official audio descriptions of 20 exhibits available for blind patrons of the Victoria and Albert Museum (henceforth V&A) in London. Table 1 is based on the CDWA guidelines³ produced by *The Jean Paul Getty Trust* (2014), indicating the various aspects that could inform the description of works of art.

6 The Victoria and Albert Museum

At the V&A, twenty out of the thirty-one categories listed in the table (including title, measurements, related works, related visual documentation) are applied in the AD descriptions (see Piazza, 2017, 83). Interestingly, the other eleven (including the

³ CDWA is a set of guidelines for the description of art, architecture, and other cultural works. CDWA also provides a framework to which existing art information systems may be mapped, upon which new systems may be developed, or upon which data may be linked in an open environment.

Tab. 1: Categories for the Description of Works of Art (CDWA), (2014) The Jean Paul Getty Trust

Object/Work	Context
Classification	Descriptive note
Title	Critical responses
Creation	Related works
Style/period	Current location (country, museum)
Measurements	Copyright
Materials	Ownership
Inscriptions	Exhibition (loan history)
State (damaged?)	Cataloguing history
Edition	Related visual documentation
Facture (production process)	Related textual reference
Orientation	Corporate authority
Physical description	Place authority
Condition/examination history	Generic concept authority
Conservation	Subject authority
Subject matter	

various ‘authorities’) do not appear in any of the exhibits. But the question remains as to whether even this amount of detail is necessary. In an attempt to whittle down the elements to describe, the most common indications suggested by experts such as Snyder (2008); Fryer (2016), and particularly Neves (2015) in her contribution to “Pictures Painted in Words”, the publication containing the afore-mentioned guidelines resulting from the ADLAB project, are (1) an initial panoramic from the general to the specific; (2) a more detailed logical and sequential description; and (3) the use of objective and jargon-free language. There is a pictorial density that needs to be rationalized. Many works, both natural and abstract, contain considerable detail, not all of which is necessary for the viewer to absorb in order to appreciate the artifact in question. Consider the detail in Velazquez’s famous painting *Las Meninas* (1646) in Figure 3 and how selecting the salient items becomes of paramount importance.

The painting is well known and many viewers of this work will already have a mental image of its contents, at least in broad detail. A blind audience, on the other hand, even if they know of the painting, will need a little more direction in appreciating the essential aspects. Apart from the historical and political background to the work, which can be briefly outlined, the painting itself requires some dissection. It portrays members of the Spanish royal court surrounding the diminutive figure in the foreground of the Infanta Margherita. The other figures, including Velazquez himself, in painter mode, are half-hidden in the background. Velazquez’s presence is presumed to be an affirmation of his importance at court,



Fig. 3: Las Meninas (1646), Diego Velazquez, Museo del Prado, Madrid. Public Domain.

while the man on the stairs is the courtier José Nieto, the reason for whose presence in the picture is not really known. His positioning, however, provides the sense of perspective and depth. The prevailing colors are dark grey and brown providing a chiaroscuro effect except where the light highlights the clothing, especially the Infanta's white dress and the peoples' faces. The child with his foot on the family dog, the dwarf figure and the maids of honor ('las meninas'), together with the Infanta, make up the horizontal front line completed on the left by the painter's

giant easel. The mirror in the center of the painting behind the Infanta reflects the figures of the King and Queen, who are witnessing the creation of the work. These are the essential elements that need to be described. The geometrical positioning of each figure and the sense of perspective, along with the role of the mirror, create in the viewer the sense of actually being in the room, that sense of ‘presence’ and ‘stepping into a painting’ mentioned earlier.

Details such as the contents of the paintings on the wall, the type and features of the dog (though the size may be of interest), the clothes of all the characters and information as to the names of the minor figures, the role of the dwarf and speculation as to what Velasquez is painting should be considered only if the blind patron, in the same way as a sighted visitor, is interested as an art enthusiast. Similarly the various layers of meaning attributed to the work, such as the extolling of the importance of art in society or the heralding of the future monarch, may be explored if the visitor shows the necessary interest. The aspects mentioned above are what the majority of sighted visitors take in, some of them absorbing much less. The painting is hung in the Prado museum in Madrid and, notwithstanding the importance and fame of the work, many visitors will be pleased to have seen *Las Meninas*, but if it comes towards the end of their visit, they may give it a rather superficial viewing, merely appreciating its overall beauty. Certainly, the blind visitor may only select a limited number of exhibits, and have a greater attention span per painting, but for the average member of the blind community an AD of a work such as *Las Meninas* can remain within the limits outlined above. If no tactile or musically accompanied version is available, and this is the case at the Prado, the spoken AD must suffice by replacing the one sensory channel with another. In order to do justice to the complete visual experience, both physical and intellectual, the description outlined above will also include mention of the general ambience, the other visitors, and any relevant surrounding action.

In fact, many of the descriptions of items at the V&A are coherent with the kind of categories described previously. A logical order is pursued, developing from the generic to the specific. Persons are described according to their compositional prominence (see Figure 3 above) and landscapes from left to right, and an objective though jargon-free language is employed.

Take the audio description of Constable’s painting *Boatbuilding at Flatford Mill* (1814) (Figure 4):

This oil painting, almost square in shape, was painted in the late summer and early autumn of 1814, and according to Constable’s biographer, C. R. Leslie, was “one which I have heard him say he painted entirely in the open air.” It shows a ‘lighter’—a large open boat used to load and unload



Fig. 4: Boatbuilding at Flatford Mill (1814), John Constable, ©V&A Museum, London. Public Domain.

ships—being built in Constable’s father’s boatyard, just by Flatford Mill in Suffolk.

The river Stour which divides Suffolk from Essex—runs across the center canvas—a blue, glistening line. To the left, a clump of trees, and the wooden gates of Flatford lock. Making its way along the river from the lock is a narrow barge, steered by a man in blue. It tows a smaller boat behind it.

The description adheres to the ‘rules’, and has in the past been well received, and there is now a tactile version available:

Locate the left edge of the frame. About half way two markers indicate the position of the river Stour. This runs horizontally across the image and is textured with lines. There is a tall tree to the left, below which is the wooden gate of Flatford lock. Making its way along the river, about a third of the way across the image from the left, is a barge, towing a smaller boat behind

it. Above this find the gently undulating horizon line. As you follow this low horizon towards the right, you will find the forked tree. Other taller trees further to the right frame the side of the image, their trunks extending down to the bottom right corner of the image. About three quarters of the way down — just to the left of the trunk — is the figure of the little girl.

Clearly a blending of the versions should provide an excellent service for blind and sight-impaired persons.

7 The Revoltella Museum

Another University of Trieste project (FRA 2017; financed by the Friuli-Venezia-Giulia Region) concerned the sculptures and paintings housed in the Museo Revoltella, actually in Trieste. This museum has no audio description and thus the objective pursued (by Rosanna Lopodota, a student writing her graduate thesis) was to provide such a service. So how could an amateur cope? First, she availed herself of another set of guidelines provided by an organization called “Art beyond Sight”. These consisted of the following pointers:

- establish the length of your description with simple syntax/lexis;
- provide basic information (artist, nationality, title, date, technique, materials, size, history, place);
- describe dimensions e.g., the size of an egg; entering St Paul’s Cathedral (20m, 50m, 100m?)
- provide a general picture (subject, shape, color)
- orient the patron, possibly using touch;
- describe technique and material;
- describe style;
- integrate sound effects; (Giansante, 2015).

And for good measure, she also consulted the ADLAB project guidelines, which cover much of the same ground but in a more user-friendly style:

Present the painting you are going to describe with a few facts (identification, artist, date, style, technique ...);
 highlight what makes it special or unique;
 give a general impression of the whole picture and then take the listener through a “journey” that may build a narrative or simply go through the



Fig. 5: *Estate* (1936), Marcello Mascherini, Museo Revoltella, Trieste. Reprinted by permission of Museo Revoltella.

elements that make up the painting, a kind of visual transitivity; describe important and interesting features and highlight details by relating technique, color, stroke and other technical features to the effect that is produced (be careful not to be over technical and minute); relate the painting to other pieces in the exhibition or the work of the artist and, if possible, attract the listener to other related works (Remael et al., 2007).

The methodology then employed began with a visit to the chosen museum to select the works to be described. The exhibits were chosen in consultation with the Museum Director and the Project Organizer and based largely around which items were considered the most prestigious, but also which were more suitable for audio description. This was followed by documentary research in the Museum Library and further research in other sources, including contact with an art expert. Only at this point was work on the description initiated. The first artifact described was a sculpture by Marcello Mascherini named *Estate* (1936) (see Figure 5). The phases outlined above (consulting guidelines, studying the museum catalog, meeting with art expert) are partially presented in Table 2.

Tab. 2: *Estate* (1936) by Mascherini

Guidelines	Catalog	Standard Audioguide	Art Expert	Other
Basic information (title, author, etc.)	Estate, Marcello Mascherini (Udine 1906) etc.	Not available	Not requested	Not researched
General summary description	Not available	Not available	The work is of a life-size young woman...	Not researched
Technique and material	Modelled in wax, fired in bronze (Arch. Amm.) <i>Estate</i> : 1936 bronze merges realism and knowledgeable evocation in the type of material used. (Salvagnini 2007)	Not available	The movement of the figure, expressed in its flexibility and gestuality has the power to invite us into the scene...	Not researched
Style, use of colors and tones, choice of motifs	(De Micheli in Mascherini sculpture europeo 1988, 16) the idea of the myth clearly connected to life, enables him to experience the simple live emotions which even the search for more complex stylistic solutions would not be able to affect ...	The artist's intention is to capture an instance of life emblematic of the summer season and relive it every time the viewer observes the work, catapulting him into the scene in the full vivacity and joy of a fragment of a summer's day...	To produce a modernly Martini style sculpture it was not enough to copy Martini. You had to follow the route Martini had taken. For this reason, between 1931 and 1935, Mascherini looked intensely at Etruscan bronzes and ancient sculpture	

The resulting AD took the following form (English translation in brackets):

Estate é un'opera dello scultore udinese Marcello Mascherini. Il bronzo é stato realizzato nel 1936 e nello stesso anno il Duca d'Aosta lo rese in dono al museo. L'opera ha un'altezza di 142 cm ed é stata esposta per la prima volta alla XXI Biennale di Venezia in una sala attigua a quella personale dell'autore. L'opera raffigura a grandezza naturale una giovane bagnante seduta e intenta ad attirare con il suo grido l'attenzione di qualcuno. Il modo in cui é rappresentata suggerisce che sia appena emersa dall'acqua. La ragazza è nuda. I capelli sono raccolti in una coda disordinata e poggiano sulle spalle come se fossero bagnati. La posizione degli arti inferiori ci lascia immaginare che sia seduta su un muretto con le gambe a penzolini....

(*Estate* is a work by the Udine sculptor Marcello Mascherini. The bronze statue was cast in 1936 and in the same year the Duke of Aosta donated it to the museum). It is 142 cm high and was first exhibited at the XXI Venice Biennale in a room next to that of the sculptor himself This work is a life-size statue of a young seated bather trying to attract someone's attention by shouting. The way in which the subject is presented suggests she had just come out of the water. The girl is naked. Her hair is tied up in an untidy pony-tail and falls on her shoulders as if wet. The position of the lower limbs seems to show that she was seated on a wall with her legs dangling (dots). (Lopedota, 2017, 113)

From the outset further accessibility can be achieved immediately by an invitation to touch the statue (there being no danger of damage to the bronze), and by the providing of a background of atmospheric music of a classical nature.

8 The Guggenheim Museum, Venice

The last example presented here deals with the particularly difficult task of describing and providing access to abstract art. Paul Klee's *Portrait of Mrs. P in the South* in Figure 6 is recognizable as a portrait of a woman but is somewhat abstract in its composition. The Guggenheim Museum in Venice provides an audio description of this work; the research project detailed here (Zilio, 2017) had the aim of examining this description, testing its effectiveness and possibly modifying it.

The methodology employed consisted of a five-phase approach. Firstly, the official audio description would be transcribed and tested out on sighted patrons, gaining their feedback. As a result of this feedback the text would be modified and



Fig. 6: *Portrait of Mrs. P. in the South* (1924), Paul Klee, Guggenheim Museum, Venice. Public Domain.

then presented to a blind audience. Their feedback would then provide input for a final modification. Relevant excerpts from the original AD of both the painting and the relief form designed for tactile exploration can be seen below (English translation in brackets):

Ritratto di Frau P. nel Sud, dipinto nel 1924. Disegno ad acquerello e ricalco a olio su carta montata su tavola dipinta a guazzo. La tecnica del guazzo impiega un tipo di pigmento simile alla tempera, reso più consistente dall'integrazione di gesso più un composto di gomma arabica. Le dimensioni sono 42,5 centimetri di altezza per 31 centimetri di lunghezza, compresa la montatura. Si trova nella Collezione Peggy Guggenheim di Venezia.

Guida all'esplorazione tattile: La vacanza in Sicilia dell'estate del 1924 fornisce a Paul Klee gli spunti per la realizzazione di vari acquerelli. Questo

dipinto, riprodotto a rilievo in resina e termoform per la lettura tattile, rappresenta la composizione stilizzata del busto di una donna con cappello su uno sfondo lievemente ruvido per evocare e ricordare l'atmosfera calda del Mediterraneo ... I tratti somatici sono resi con linee abbozzate. I capelli, vagamente ricci, sono appena accennati sul lato sinistro ... Gli occhi sono rotondi e appena accennati ma riconoscibili ... Il busto ha una pettorina semicircolare, solcata a metà da una fascia orizzontale ruvida che ricorda la sabbia ... La forma a cuore sul petto di Frau P. é un motivo ricorrente nell'opera di Klee, dove, a seconda dei casi, rappresenta una bocca, un naso o un busto.

(Portrait of Mrs. P. in the South, painted in 1924. Watercolor and oil tracing on paper 'a gouache'. The 'gouache' technique employs a type of pigment similar to tempera, made more consistent by the integration of plaster plus a mixture of gum Arabic. It is 42.5 cm high and 31 cm wide, including the frame. It is to be found in the Peggy Guggenheim Collection in Venice.

Guide to the tactile exploration: His 1924 Sicilian holiday provided Paul Klee with the stimulus to produce a number of watercolors. This painting, reproduced in resin relief for tactile reading, is a stylized composition of the bust of a woman in a hat on a slightly rugged background that evokes and recalls the warm atmosphere of the Mediterranean ... The facial features are sketched in. The hair, vaguely curly, is just visible on the left... the eyes are round and small, but recognizable... There is a semicircular collar round the neck crossed by a rough line recalling sand... The heart shape on Frau P.'s breast is a recurring theme in Klee's work where it can represent a mouth, a nose or a bust.)

The eleven sighted viewers, young men and women, most of whom studying translation and audio description at the University of Trieste, were asked to express their views as to whether elements of the description were described in a satisfactory or unsatisfactory manner. The elements regarded as key or at least relevant were background, explanation of terms, the face, the head, the hair, the hat, the posture, the heart on the breast, the colors, and the organization of the information.

Table 3 shows the comments of three of the sighted viewers.

Tab. 3: Opinions of sighted patrons on the audio description of *Portrait of Frau P. in the South*

1	The AD was too fast. It is not clear what is meant by ‘the warm atmosphere of Sicily’, I imagined the sea. The heart on the breast isn’t made clear, perhaps it would suffice to say that there is a design on the chest. The line representing sand and its relevance are not made at all clear.
2	The colors are not described, should I imagine the painting in black and white? The lack of realism is not made sufficiently clear. The speed is OK. I didn’t understand the reference to the line representing sand. The background is poorly described. The top to bottom style of the description was followed. What does ‘the eyes are recognizable’ mean?
3	I don’t know why I imagined her laid out on the beach. I was able to visualize the colors in the background. I didn’t understand the ‘heart on the breast’ bit. Too much detail. Perhaps more information on the artist.

As can be seen, some opinions differed, but taking the sample as a whole, there was considerable agreement. One concrete example can be seen in the text extract:

uno sfondo lievemente ruvido per evocare e ricordare l’atmosfera calda del Mediterraneo

Most viewers agreed that the reference to Sicily and the warm atmosphere of the Mediterranean had induced an image of landscape, whereas the description is merely of a background. The AD was thus modified:

Si trova su uno sfondo uniforme di colori caldi, prevalentemente arancione e marrone. Lo sfondo è stato reso con lievi sbalzi e macchie di colore rarefatte, poco dense, che cercano di riprodurre l’irreale. Esso è inoltre lievemente ruvido per evocare e ricordare l’atmosfera calda del Mediterraneo.

(it is on a background of uniform warm colors, prevalently orange and brown. The background is rendered by slight movements and patches of rarefied, not very dense colors, in an attempt to reproduce the unreal. It is also a little rugged to evoke and recall the warm atmosphere of the Mediterranean)

The modified description of the background was well received by the blind group consisting of eight members of the Trieste and Padua branches of the Italian Blind Association. It was judged satisfactory by all with only the request for a few more details e.g., the color of the woman’s hair. Of course, this particular feedback from the blind people themselves was crucial in that “no-one can imagine another’s life

well enough to develop services for them without involving them directly in that development” (O’Neil, 2008, 25).

And the final version:

Questo dipinto è anti-naturalista e rappresenta la composizione notevolmente stilizzata del busto di una donna, posta frontalmente. La donna indossa un cappello e si trova su uno sfondo uniforme di colori caldi, prevalentemente arancione e marrone. Lo sfondo è stato reso con lievi sbalzi e macchie di colore rarefatte, poco dense, che cercano di riprodurre l’irreale. Esso è inoltre lievemente ruvido per evocare e ricordare l’atmosfera calda del Mediterraneo. I tratti somatici sono spesso resi con linee abbozzate. I capelli, vagamente ricci e biondo-arancio ...

(This painting is anti-naturalist and is a notably stylized composition of the bust of a woman, seen from the front. The woman is wearing a hat and is seen on a background of uniform warm colors, prevalently orange and brown. It is rendered by slight movements and patches of rarefied, not very dense colors, in an attempt to reproduce the unreal. It is also a little rugged to evoke and recall the warm atmosphere of the Mediterranean. The somatic features are often rendered by sketched lines. The hair, slightly curly, is blonde-orange ...)

Again, the spoken and tactile experience could be augmented by a musical background (more contemporary in this case) and even by a sense of smell evoking the Mediterranean atmosphere, that of the sea and typical vegetation for example.

Other elements that received noticeable modification were the descriptions of the hair, the hat, the line representing sand, the heart shape, the colors, and the contextual information regarding, for instance, the symbols used by Klee. One of the encouraging aspects of the experiment was that the modifications implemented following the feedback from the sighted viewers were appreciated by the blind, thus dispelling the fear that the latter would have radically different needs.

Thus, in all these ways, the blind patrons receive a product that is the result of a continuous process which could be extended further ad infinitum, but there comes a point when feedback, intuition and analysis tell us we have reached an optimum, or at least a more than satisfactory level of accessibility, a kind of critical multimodal awareness. This is, at least, to be hoped.

9 Conclusion

Through a series of examples this paper has sought to demonstrate how the new linguistic genre of audio description can be integrated by elements of the other senses in order to provide a fully accessible product for the blind and sight-impaired community, and not only. ‘Talk and Touch’ tours provide the already quite widespread blueprint for future AD in museums, but the use of other senses such as smell and even taste (imagine a glass of mead being offered at a medieval history exhibition) can be exploited and go even further in the attempt to provide an experience equal to that of the sighted visitor. At a recent exhibition at the Archeological Museum Udine, Italy (*Women, Mothers, Goddesses*), a strip similar to the kind used to advertise perfumes in duty free shops, exuding the essence of a Roman scent, was made available to patrons as they moved round the ancient artifacts.

In the first example presented above, the audio description of Constable’s painting refers to both its original two-dimensional form and its relief model. The statue in the Revoltella was described by an amateur ‘learning the ropes’ following a number of published guidelines that are now available, but also using a methodology including research in the museum catalog and contact with a history of art expert. The statue can be touched, and a musical background can be provided in request. Finally, it can be seen in the AD of the Paul Klee painting that the multimodal form of the art work provides more than just color and image but also atmosphere in order to provide the blind person with as authentic an experience as possible. Audio description provides practical instances of the interplay of sensory channels, which must in some way feed into any theory of multimodality.

Bibliography

- Benecke, Bernd. 2004. Audio Description. *Meta. Translator’s Journal* 49(1). 78–80.
- Braun, Sabine. 2011. Creating Coherence in Audio Description. *Meta* 56(3). 645–662.
- De Coster, Karin & Volkmar Mühleis. 2007. Intersensorial Translation: Visual Art Made up by Words. In Jorge Diaz-Cintas, Pilar Orero & Aline Remael (eds.), *Media for All: Subtitling for the Deaf, Audio Description, and Sign Language*, Amsterdam: Rodopi. https://doi.org/10.1163/9789401209564_014.
- Dobbin, Clair, Alison Eardley & Joselia Neves. 2016. Ektashif – Art through the Senses: Families Shaping Museum Programming in Qatar. *Multaqa: Professional Journal of the Gulf Museum Educators Network* 2. 11–19.
- Eardley, Alison, Clara Miniero, Joselia Neves & Peter Ride. 2016. Redefining Access: Embracing Multimodality, Memorability and Shared Experience in Museums. *The Museum Journal* 59(3). 262–286.
- Fryer, Louise. 2016. *An Introduction to Audio Description* Abingdon. London: Routledge.

- Fryer, Louise & Johnathan Freeman. 2012. Presence in Those With and Without Sight: Audio Description and its Potential for Virtual Reality Applications. *Journal of Cyber Therapy and Rehabilitation* 1(15–23).
- Giansante, Lou. 2015. Writing Verbal Description Audio Tours. Last accessed: 17 September 2019. <http://www.artbeyondsight.org/mei/verbal-description-training/writing-verbal-description-for-audio-guides/>.
- Grimberg, Ruth. 2015. *Across Still Water*. London: BFI Network.
- Hirose, Kojiro. 2013. Research on Methods of ‘Touching the World’ – The Aim of the Exhibit Area of Tactile Learning in Japan’s National Museum of Ethnology. *Disability Studies Quarterly* 33(3). 170–185. <https://doi.org/10.18061/dsq.v33i3.3743>.
- Holland, Andrew. 2009. Audio Description in the Theatre and Visual Arts: Images into Words. In Jorge Diaz-Cintas & Gunilla Andrman (eds.), *Audiovisual Translation: Language Transfer on Screen*, London: Palgrave Macmillan.
- Holsanova, Jana. 2016. A Cognitive Approach to Audio Description. In Anna Matamala & Pilar Orero (eds.), *Researching audio description*, 49–73. Basingstoke: Palgrave Macmillan.
- Hutchinson, Rachel. 2016. *Enhancing Museum Visits with Audio Description: Challenges and Opportunities*: University of Westminster dissertation.
- Independent Television Commission. 2000. ITC Guidance on Standards for Audio Description. Last accessed: 17 September 2019. <http://www.ofcom.org.uk/static/archive/itc/itcpublications/codesguidance/audiodescription/index.asp.html>.
- Kress, Gunther. 2003. *Literacy in the New Media Age*. London: Routledge.
- Kress, Gunther & Theo van Leeuwen. 1996. *Reading Images: The Grammar of Visual Design*. London and New York: Routledge.
- Kubrick, Stanley. 1980. *The Shining*. Warner Bros. and Hawk Films.
- Lopedota, Rosanna. 2017. *Tradurre l’arte in parola: L’audiodescrizione al Museo Revoltella di Trieste*. Trieste: Museo Revoltella.
- Mazur, Ivona & Agnieszka Chmiel. 2012. Towards Common European Audio Description Guidelines: Results of the Pear Tree Project. *Perspectives: Studies in Translatology* 20. 5–23.
- Neves, Joselia. 2015. Additional Hints for Designing Descriptive Guides. In Aline Rемаel, Nina Reviere & Gert Vercauteren (eds.), *Pictures Painted in Words: ADLAB Audio Description Guidelines*, Trieste: EUT.
- O’Neil, Mark. 2008. Advancing Museum Practice. In Jocelyn Dodd, Richard Sandell, Debbie Jolly & Ceri Jones (eds.), *Rethinking Disability Representation in Museums and Galleries*, Leicester: RCMG, University of Leicester.
- O’Toole, Michael. 1994. *The Language of Displayed Art*. London: Leicester University Press (Pinter).
- Piazza, Gala. 2017. “non si vede bene che con le orecchie: L’arte è invisibile agli occhi”: Le audiodescrizioni per ipovedenti delle opere del victoria and albert museum di londra: University of Trieste dissertation.
- Remael, Aline, Nina Reviere & Gert Vercauteren (eds.). 2007. *Pictures Painted in Words: ADLAB Audio Description Guidelines*. Trieste: EUT Edizioni Università di Trieste.
- Samis, Peter & Mimi Michaelson. 2017. *Creating the Visitor-Centred Museum*. London: Routledge.
- Secchi, Loretta. 2004. *L’educazione estetica per l’integrazione*. Roma: Carocci Editore.
- Secchi, Loretta. 2014. Between Sense and Intellect. Blindness and the Strength of Inner Vision. In Derrick de Kerckhove & Cristina Miranda de Almeida (eds.), *The Point of Being*, 197–212. Newcastle upon Tyne: Cambridge Scholars Publishing.

- Snyder, Joel. 2008. Audio Description: The Visual Made Verbal. In Jorge Diaz-Cintas (ed.), *The Didactics of Audiovisual Translation*, 191–198. Amsterdam: John Benjamins.
- Taylor, Christopher. 2012. From Pre-Script to Post-Script: Strategies in Screenplay Writing and Audio Description for the Blind. In Fiona Dalziel, Sara Gesuato & Mariateresa Musacchio (eds.), *A Lifetime of English Studies: Essays in Honour of Carol Taylor Torsello*, Padova, Italy: Il poligrafo.
- Taylor, Christopher. 2017. Reading Images (Including Moving Ones). In Tom Bartlett & Gerard O'Grady (eds.), *The Routledge Handbook of Systemic Functional Linguistics*, London: Routledge.
- Thibault, Paul J. 2000. The Multimodal Transcription of a Television Advertisement: Theory and Practice. In Anthony P. Baldry (ed.), *Multimodality and Multimediality in the Distance Learning Age*, 311–385. Campobasso, Italy: Palladino Editore.
- Udo, JP & Deborah Fels. 2009. Re-Fashioning Fashion: An Exploratory Study of a Live Audio-Described Fashion Show. *Uni. Access Inf. Soc* 9(1). 63–75.
- Vandaele, Jeroen. 2012. What Meets the Eye. *Cognitive Narratology for Audio Description. Perspectives: Studies in Translatology* 20. 187–102.
- Vercauteren, Gert. 2007. Towards a European Guideline for Audio Description. In Jorge Diaz-Cintas, Pilar Orero & Aline Remael (eds.), *Media for All: Subtitling for the Deaf, Audio Description, and Sign Language*, 139–149. Amsterdam: Rodopi.
- Vercauteren, Gert & Pilar Orero. 2013. Describing Facial Expressions: Much More than Meets the Eye. *Quaderns. Revista de Traducció* 20. 187–199.
- Vercauteren, Gert & Aline Remael. 2014. Spatio-Temporal Setting. In Anna Maszerowska, Anna & Pilar Orero (eds.), *Audio Description: New Perspectives Illustrated*, 61–80. Amsterdam: John Benjamins.
- Wood, Elizabeth & Kiersten Latham. 2014. *The Objects of Experience: Transforming Visitor-Object Encounters in Museums*. Walnut Creek: Left Coast Press.
- Zilio, Serena. 2017. *Audiodescrizione e pinacoteche: elaborazione di modelli pratici*: University of Trieste dissertation.

Edward Larkey

Narrative as a Mode of Communication: Comparing TV Format Adaptations with Multimodal and Narratological Approaches

Abstract: This chapter illustrates how a multimodal approach can be combined with narratological knowledge to compare the narrative sequencing, structure, and content of culturally different versions of a sketch television comedy series. The series features archetypical interactions and conflicts of a 30-something heterosexual couple. By combining both narratological and multimodal approaches to compile qualitative and quantitative data on durations of scene and shot sequencing, narrative content, and culturally specific multimodal relationships, the chapter compares television format adaptations cross-culturally to distinguish a variety of discursive positionings towards gender roles, masculinity, and family conflict management strategies as a first step in the process of cross-cultural comparisons. These methodological innovations make a substantial contribution to determining the intercultural and cross-cultural dimensions of multimodal analyses and help concretize and more precisely delineate the nature of globalization processes, global discourses, and hybridization. Furthermore, it is suggested that combining computer software-driven multimodal analyses with narratological approaches offers a more precise and objective foundation for cross-cultural comparative analyses of increasingly prevalent localizations of television formats.

Keywords: multimodality, TV series, adaptation, narrative, cross-cultural

1 Multimodality and Television Format Adaptations

A semiotically based notion of multimodality is an appropriate approach to cross-culturally analyzing television format adaptations for a variety of reasons. Aveyard & Moran (2015) mention the dual nature of the television format as a universal, mobile, and adaptable global cultural product on the one hand, with a capacity for domestic and local modification to achieve cultural proximity to the intended audience on the other:

[...] the new programme's performers and participants will, for the most part, be ethnically familiar, speak one or the other dominant territorial languages; be visually and culturally

<https://doi.org/10.1515/9783110608694-009>

anchored in recurring, everyday locations; deal with recognizable situations and issues and behave in customary and familiar ways. (Aveyard & Moran, 2015, 689)

Television formats contain various modes of communication as socially shaped and culturally given semiotic resources for making meaning (Kress, 2010). Multimodal scholars such as Gunther Kress point to the intercultural dimension of multimodality by emphasizing that modes result from the social and historical shaping of materials that differ from one society to another (Kress, 2010, 11). Kress also reminds us that individuals “with their social histories, socially shaped, located in social environments, using socially made, culturally available resources, are agentive and generative in sign-making and communication” (Kress, 2010, 54). Kress furthermore mentions that semiotic resources function differently in different societies since they “differ from culture to culture” (Kress, 2010, 168) such that a particular semiotic resource in one culture may be different to that resource in another (Kress, 2010, 81). Television format adaptations illustrate this admirably, but we need the help of a multimodal analysis to uncover the interrelationships between culture, society, and representation at work in the adaptation of the television format in different countries and societies.

2 Narrative and Cross-Cultural Multimodality

Jewitt (2009, 13) underscores the potential parity among all modes but also the variety of complex interactions and relationships between different modes of communication, concluding that “the interaction between image and writing in a text” has driven much existing multimodal research. Jewitt also provides an approach to intercultural and cross-cultural multimodality by positing that people “orchestrate meaning through their selection and configuration of modes” (Jewitt, 2009, 15). Similar to Kress, we can thus postulate that television format adaptations exhibit inter- and cross-culturally salient and varying configurations of the ensemble of modes, as I will demonstrate further below.

Our approach will therefore follow Jewitt’s advice for multimodal research to investigate the “interplay between modes to look at the specific work of each mode and how each mode interacts with and contributes to the others in the multimodal ensemble” (Jewitt, 2009, 25). In the context of transnational format adaptations, the interplay between modes and their interactions with each other may exhibit culturally specific patterns and interconnections. While I will not be going into detail into each culturally specific pattern in this chapter, I will show at which point a multimodal comparative analysis would be initiated. I will illustrate how

this happens in various versions of the Canadian TV series *Un Gars, Une Fille* ('A Guy and a Girl') produced from 1997 to 2002 in the francophone Canadian province of Quebec. The two main protagonists of the series will be designated as 'the Guy' and 'the Girl' and, in the specific scene I will be investigating, I will also refer to the 'Mother-in-Law' when referencing the mother of the Girl from the standpoint of the Guy, and as the 'Mother' when referencing the character from the standpoint of the Girl. The analysis will be carried out particularly with regard to the dialog and the music in a shot in which the Girl confronts her Mother about traumas induced by the Mother during her childhood.

For the purposes of the chapter, I will follow the suggestions of Bateman et al. (2017, 132) about the nature of 'texts' as concrete artifacts utilizing semiotic modes to produce a predesignated and anticipated organization of material to guide an interpretation based on those modes. I will consider each different version of the series *Un Gars, Une Fille* to be just such an 'audiovisual text'. The collection of texts investigated in this study can then be considered a "corpus" (Bateman et al., 2017, 152) available for computer annotation with the help of software, even though these texts are technically speaking not necessarily 'naturally occurring' but scripted in advance and intentionally adapted to culturally specific situations. Applying the dual concept of discourse proposed by Bateman et al. (2017, 133) which the authors call the 'big D', defined as ways of thinking about a particular concept or principle (they mention 'discourse of gender'), delineated from 'small d' discourse as a local, text-based, and fine-grained analysis. In our investigation, therefore, I will look at the instances of 'small d' discourse in each of the versions of the series contained in the particular scene, while I will take a comprehensive view of that same scene in several versions to characterize various positionings within contested 'big D' discourses of gender roles and family relationships.

The multimodal part of this investigation thus focuses on the micro-level, collecting and discussing "observable traces of meaning-making" (Jewitt et al., 2016, 7) in a particular scene in the various versions of the television format adaption of *Un Gars, Une Fille*. I selected a scene common to most versions and used digital tools (see Bateman et al., 2017; Lim Fei et al., 2015) to compile and correlate quantitative and qualitative data on the durations of segments within the scene, as well as the scene segments and their placement themselves. This shot scene is significant in the narrative not only due to its greater length, but also because it most vividly depicts the values, personalities, gender expectations, and relationships of the Girl with her Mother to a greater extent than other scenes.

Finally, in arguing for the inclusion of narrative structure, content, and sequencing into our analysis, I will rely to a large extent on guidance provided by Bateman & Schmidt (2012), who not only define basic units of measurement in film

and televisual analysis such as shot, scene, and scene shot¹ but also supply the conceptual foundations for approaching multimodality beyond the micro-level. They do this by introducing and reflecting on the Metzian theories focusing on spatiality, temporality, and sequentiality. These are especially crucial for cross-cultural multimodal analyses since they highlight a mesolevel of analysis which, in addition to the micro-level, looks at structures and sequences of televisual narratives—a process they call ‘layouting’—which

are constructed in ways that guide interpretation *even prior* (italics in original, EL) to handing over the task of understanding to some viewer’s ‘common sense’—such that lines of interpretation are—not closed off, remaining open and potentially relevant for understanding” (Bateman & Schmidt, 2012, 1).

Our depictions of the narrative structure of the shot in the Mother-in-Law scene(s) of *Un Gars, Une Fille* will focus on shot sequencing as the prime textual organization of the narrative. Our task in multimodally comparing the selected scene and its position in the episode is simplified to a certain extent in that the scene is monospacial and explicitly staged, taking place in the dining room of the Mother-in-Law, while the temporal sequencing is implicitly constructed but made more explicit in our re-constructed sequencing. Our analysis is further simplified by the fact that the camera—in most versions—is largely immobile, with either a point-of-view shot of the Guy and the Girl facing the Mother-in-Law, or an over-the-shoulder shot of them with the Mother-in-Law either not in the frame, or only barely visible. I will also highlight the complicated and ambiguous role of the transitions between the shots for creating emotional, topical, temporal, and sequential bridges.

Un Gars, Une Fille has been reproduced with domestic casts, settings, and languages in 28 other countries and regions to date. Table 1 contains a list of official adaptations produced up to the present, while Table 2 contains a list of more or less blatant unofficial and unauthorized versions ‘inspired by’ *Un Gars, Une Fille*.

¹ It should also be mentioned here that I am not employing these concepts exactly as they are defined by Bateman and Schmidt. I will be using ‘segment’ in a more general sense to mean any kind of measurable spatiotemporal portion of the televisual narrative, while the ‘scene’, due to the idiosyncratic nature of the shot structure and sequencing which includes both topical as well as spatiotemporal configurations in the television series, is loosely defined by the transitions inserted into the narrative to both distinguish as well as bridge topical, emotional, and temporal gaps.

Tab. 1: All officially recognized versions of *Un Gars, Une Fille* and their adaptations until the present Courtesy of Avanti Cine Groupe

Number	Country	Date of First Broadcast
1	Canada	1 May 1997
2	Belgium	August 1998
3	France	11 October 1999
4	Sweden	November 1999
5	Portugal	May 2000
6	Greece	Autumn 2000
7	Spain	22 September 2000
8	Bulgaria	2002
9	Hungary	2002
10	Netherlands	2002
11	Israel	9 July 2002
12	English Canada	September 2002
13	Poland	3 September 2002
14	Germany	13 October 2002
15	Mexico	31 July 2003
16	Russia	20 September 2003
17	Italy	15 December 2004
18	Ukraine	14 January 2005
19	Lebanon	27 February 2006
20	Latvia	20 October 2006
21	Lithuania	29 August 2007
22	Turkey	2 October 2008
23	Cyprus	8 October 2010
24	Kazakhstan	8 October 2012
25	Abu Dhabi	July 2013
26	Czech Republic	10 September 2013
27	Serbia and Montenegro	5 October 2015
28	Slovenia	6 March 2016
29	French Africa	13 November 2017

The chapter will compare the particular scene selected for analysis across 14 different versions² in which the couple visits the Girl's Mother. The scene is divided into two parts, each positioned differently both within the approximately 24-minute

² Besides the original Quebec version this paper will draw on the Spanish, Italian, Turkish, French, Polish, Russian, Ukrainian, Latvian, Serbian, Slovenian, Greek, German, and Israeli versions. The French, Spanish, and Italian versions were produced as short segments between longer series in the respective broadcast schedule and are therefore strictly speaking not episodic. In spite of that, I will draw upon relevant examples from the Spanish version to illustrate relevant analytical points.

Tab. 2: These are “non-official” versions of *Un Gars, Une Fille* produced in various countries without authorization or legal foundation

Title of Series	Country of Origin	Year(s) of First Broadcast
Sousou et Nounou	Algeria-Arabic	2007
Timsal n Wexxam	Algeria-Kabyle	2014
Le Couple	Morocco	2013
Ilyz m'tsam	Madagascar	2013–2018
Love Bytes	India	2015–2017
Mann/Frau	Germany	2014–2015
Bent Walad	Tunisia	2012
Polovinki	Russia	2012–2013
Sasho i Sashka	FYR Macedonia	2005–2006

episodes, and distributed differently among the episodes of those versions, as can be seen in Table 3.

All 29 official versions of the series feature interactions and conflicts between the non-married, yet cohabitating childless 30-something heterosexual couple as it negotiates its relationship between the pressures of what family relations researchers have called the ‘de-institutionalization of marriage’ and the erosion of the ‘heterosexual marriage monopoly’ (de Vaus, 2011). Contemporary couples, including those depicted in various versions of *Un Gars, Une Fille*, must continually re-negotiate both the division of domestic labor as well as the degree of autonomy of each participant in the relationship, without necessarily relying on behavioral patterns and ways of thinking of previous generations. Still, Allen & Walter (2000, 4), summarizing their own and others’ previous research, have determined that “both old and new ideas about gender and old and new gender practices coexist” in the same families simultaneously. Therefore, various versions of *Un Gars, Une Fille* display a wide variety of culturally specific negotiating strategies over gender roles, conflicts, and negotiations within the narrative framework of the television format. While I will not delve into the cultural specifics of each depiction, I will display the differences which form the starting point of further analyses of culturally specific negotiating strategies.

The life of the Guy and the Girl deviates from the traditional sequence of events for family formation, consisting of courtship, engagement, marriage/co-residence, sex, children, empty nest, etc. The couple illustrates what de Vaus points to as a re-ordering of those sequences which depart from conventional social scripts while pursuing the construction of one’s own life narrative. De Vaus underscores the ambivalent nature of this undertaking: on the one hand, its complexity can

Tab. 3: Table showing the lengths of the episodes containing versions of *Un Gars, Une Fille*, the lengths of the Mother-in-Law scenes within those two episodes, and the percentage proportion of the Mother-in-Law scenes within those episodes, along with the percentage of the pre-visit scenes to the both Mother-in-Law visit scenes. Notice the outlier German version which will not be scrutinized in this chapter. (Time details: mm:ss)

Versions	Total Length of Episode(s)		Total Length both Episodes	Length of Mother-in-Law Visit(s)	Proportion of Mother-in-Law Visit Segments to Episodes	Previsit Segment Lengths	Proportion of Pre-Visit Scene(s) to Visit Scene(s)
German	24:01			08:14	34.3%	07:41	93.3%
	Part 1	Part 2					
Quebec	23:24	22:50	46:14	14:20	31.0%	00:59	6.9%
Israeli	23:45	24:34	48:19	14:31	30.0%	00:51	5.9%
Bulgarian	23:14	25:21	48:35	15:55	32.8%	01:04	6.7%
Slovenian	27:24	27:26	54:50	14:17	26.0%	01:11	8.3%
Turkish	24:50	27:30	52:20	15:44	30.1%	01:25	9.0%
Ukrainian	25:04	27:58	53:02	16:26	31.0%	01:07	6.8%
Russian	22:07	21:37	43:44	14:53	34.0%	01:12	8.1%
Serbian	26:19	26:07	52:26	11:05	21.1%	01:10	10.5%
Latvian	21:09	23:06	44:15	15:12	34.4%	00:55	6.0%
Polish	21:38	21:30	43:08	14:39	34.0%	01:09	7.8%
English	23:00	23:00	46:00	14:07	30.7%	00:53	6.3%
Canada							

be welcoming and liberating, but it can also be “profoundly unsettling” (de Vaus, 2011), particularly without the guidance of those established social scripts. In addition, the continual re-negotiation of domestic roles can be both exhilarating yet exhausting and complicated for the couple.

The study will incorporate several of the narratological notions of exposition, narrative dilemma, cause-effect chain, and protagonist and antagonist as described by Butler (2012) to explain how the series achieves a genre-specific degree of narrative cohesion. The study will apply notions of chronicity, spatiality, and sequentiality (Bateman & Schmidt, 2012) to a multimodal comparison of the narrative structure, sequencing, and content to elaborate a loosened narrative cohesion compared to the tighter narrative of a situation comedy.

As film and television studies scholar Jason Mittell (2004, xiv)³ has stated, genre pre-determines the cultural and discursive expectations of viewers about the content and structure of the narrative, its cohesion and the outcome of narrative dilemmas. Scholars in the fields of television studies as well as in multimodality concur in this general outlook on the notion of genre. Bateman et al. (2017, 129) emphasize that genre creates patterns and conventions to accomplish communicative work, whereby narrative fulfills “storytelling work.” According to them consumers allocate artifacts to particular communicative events containing interpretive frames and expectations which assist in their meaning-making process. This is evident even in the idiosyncratic combination of the sketch comedy with a situation comedy found in *Un Gars, Une Fille*.

The narrative cohesion of a situation comedy features a return to an original state of equilibrium after the resolution of at least one of perhaps several dilemmas within the narrative. A sketch comedy, however, neither aims for a temporally configured narrative cohesion nor equilibrium consisting of, for instance, the beginning and a definitive end point of a storyline in each episode. Instead, viewers expect brief and perhaps non-sequential fragmentary comedic (parodic, satirical, ironic, or hyperbolic, stereotypical) illustrations of different disputes and encounters between the protagonist(s) and antagonist(s), without a resolution. *Un Gars, Une Fille* is a peculiar combination of sitcom and sketch comedy in that the short sketch segments may relate a coherent narrative, but these may be either incomplete or unresolved, in addition to their lack of diegetic sequentiality and linear chronicity. One of the crucial devices for creating narrative cohesion in *Un Gars, Une Fille* is paradoxically the specific type of *transition* between shots and scenes incorporating audio and visual elements.

The addition of multimodal quantitative and qualitative data can make a unique contribution to cross-cultural comparisons of the different versions of *Un Gars, Une Fille*. When combined with computer software-compiled quantitative and qualitative data, a multimodal approach achieves a precise measurement of durations of additions, modifications, and omissions in relevant narrative segments, particularly within the modes of dialog, music, voice, and length of shots and scenes.⁴ Multimodal quantitative and qualitative data also allow for more definitive and objective supplemental information for a cross-cultural comparative interpretive analysis. These data may also facilitate identifying previously unknown or obscured patterns and systems of multimodal interconnections and

³ Mittell states that genres “work to categorize texts and link them into clusters of cultural assumptions through discourses of definition, interpretation, and evaluation.” (Mittell, 2004, xiv)

⁴ For a more detailed study of different kinds of adaptations based on narrative structure, sequencing, and content see Larkey et al. (2016).

interplay in support of hermeneutic interpretation. The unique ensemble of communication modes works on a qualitatively different level than merely the additive sum of each separate mode. Finally, quantitative data on the durations of narrative segments and structure may help pinpoint and determine subtle but culturally relevant narrative and cultural differences and similarities.

In the narratological mesoanalysis I will utilize quantitative data collected with computer software, primarily Adobe Premiere Pro and Atlas.ti, with which approximately 10 narrative segments—shots and conversations—of multiple modes of communication were measured, correlated, and compared from one particular shot within the Mother-in-Law scene. A ‘meso’-description of shot sequencing will first describe a bathroom scene prior to the couple’s visit to the Mother-in-Law, in which the Girl has to seduce the Guy in order to convince him to visit her Mother due to the antagonism between him and her Mother. This shot in a sequence of others in the bathroom prepares the viewer for the later visit in the episode and pinpoints the emotional bond between mother and daughter as the primary motivation for the visit. The next step will be to examine the longest shot in most versions, a confrontation between the female protagonist, the Girl, and the female antagonist, the Mother or Mother-in-Law depending on which person in the couple’s relationship is being referenced. In this shot, the Girl initiates a conversation with her Mother about mother-induced traumas from her past. The detailed multimodal analysis will focus on the confrontation between the mother and daughter, during which the daughter breaks down and cries after encountering her mother’s insensitivities to her traumas as an adolescent. The addition of a ‘sad music’ cue during the segment can be read as guiding the viewer toward an identification with the Girl and directs the viewer’s sympathy away from the Mother.

2.1 Narrative Sequencing and Structure as ‘Meso’ Level

In the approximately 21- to 24-minute episodic versions of the series, the sequencing of the encounter in the Mother-in-Law scene(s) is divided into three distinct—and non-consecutive—segments:

1. the Girl’s aforementioned pre-visit seduction of the Guy to entice him to visit her mother, usually located in the same episode as Part 1;
2. Part 1 of the Mother-in-Law scene(s), which features an afternoon coffee and cake sitting at the Mother-in-Law’s, during which a major confrontation between the Guy and his Mother-in-Law erupts whereby the Guy attempts to get

at the heart of their animosity towards each other (for a detailed analysis of Part 1 see Larkey 2019)

3. Part 2 of the Mother-in-Law scene in a later episode in which the daughter tries to communicate with her mother about what the Girl asserts are mother-induced traumas in their relationship during the Girl's adolescence.

These three Mother-in-Law segments therefore have a beginning, i.e., the seduction of the Guy to induce him to visit the Girl's Mother, and an end point, in which the Guy and Girl leave the Mother-in-Law's home laden with leftover food from the visit. For the purposes of this investigation, I will consider these segments as a coherent narrative, even though both segments themselves are neither chronologically nor temporally consecutive, as can be evidenced by the different clothing worn by the couple in Parts 1 and 2. In addition, there is a temporal gap in the broadcast of the episodes between Part 1 and Part 2. In effect, therefore, we are re-constructing a chronologically coherent narrative from a non-sequential chain of shots and segments.⁵

A large portion of shots is devoted to discussions of and reactions of the Girl to the expectations and desires of her Mother for a grandchild. The Mother makes several references to her wish for family offspring, even going so far as to inquire about when they have sex, urging them not to use contraception. The lack of this step in the expected life-course of the couple generates some of the animosity between the Guy, who is less than enthused about the idea of his partner's possible pregnancy, and his Mother-in-Law. Several shots in the narrative center around a presumed pregnancy on the part of the Girl, whose assumption is based on a merely two-hour delay in the onset of her period. The Guy displays great relief when, at the end of the first part of the scene, the Girl announces the start of her period, to her great disappointment.

In almost all of the versions in the corpus, the selected scene takes place—monospatially (Bateman & Schmidt, 2012, 205)—in the dining room of the Mother-in-Law's home. In more recent versions such as the Slovenian, the shots take place in different rooms of the Mother-in-Law's house and are therefore 'multispatial', and later versions also move away from using a stationary camera. However, most

⁵ The Greek version diverges from this sequencing by adding two longer segments not part of the other versions: a longer segment discussing the internet in Episode 9, and a 07:40 minute segment in Episode 18 in an Italian restaurant visited with the Mother-in-Law. The confrontation between the daughter and her mother, i.e., the focus of this comparative study that is a contiguous sequence with other shots in Part 2, is inserted as a single shot into Episode 12 and therefore does not precede the room-renting segment as in the other versions.

versions feature a narrative sequencing of the segments of the scene which closely follows that of the original version but with minor modifications, additions, or removals which may or may not result from culturally specific factors.

Each shot—separated by a distinctive transition—of the Mother-in-Law scene(s) forms a part of what Jeremy Butler would call the ‘cause-and-effect-chain’ of the narrative dilemma, which would be the antagonism between the Guy and the Mother-in-Law as the protagonist and antagonist, first revealed in the seduction scene. The animosity serves both to justify the male protagonist’s unwillingness to visit the Mother-in-Law, and also raises the question as to why the antagonism exists between them. This animosity prepares the narrative foundation for the confrontations later on in the episode. Each segment of the scenes assigns culpability or justification for the mutual animosity between the two. Some segments appear to designate the one or the other ‘victorious’ in the contest, although ultimately there are no clear ‘winners’, and the relationship is never completely resolved in the segments under scrutiny here. This is particularly the case with the confrontation between the Guy and the Mother-in-Law in Part 1 (see also Larkey, 2018), but this also characterizes the dispute between the daughter and her mother in the segment of Part 2 in this study. Emerging from these squabbles is a multi-faceted field of conflict between the younger-generation couple and the older-generation Mother-in-Law, which assigns culpability for perceived deficits in family organization and performance to different actors as will be pointed out further below.

The bar graph (Figure 1) illustrates the narrative sequencing of the Part 2 segments of the Mother-in-Law scene(s) of 10 different versions. The larger bars of each version represent the different shots, while the narrower segments comprise the transitions. The transitions are an integral part of the narrative and consist of

1. a fade-out to black and fade-in to the next shot;⁶
2. fade-in of graphics of writing and objects commenting on the previous shot and fade-out of these graphics;
3. specific transition music cues and sound effects which act as punctuation and commentary to the previous shots.

6 The one exception to this pattern of transition is the English Canadian version, which has no fade-to-black, but instead superimposes the white writing portions of the transitions onto the non-faded clip as transition. One of the reasons for the varying lengths of the bar graph elements of each version is the imprecise location of the transition boundaries. Different modes of audio and visual communication may bleed into the next shot, or into the preceding shot. I have set the transition boundaries when at least two modes are still operating at once. The bleeding of single communication modes into the shot, e.g., dialog, signals the beginning of the new shot.

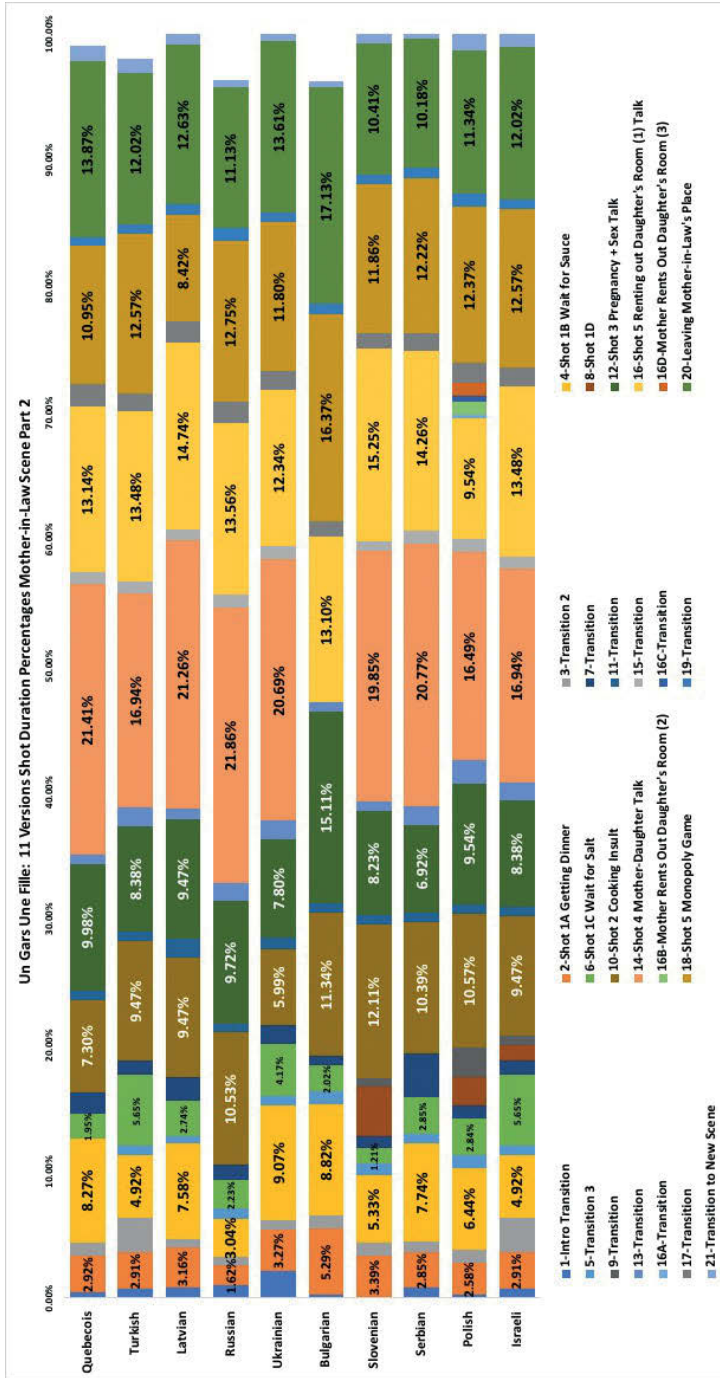


Fig. 1: Bar graph laying out the shots in 10 versions of *Un Gars, Une Fille*, Part 2, containing the confrontation between the daughter and her mother located in the middle of the scene. Notice the lack of this shot in the Bulgarian version.

Each of the scenes in the episodes is introduced by a complete fade-out to black of the previous scene, insertion of the series logo between the scenes, and fade-in to the first shot of the new scene. The bar graph, which is based on a table of extensive data compilation, demonstrates not only the lack of a Mother-Daughter conversation shot in the Bulgarian version, but also the additions in the Polish version, along with the relative lengths of the shots and transitions common to all versions. The transitions form and fill temporal, topical, emotional gaps to bridge and help construct a coherent televisual narrative by additively signaling the beginnings and ends of the shots. In the face of the monospaciality of the scene, the transitions are what Bateman & Schmidt (2012, 251) would call hypotactic, i.e., they help construct temporal and sequential relations between the shots, even as they—paradoxically—implicitly distinguish and separate each shot.

The final shot of the scene(s) culminates in a seemingly never-ending departure of the couple from the Mother-in-Law's, laden with leftovers from the meal, while the Mother-in-Law delays the departure, extending the conversation further by urging the couple to call her by telephone every day, even if she is not at home, leave a message on the answering machine, etc. The mounting impatience and exasperation of the Guy at the continually delayed departure climaxes in his declaration that he will go outside and warm up the car to prepare to leave. Replying to the Mother's and Girl's look of astonishment and surprise at the reference to the winter—the scene takes place in the summer—the Guy states that by the time they are finished the cold winter weather will already have arrived.

This analysis will now focus on the multimodal micro-level of the dispute between the mother and the daughter in the Part 2 section of the Mother-in-Law scene(s). As one can see in Table 3, the Mother-in-Law scenes comprise approximately 30% of the respective episode, with the Quebec original versions approximately 31% for both parts, while the adaptations vary between 26% for the shortest Part 1 segment (Israeli version) and 36% for the longest Part 1 segment (Bulgarian version). The differences in the Part 2 segments range from 25% for the shortest version (Slovenian) to 38% for the longest version (Russian). Due to the idiosyncratic nature of the Greek version explained above, I have not included this in the calculations of the table.

The components of the episodic versions of *Un Gars*, *Une Fille* exhibit the sketch comedy 'heritage' of what I call the 'non-episodic' version of between 5–10 minutes. The sitcom-like episodic versions usually consist of three different and unrelated scenes in the 21- to 26-minute episodes which are deployed as autonomous segments in their own right in the non-episodic versions, as bridges between longer series in their respective national television broadcasts. The seemingly random assemblage of scenes into "episodes" can be found in the non-episodic versions as well. Early adapted versions of the series—such as the Israeli version—combined

the different shorter segments to construct a 24-minute “episode” long enough to occupy a typical broadcast slot in that country, similar to the Italian, French, and Spanish versions.

As a part of this multimodal comparison of the different versions, we compiled and correlated both quantitative and qualitative data which include aspects of the dialogs, content, music, the behavior of the Guy, who is at first largely indifferent to the Girl’s growing emotional crisis but later slowly engages with his partner (the Girl), and the physical positions and movements of the Girl, the main protagonist during the discordant exchange. In all but the Greek version of the series, the Mother-in-Law is the clear antagonist in the conflict, while she ultimately deflects responsibility for the daughter’s increasing distress to accuse the Guy of culpability. Instead of assuming at least part of the responsibility for her daughter’s troubles, as in the Greek narrative, the Mother-in-Law in other versions reinforces her animosity towards the Guy by accusing *him* of causing the Girl’s anguish. The common thread running through the Mother-in-Law scenes is the performance and display of the animosity between the Mother-in-Law and the Guy, already signaled in the seduction scene prior to the actual visit.

Further below I will contextualize the Mother-in-Law scene(s) within the framework of the Girl’s contradictory efforts at attaining greater autonomy from her mother on the one hand, yet retaining the emotional bond with her on the other. The Guy, the male protagonist, exhibits qualities I attribute to what I call the ‘new patriarchy’ in which the Mother-in-Law disparages him for not fulfilling his role as main breadwinner and head of the family, seen as the result of greater agency and autonomy of the female protagonist prompted by feminist gains in re-defining gender roles. However, the new role, a result of more recent developments of families which deviate from older models of convention (same-sex couples, grandparent families, single-parent families, childless families, etc.), are less able to rely on conventional life-course scripts for gender role performance, which brings them into conflict with the conventional role models of the previous generation of parents. The tension between these generations is played out in the various versions of this series in a variety of culturally specific and different ways.

In all versions, the Mother-in-Law is out of frame and is therefore participating in the conversation only off camera as the sole mode of communication in the dialog. She therefore has no visible physical position, no gesture, no facial expression, no movement, with the exception of the Turkish version, which I will discuss further below. The only perception of the Mother-in-Law for the audience is a disembodied voice in the conversation and the reactions and movements of the Girl, who is the most important figure in the scene. This lack of embodiment places the Mother-in-Law at a distinct disadvantage in the conversation with regard to her visible authority and presence—with the exception of the Bulgarian version,

in which the Mother-in-Law is in the frame, albeit from a medium side shot. The *mise-en-scène* also configures the conversation between the Mother-in-Law and the couple as an interrogation of the couple by the Mother, in which the interrogator is out of sight and semi-anonymized. Although the Guy never participates actively in the conversation apart from at times following the turn-taking with his head movements and eyes, the Mother-in-Law nevertheless accuses him of causing the daughter's hardship. In the Ukrainian version, the viewer sees the extended arm of the Mother-in-Law in the final seconds of the shot, pointing at the Guy whom she accuses of causing the Girl's distress. With the exception of the Slovenian version and the leaving shot in the Spanish version, all the shots in Part 2 take place monospacially in the dining room of the Mother-in-Law.

The mother represents what Klein & Milardo (2000) define as a 'third party' in couple conflicts, i.e., friends or family members to whom authority is delegated in decisions of whom to ask for advice. According to Klein and Milardo the 'third party' offers informal guidance and helps interpret the couple's conflicts, while also endorsing strategic choices to promote measures and attitudes conducive to more accommodative strategies for resolving conflicts. In this connection I will ask what power the couple has delegated to the mother as a 'third party' participant, for whom does the mother advocate, and how is the mother's advocacy multimodally staged in particular versions?

2.2 The Girl Argues with her Mother: Multimodal Escalation of Emotionality

The major conflict in Part 2 of the Mother-in-Law scene consists of the daughter's attempt to justify and circumscribe her autonomy and independent selfhood and subjectivity vis-à-vis her mother by initiating a discussion about a (real-life) book on the topic entitled *My Mother, My Mirror*, by psychologist Laura Arens Fuerstein. Fuerstein proposes that the mother's—necessarily distorted—self-image is implanted in the daughter's mind and imprints her behavior during childhood as what she calls a "carnival-mirror self-image", in which the mother sees the daughter "through her own inaccurate lens" (Fuerstein, 2008, 4).⁷ Fuerstein's psychoanalytical approach is intended to help the daughter achieve greater autonomy

⁷ The original printing of the book was in 1996 and has had at least 10 printings, judging from the information in the book itself, even though 'first printing' is also imprinted. However, the actual book used by the screenwriters was by Nancy Friday (1977) entitled "My Mother, My Self. The Daughter's Search for Identity". This book goes into greater academic depth about the daughter's struggle for autonomy vis-a-vis the mother, focusing specifically on the discrepancy between

and a ‘truer self-image’: “When you see that your mother’s skewed image of herself no longer has to cause your skewed image of yourself, you gain a truer self-image” (Fuerstein, 2008, 4). Mentions of this book in the original and all other versions of the series—one of the longer shots in the series—initiates the discussion between the daughter and her mother.

A multimodal approach in analyzing the different versions of this shot documents the variety in the interplay of communication modes underlying the escalating emotionality in the shot and prepares the foundation for a comparative interpretative analysis of the different versions: the voice of the Girl, the behavior of the Guy, the music, the dialog content, and the positions and gestures of the Girl’s body during the conversation. The Girl’s voice evolves from a cheerful beginning through at least two stages—a stage in which her voice seems insecure and hesitant, and a further stage in which her voice starts to break—until she breaks down and cries at the table, bowing her head down and, in some cases, turning her body toward the Guy for consolation in his arms. Examples of these different responses can be seen in the screen shots in Table 4.⁸

The Guy’s body posture and actions shift from indifference—mild to extreme (looking at his mobile phone, eating, sitting silently at the table watching the two others)—at the outset of the shot into more or less actively consoling the Girl when she is upset, distressed, and crying. Only the Serbian version differs. In this version, the Girl herself de-escalates the emotionality and the Guy remains entirely unmoved—in all senses of the word⁹—throughout the entire conversation. With one exception—the Greek version—the Guy does not participate in the dialog between the Mother and the Girl but turns to the Girl to console her after she starts crying towards the end of the segment. In the Turkish version the Guy also verbally tries to calm his partner down by urging her to act appropriately for dinner. The Guy in the Ukrainian version also does not turn his body toward the Girl in

what the Mother says and what she does with regard to enjoying sexual pleasure as a woman and denying it as mother. Friday describes how mothers attempt to avoid this confusing deception when communicating with the daughter.

8 The screenshots on the left are taken at the beginning of the conversation, whereas the screenshots on the right are taken at its conclusion. Notice the Serbian version at the bottom in which the daughter de-escalates the emotionality while the other versions conclude with the daughter in tears, with the Mother blaming the Guy.

9 That is: no change in his indifferent expression on his face, no change in propping up his head with his elbows planted on the table and his hands under his chin, and staring empty mostly at the Mother-in-Law, but sometimes at the daughter. Since the daughter de-escalates the emotionality of the conversation in this version on her own, there seems little need for him to offer consolation.

Tab. 4: Screen shots of the mother-daughter conversation initiated by the daughter’s mention of the book *My Mother, My Mirror*



Quebec: Beginning of Mother-Daughter dispute. Guy is mildly disengaged.



Quebec: End of Mother-Daughter dispute. Guy is engaged in physical consolation of Girl.



Ukraine: Guy is mildly interested in the conversation between Mother and Daughter, but looking at the Mother-in-Law.



Ukraine: Guy exhibits no change in facial expression nor body position and offers no consolation to Girl.



Poland: Girl initiates conversation while Guy is absent.



Poland: Guy returns to table to find Girl in tears and attempts consolation without knowledge of conversation.



Serbia: Guy is completely disengaged in conversation and appears to be day-dreaming.



Serbia: Girl de-escalates emotionality. Guy continues to be disengaged.

consolation and retains a detached but astonished facial expression while looking at the Mother-in-Law during her accusation.

2.3 The Monolog and Music Cues

The diegetic music underscores the Girl's performance in the face of her Mother's insensitivity during a part of the segment I am calling the monolog. The monolog is an uninterrupted and extended speech segment of the Girl describing the trauma she has experienced during her adolescence but also relativizing the role of the Mother by referring to the 'difficulties' faced by the Mother and the fact that both seem to be 'stubborn'. The monolog is initiated after the Mother states that she is not familiar with the book *My Mother, My Mirror* and has not read it. The sad music cue sets in when the Mother replies that she cannot recall the traumatic experiences of the daughter. The music therefore signals to the audience what emotions are appropriate in the scene and with whom the audience should sympathize and empathize during the Girl's remarks. The Girl's initial emotional state of sadness stands in stark contrast to the Girl's explicit and insistent denials that her Mother's insensitivity affects her.¹⁰

The music cue therefore creates a tension between the increasing emotional content of the monolog on the one hand, the lack of appropriate response of empathy by the Mother, and the Girl's ultimately unsuccessful attempt at de-escalation, at least in most versions. The music cues prepare and accompany the escalating emotionality in the Girl's voice and brace the audience for her ultimate breakdown into crying at the end of the segment.¹¹ There are six parts to the music:

1. a slow, two-note introduction and upward progression covering an octave or 5th;
2. a single note downward progression for a measure or two;
3. a downward chord progression of 2-3 measures;
4. a one-measure musical pivot point which interrupts the downward progression;
5. a continuation of the previous downward chord progression (as coda);
6. the finale consisting of a final chord.

The monolog indicates which topics and behaviors in the discussion upset and distress the Girl. They are rooted in the insensitivity and lack of the Mother's

10 The Russian version is an exception to this in that the Girl states explicitly that she is angry at the Mother for causing the Girl's trauma.

11 The absence of this music cue in the Slovenian version means that the actress must produce the emotionality without the help of the music. This lack of this subtle but crucial mode of communication is especially glaring when compared with the other versions.

empathy for the daughter's traumatic experiences while growing up. The dialog demonstrates two problems in the relationship with the Mother: not only does the Guy have a problem with his Mother-in-Law, but the Girl also has a problem with her Mother, who in most of the versions denies any memory or knowledge of wrongdoing. The two most prominent traumatic experiences that the Girl mentions are the Mother washing her teddy bear in the washing machine with bleach and the Mother's prohibition for the daughter to bathe with the father.

While a quantitatively significant part of the dialog consists of the daughter's questioning the Mother about what the daughter calls past traumatic events, along with the monolog, a qualitatively significant (but quantitatively insignificant) portion of the dialog comprises the response of the Mother to the daughter's questions. In most versions, the Mother denies with monosyllabic responses awareness, or being the cause, of the daughter's traumatic experiences. In most versions, the Mother replies with a simple 'no' to the question if her Mother remembers washing the teddy bear with bleach. In the Polish version, the Mother tries to excuse her actions as a 'careless mistake that can happen' to anyone, but in most of the other versions the Mother does not acknowledge her responsibility. Contrary to almost all the other versions, the daughter in the Serbian version refuses to accept the Mother's denial and keeps insisting that the Mother remember the event, which the Mother grudgingly acknowledges after her daughter's repeated insistence.

The negative response of the Mother increases the insecurity of the daughter to speak with her Mother about intimate issues, manifested in some versions with a change in voice of the daughter through hesitations, stuttering, repetitions, a reduction in the rate of speech, and a switch to a lower (Serbian) or higher (Russian, Ukrainian) pitch. The music cue sets in in most cases soon after the first denials by the Mother, and accompanies the continued monolog of the daughter until the daughter breaks down crying. The crescendo of music signals not only the daughter's increasing emotional agitation at the Mother's lack of empathy with and sensitivity to the daughter's plight, but also the guilty conscience of the Mother at the daughter's insistence of the Mother's responsibility for the daughter's trauma. Looking at the beginning and end points of both the dialog and the sad music cues in Table 5, one notices considerable variation in both the duration of the dialog in the different versions, as well as the duration of the corresponding music cue, with the Greek version containing the longest dialog and the most extended sad music cue of all the versions with 01:35 (mm:ss) in the segment lasting 02:09, in which the dialog encompasses almost the entire segment at 02:06. The shortest segment duration is the Polish at 01:08, while the dialog in that segment is 00:53 long (i.e., 78% of the segment) and the sad music portion 00:51 in length, or 75% of the segment. The shortest extent of sad music can be found in the Canadian English version with only 00:35, or 42% of the 01:23 dialog portion of the segment.

Tab. 5: This shows the length of the sad music segments in the Mother-daughter conversation in Part 2 of the Mother-in-Law scenes, with the beginning and end points, the percentages of music in the total shot, and the length of the dialog in which the music was embedded (Time details: mm:ss)

	Total length of segment	Start of Sad Music Segment	End of Sad Music Segment	Total Length of Sad Music in Segment	Length of Sad Music in Segment in Percent	Beginning of dialog (daughter)	End of Dialog (daughter)	Length of Dialog Segment	Length of Dialog in Segment in Percent	Proportion of Music to Dialog in Percent
Quebec	01:30	00:38	01:25	00:47	52.2%	00:01	01:24	01:23	92.2%	56.63%
Canada-ENG	01:23	00:42	01:17	00:35	42.1%	00:03	01:16	01:13	87.95%	47.95%
Israel	01:36	00:43	01:28	00:45	46.88%	00:01	01:29	01:28	91.67%	51.14%
Turkish	02:07	00:47	01:52	01:05	51.18%	00:02	01:55	01:53	88.98%	57.52%
Greek	02:09	00:27	02:05	01:38	75.97%	00:01	02:07	02:06	97.67%	77.78%
Russian	01:54	00:43	01:37	00:54	47.37%	00:08	01:44	01:36	84.21%	56.25%
Ukrainian	01:59	00:48	01:41	00:53	44.54%	00:01	01:43	01:42	85.71%	51.96%
Latvian	01:41	00:41	01:35	00:54	51.92%	00:10	01:35	01:25	81.73%	63.53%
Polish	01:08	00:07	00:58	00:51	75.00%	00:02	00:55	00:53	77.94%	96.23%
Spanish	01:42	00:44	01:38	00:54	52.94%	00:01	01:36	01:35	93.14%	56.84%
Serbian	01:48	00:47	01:39	00:52	48.15%	00:03	01:41	01:38	90.74%	53.06%

The lengthy Greek version avoids the brief and superficial mention of the teddy bear and the prohibition of bathing with the father. Instead, the daughter opts for a much more detailed and extended listing of ‘oppressions’ perpetrated by her unaware Mother, but only after emphasizing that she is not criticizing the Mother, since the Mother raised her to be a ‘good Christian’ and a Greek patriot, remarks culminating in reciting a poem in unison that the Mother taught her when she was younger. However, the Greek monolog contains the most detailed and extensive list of all versions of how the Mother oppressed the Girl. In addition, the Greek version’s longer duration is the product of several snide interjections by the Guy wondering about his partner’s recalcitrance. Furthermore, the Greek version features a series of self-reflections by the Girl in which she states how difficult it is to discuss the topic with her Mother. The Greek version diverges in important ways from the other versions in that the Mother not only explicitly sympathizes with her daughter, but she apologizes for having unwittingly oppressed her daughter in the past. Another divergence from other versions is the Guy’s lack of support for his partner’s distress, so instead of consoling her in his arms at the end of the conversation, he tells the Mother-in-Law that she should continue oppressing the daughter since the daughter ‘lives in her own world’ and deserves the oppression. The Guy in the Greek version is the most active of all versions not just due to the multiple snide interjections against the Girl, but also because he opposes both the Girl and her Mother in the conversation.

The music cue in this Greek segment is extended to accompany most of the extended dialog, and especially the monolog of the Girl. In order to do this, the entire music cue is repeated for a second time, while in other versions only a coda of the third part, the downward chord progression, is inserted after the pivot point. In the short Canadian English version of this segment, conversely, the music cue dispenses with the coda of the third downward progression altogether, also dropping all mention of bathing with the father. The dialog diverges in another way as well in the Canadian English version. Instead of either implicitly or explicitly blaming the Mother for the daughter’s oppression as in most of the other versions, the daughter in the Canadian English version apologizes for not being around to support the Mother in her ‘very difficult’ life. The daughter perceived her Mother’s vacillation between wanting to be either her Mother or her friend, or both at the same time, as a cause of the daughter’s confusion. This statement was also included in other versions of the series. Thus, in most versions, the daughter blames *her Mother* for the confusion, while in the English Canadian version the daughter blames *herself* for being confused.

The increasing anguish of the daughter throughout the conversation is evinced in a variety of different—multimodal—ways in the various versions under scrutiny in this study. In the Russian version, for instance, the daughter starts to hyperventilate

during her monolog while nervously jiggling a spoon back and forth in her right hand during the tensest moments. In most versions, the most frequent physical movements of the Girl are her use of hand gestures, opening and closing the palms in front of her torso for emphasis, while the elbows and forearms are usually resting on the table. In the majority of versions, the Girl's body is mostly rigid and immobile with the exception of the forearms and hands, while only the upper body and torso are visible to the audience.

In the unique situation of the Turkish version, the Mother leaves the dining room altogether to go to the kitchen when the conversation begins to become uncomfortable, and the Girl has to turn her head to her left towards the kitchen, lean over the table to look in that direction, and yell in a loud—and breaking—voice to make her point. The Turkish version also modifies the causes of the daughter's trauma in mentioning that she was not allowed to sleep with the Mother (*not* take a bath with the father) when much younger. In the Ukrainian version, the Mother ends up crying at the end of her daughter's monolog in addition to the daughter, but still accuses the Guy of causing the hardship and distress. In the Polish version, the entire conversation between the Mother and her daughter takes place in the absence of the Guy, who went to the bathroom and only returns after the Girl has broken down crying, rendering the accusation of the Mother against the Guy—that he is to blame for her anguish—patently absurd. The Polish version is distinguished from the others in that the Girl mentions not the disfigurement of her teddy bear and the prohibition of taking a bath with the father, but instead that her father got first portions of porridge at breakfast and was the first to read the Father Christmas poem during the Christmas holidays.

These examples illustrate that subtle and not-so-subtle differences in the same scene and shots may lead to different kinds of interpretive conclusions based on the descriptive cataloging of the similarities and differences in the various versions within the framework of a micro-level multimodal method combined with a macro- and meso-level narratological approach. The behaviors of the Girl, her Mother, and the Guy reflect culturally specific discourses of gender identities, family relations, and ultimately, the meaning of families in the reproduction of the nation in each of the different versions.

The functionalist sociological literature on family relationships claims that the family is the smallest unit of the nation in which the gendered values, customs, and behaviors are transmitted to the younger generation through socialization (Dempsey & Lindsay, 2014, 51). In this view, the masculine is oriented toward the public sphere and economic sustenance of the family, assuming an 'instrumental' and active role. The feminine, on the other hand, is oriented toward the private realm and unpaid labor, and adheres to an 'expressive' and passive role. While the female protagonist in the various versions of *Un Gars*, *Une Fille* performs the

expressive and emotional gender role by means of her crying with the support of the sad music cue, it could be argued that in most versions the female character is the active person in the confrontation with her Mother. The male character, on the other hand, exhibits an uncharacteristic passivity towards the entire conversation he is witnessing and only acts when his partner is in her most distressed state.

This can also be seen in the traumas detailed in the criticism of the Mother by the daughter. Both the Mother and daughter help construct the culturally and politically specific discursive context of competing narratives of national identity as they are realized in the description of the daughter as well as the response of the Mother.

3 Concluding Thoughts: Multimodality, Interculturality, and Family Conflict Management Strategies

In an essay on intercultural screenwriting, Patrick Cattrysse (2017) suggests applying Gert Hofstede's dimensions of cultural variability to adaptations of television formats and films to create greater identification with characters and cultural proximity to new audiences in the adapted versions. These dimensions of cultural variability, which are the mainstays of empirical positivist orientations of intercultural communication research, are binary pairs which supposedly categorize national cultures and assist in comparing cultures of different countries and regions: high and low power distance, individualism and collectivism, masculinity and femininity, high context and low context styles of communication, and others. Critics (Holliday, 2013; Martin & Nakayama, 2012; Sorrells, 2016) have noted the essentializing and reifying nature of these categories in denoting whole cultures, but Cattrysse (2017, 14) proposes to deploy them as hermeneutic tools of intercultural analysis and screenwriting, a path I have pursued in previous studies (Larkey, 2009) of various versions of the TV series *The Office* (BBC 2001–2003; Gervais & Merchant, 2004), in which I have examined the countervailing use of camera position, camera movements, and dialog in a particular scene to parody gender and power distance violations between the main male protagonist and his female superior. Er, in another co-authored article (Larkey et al., 2016), examined how a humorous music cue helped mitigate the violation of power distance conventions among two characters in the Turkish version of the US TV series *Monk*.

Cattrysse (2017) scrutinizes the binary pair of universalism and particularism, which is the most relevant for our purposes and suitable for closer study.

According to Cattrysse, universalist and particularist attitudes are manifested in people being “rule-oriented” or “person-oriented”: “Whereas the universalist stipulates that society must be ordered by laws that apply equally to everyone always, the particularist will say that interpersonal relationships prevail” (Cattrysse, 2017, 12). I would argue that underlying the Mother-in-Law scene are the Mother’s universalist expectations about the conventional life-course progression of institutionalized marriage toward the couple on the one hand, and her particularist critique of the behavior of the male protagonist on the other, who is made personally responsible for the violation of those conventions within the framework of the family in question in this chapter. The Mother’s expectations correspond to older generations’ expectations of a conventional life-course management of the couple: courtship-marriage, children/child-rearing, male breadwinner, etc. Wilcox et al., who investigate “America’s retreat from marriage” along with the resulting “growing class divide in marriage” (Wilcox et al., 2015), also identify factors they believe influence the shift away from the conventional institution of heterosexual marriage:

[s]hifts in attitudes, aspirations, and norms, coupled with declining participation in secular and religious institutions, have undercut the social pressure to marry, to have children within marriage, and to stay married. (Wilcox et al., 2015, 113)

Wilcox and his fellow co-authors present a ‘deficit’ attitude toward the demise of the institution of marriage in US society¹² (as well as others) by asserting the increasing prominence of a “less family-oriented, more individualistic approach to relationships, marriage, and family life” (Wilcox et al., 2015, 112) since the 1960s. This development is characterized by the transition from a “cornerstone” to a “capstone” marriage model proposed by (Cherlin, 2010, 114), whereby “[m]en and women become less likely to see marriage as a foundation for adulthood, as the exclusive venue for sexual intimacy and parenthood”. This is also demonstrated in the interactions of the Guy and the Girl in *Un Gars, Une Fille* and is at the heart of the conflicts between the Mother and the Guy, as well as the Mother and the daughter in almost all versions of the series. The characters seem to uphold the conclusion about the more modern forms of marriage that have become:

¹² This is manifested in their view “that adults are less likely to thrive emotionally, physically, and economically outside of marriage, and because children who grow up outside of an intact, two-parent married family are more likely to suffer from psychological and social problems, and less likely to acquire the education and life experiences they need to realize the American dream of stable work and comparatively high income” (Wilcox et al., 2015, 112).

an opportunity for men and women to consecrate their arrival as successful adults, to signal that they were now confident they could achieve a fulfilling romantic relationship built on a secure, middle-class lifestyle (Wilcox et al., 2015, 114).

The male protagonist adheres to this different set of universalist assumptions based on the de-institutionalization of marriage, the Girl's ambivalent attempts at autonomy, and the lack of validated male gender life-course scripts, which necessitates continuous negotiations between the couple. The Guy is only able to offer a particularist response to the Mother's universalist critique. Since there is no historical script for the more enlightened male character who nonetheless still comprises a pillar of patriarchy, there is a need for negotiation, conflict management, and experimentation. Even though the couple is unmarried, their relationship closely resembles that of a married household in that they cohabit and have been committed to each other for many years (cf. Allen & Walter, 2000, 13).

As the 'third party negotiator', the Mother character in the various versions displays a wide variety of attitudes toward accepting responsibility for the traumas imposed on the daughter. The attitude prevalent in most of the versions is the denial of responsibility for the traumas of the daughter (Israeli, Latvian, Russian, Slovenian, Spanish), while the Mother in the Turkish version avoids the conflict altogether by escaping the conversation into the kitchen. The Serbian Mother only grudgingly admits to causing the trauma, while the daughter de-escalates the emotionality. The daughter in the English-Canadian version assumes responsibility for her own 'confusion' and absolves the Mother of any responsibility whatsoever, while the Polish Mother confesses to responsibility but downplays its significance. Only the Greek Mother confesses to her unwitting responsibility for the traumas after the daughter's extensive listing of them. Regardless of the question of culpability for the daughter's traumas, the sad music cues and the transformation of the daughter's speech from talking in a normal voice to breaking down and crying uphold the morally justified position of the daughter and create empathy with her regardless of how the Mother positions herself with regard to causing the daughter's trauma.

Further research would be needed to interconnect the narratives in each of the different versions to local discourses on evolving gender and family relations in each of the countries investigated here. Further research would show how additional shots in the scene reinforce the Mother-in-Law's animosity towards the Guy in the Turkish version by portraying him as a loser with regard to conventional family life-course scripts. This is in keeping with the more conservative outlook of Turkish television series intent on maintaining 'Turkish' family structures and identities in the face of rapid social and economic change. Further research would also illustrate more support for the male protagonist, who, despite being a pillar of pa-

triarchy, is nonetheless willing to negotiate greater autonomy and self-realization for his female partner.

A multimodal approach to comparatively analyzing different versions of global scripted television formats, expanded to include software-compiled qualitative and quantitative data, along with supplemental data on narrative structure, content, and sequencing, might offer a path to resolving Kress' assertion on the significance of cross-cultural and intercultural differences in a semiotic object. He states: "The more pronounced the cultural differences, the greater are the differences in the resources of representation and in the practices of their use" (Kress, 2010, 10). While we are not yet ready to make pronouncements about the correlation between the degree of cultural differences and the quantity of cultural resources deployed in representations, our methodology does work in the same general direction by collecting evidence and data on what that might mean.

This touches upon notions of cultural (Moran, 2009; Aveyard & Moran, 2015; Straubhaar, 2007; Chalaby, 2015) and aesthetic proximity (van Keulen, 2016) at the heart of research on global television formats. Kress's hypothesis implies that culturally specific configurations of televisual aesthetic techniques and practices are manifested in a specific configuration in the ensemble of communication modes specific to a particular society. There is currently a gap between what Kress is envisioning on the one hand, and the sophistication of multimodal and narrative research and knowledge on the other such that many more cross-cultural multimodal comparative studies of different television formats must be conducted in order to examine how concepts such as 'cultural differences' are configured in the 'resources of representation', and what the 'practices of their use' look like. For instance, it is not clear how a scholar would quantify or provide qualitative data on 'cultural differences'. The methodology suggested in this paper can be considered some first steps towards quantifying the difficult notion of cultural differences.

A further challenge is the nature of multimodal knowledge and its visual models. Currently, each mode of communication tends to have its own model to represent the knowledge of that specific mode. Music notation, dialog transcriptions, video sequences, editing, conversation analysis, and even Labanotation for physical movement are specific to each single mode of communication and are merely linked together by simultaneous use or depiction in models of audiovisual media such as film and TV. However, what does multimodal knowledge look like when many modes are combined and correlated with each other? At present, there needs to be a way to model multimodal knowledge apart from attempts to bring together disparate (audio-)visual representations. This paper is also a product of this dilemma in that I have chosen tables and bar graphs created with Microsoft Excel to make my points. It would seem that multimodal research also needs to develop

its own, qualitatively more appropriate visual models for presenting multimodal knowledge.

Acknowledgements

I would like to thank the following colleagues, students, and friends for their help in translating the dialogs of the various versions: Yael Levy (Hebrew), Boyan Durmisević (Serbian), Drorit Zilberberg (Ukrainian and Russian), Petra Opitz (Russian), Dora Maniou (Greek), Angelo Sollano (Polish), Elisabeth Arévalo (Spanish), Ibrahim Er (Turkish), and Landry Digeon (French and Quebecois). I am also grateful to Jerome Bourdon and Sandrine Boudara from Tel Aviv University, and Paul Frosh and Tobias Ebbrecht-Hartmann at Hebrew University for the opportunity to interact with them, their students, and colleagues in developing my ideas. My extreme gratitude also goes out to Stéfany Boisvert, who was instrumental in coordinating communication with Avanti Groupe in Montreal. In addition, I would like to thank my university, University of Maryland, Baltimore County (UMBC) for its support of innovative interdisciplinary research that traverses the humanities, social sciences, and other fields of science and engineering. Finally, and last but not least, I would also like to express my thanks to my wife, Uta Larkey, for her critical comments and her suggestions for revisions and corrections.

Bibliography

- Allen, Katherine R. & Alexis J. Walter. 2000. Constructing Gender in Families. In Robert Milardo & Steve Duck (eds.), *Families as Relationships*, 1–18. West Sussex: Wiley and Sons.
- Aveyard, Karina & Albert Moran. 2015. Television Format Traffic – Public Service Style. *Continuum: Journal of Media and Cultural Studies* 29(5). 684–693.
- Bateman, John A. & Karl-Heinrich Schmidt. 2012. *Multimodal Film Analysis: How Films Mean*. London: Routledge.
- Bateman, John A., Janina Wildfeuer & Tuomo Hiippala. 2017. *Multimodality – Foundations, Research and Analysis. A Problem-Oriented Introduction*. Berlin: Mouton de Gruyter.
- Butler, Jeremy. 2012. *Television. Critical Methods and Applications. Fourth Edition*. New York: Routledge.
- Cattrysse, Patrick. 2017. Cultural Dimensions and an Intercultural Study of Screenwriting. In Carmen Sofia Brenes, Patrick Cattrysse & Margaret McVeigh (eds.), *Transcultural Screenwriting. Telling Stories for a Global World*, 8–27. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Chalaby, Jean. 2015. *The Format Age. Television's Entertainment Revolution*. Cambridge: Polity.
- Cherlin, Andrew J. 2010. *Marriage-Go-Round: The State of Marriage and the Family in America Today*. New York: Vintage Books.

- Dempsey, Deborah & Jo Lindsay. 2014. *Families, Relationships and Intimate Life. Second Edition.* South Melbourne: Oxford University Press.
- Friday, Nancy. 1977. *My Mother, My Self. The Daughter's Search for Identity.* New York: Delacorte Press.
- Fuerstein, Laura. 2008. *My Mother, My Mirror: Recognizing and Making the Most of Inherited Self-Images.* Oakland: New Harbinger Publications.
- Gervais, Ricky & Stephen Merchant. 2004. *The Office. The Complete Series One and Two and the Special.* Burbank: Distributed by Warner Home Video.
- Holliday, Adrian. 2013. *Understanding Intercultural Communication: Negotiating a Grammar of Culture.* Milton Park: Routledge.
- Jewitt, Carey. 2009. An Introduction to Multimodality. In Carey Jewitt (ed.), *The Routledge Handbook of Multimodality*, 14–29. London, New York: Routledge.
- Jewitt, Carey, Jeff Bezemer & Kay O'Halloran. 2016. *Introducing Multimodality.* London: Routledge.
- van Keulen, Jolien. 2016. Aesthetic Proximity. The Role of Stylistic Programme Elements in Format Localisation. *View* 5(9). 93–104.
- Klein, Renate C. A. & Robert M. Milardo. 2000. The Social Context of Couple Conflict: Support and Criticism from Informal Third Parties. *Journal of Social and Personal Relationships* 17(4-5). 618–637. <https://doi.org/10.1177/026540750017400>.
- Kress, Gunther. 2010. *Multimodality: A Social Semiotic Approach to Contemporary Communication.* London: Routledge.
- Larkey, Edward. 2009. Transcultural Localization Strategies of Global TV Formats: The Office and Stromberg. In Albert Moran (ed.), *TV Formats Worldwide. Localizing Global Programs*, 187–202. Bristol: Intellect.
- Larkey, Edward. 2018. Narratological Approaches to Multimodal Cross-Cultural Comparisons of Global TV Formats. *VIEW. Journal of European Television History and Culture* 7(14). 38–58. <http://doi.org/10.18146/2213-0969.2018.jethc152>.
- Larkey, Edward, Landry Digeon & Ibrahim Er. 2016. Measuring Transnationalism: Comparing TV Formats using Digital Tools. *View* 5(9). 72–92.
- Lim Fei, Victor, Kay L. O'Halloran, Sabine Tan & Marissa K.L. E. 2015. Teaching Visual Texts with the Multimodal Analysis Software. *Education Tech Research Development* 63. 915–935.
- Martin, Judith N. & Thomas K. Nakayama. 2012. *Intercultural Communication in Contexts. Sixth Edition.* New York: McGraw-Hill.
- Mittell, Jason. 2004. *Genre and Television.* New York and London: Routledge.
- Moran, Albert. 2009. *New Flows in Global TV.* Bristol: Intellect.
- Sorrells, Kathryn. 2016. *Intercultural Communication. Globalization and Social Justice. Second Edition.* Thousand Oaks: Sage Publications.
- Straubhaar, Joseph D. 2007. *World Television. From Global to Local.* Thousand Oaks: Sage Publications.
- de Vaus, David. 2011. Social Trends and Their Impact on Couple and Family Relationships. In Patricia Noller & Gery C. Karantzas (eds.), *The Wiley-Blackwell Handbook of Couples and Family Relationships*, 25–35. New York: Wiley & Sons.
- Wilcox, W. Bradford, Nicholas H. Wolfinger & Charles E. Stokes. 2015. One Nation, Divided: Culture, Civic Institutions, and the Marriage Divide. *The Future of Children* 25(2). 111–127.

Wendy Nielsen, Pauline Jones, Helen Georgiou, Annette Turney,
and Mary Macken-Horarik

Learning Science through Generating Multimodal Digital Explanations: Contributions to Multimodality in Educational Practice

Abstract: Communicational, technological, and cultural shifts within science have substantially impacted how scientific understanding is developed and communicated. Accordingly, science learners need to experience teaching and learning approaches that utilize a wide range of multimodal forms and representations to develop content knowledge and communications skills. ‘Blended media’, for example, provide opportunity for science learners to develop a range of important competencies while working with multimodal representational forms. Research in multimodality has yet to move into the complex learning space of student-generated digital explanations and we contribute to research in the field through this interdisciplinary work that examines the nature of knowledge building as university science students create multimodal digital products. In this research, we have gathered samples of digital products (n=60) created by science learners, interviewed creators and instructors, and conducted case studies to capture the creation process. We have also developed a series of analysis tools to map the field and frame a focused look into how learners use semiotic resources to create the digital media product and examine the effect of the process on the quality of students’ learning. This chapter introduces the project while bridging several research areas as we illustrate in our analyses. We draw from theoretical perspectives in science education, cognitive psychology, educational semiotics, and the sociology of education to understand meaning making processes. The chapter introduces our approaches to this interdisciplinary research and develops emerging theoretical and analytical insights as we theorize the learning processes involved in student-generated multimodal digital text creation.

Keywords: multimodality, learning, education, blended media, digital explanations

<https://doi.org/10.1515/9783110608694-010>

1 Introduction

Recent shifts in communications, technology, and culture have substantially impacted how scientific understanding is developed and communicated. The ubiquity of Web 2.0 applications, growth of digital publication and dissemination (Rigutto, 2017; Warden, 2010), and integration of participatory approaches to scientific communication like blogs and vlogs (Estrada & Davis, 2015) have expanded the communicative channels used in science. These shifts have contributed to the increasing use of interactive and dynamic elements alongside traditional visual, diagrammatic, mathematical, and linguistic representations both within the field of science and in school and university science subjects, where multimodal communication is increasingly acknowledged as an important educational outcome.

This project examines ‘digital explanation’ (Hoban et al., 2016a), an assessment task requiring tertiary science learners to generate stand-alone presentation-style digital media artifacts to explain science content to others (Hoban et al., 2016b; Nielsen et al., 2018). These others might be children if the creators are preservice teachers or non-expert peers when the creators are biology or pharmacology students. Such tasks provide opportunities for science learners to develop a range of important competencies, for example, to demonstrate understanding of biological or chemical processes, to integrate complex material in texts of different kinds, and to represent learning in a range of multimedia genres.

However, research in multimodality has yet to move into this learning space and we seek to contribute to the field through our interdisciplinary work. In this chapter, we consider an interdisciplinary view on multimodality, combining research approaches from sociology, social semiotics, and science education to widen the lens through which multimodality can be theorized. An emerging theory of multimodality useful to education would demonstrate relevance to specialist domains of inquiry, make resources for meaning more visible and tools for producing multimodal texts accessible. Perhaps most challenging, such a theory would shed light on the ‘how’ of learning—the progressive expansion of repertoires for meaning.

With this chapter, we firstly review theoretical perspectives on science learning, social semiotics, and multimodalities that drive the research. Systemic-functional linguistics (henceforth: SFL; Halliday, 1994; Halliday & Hasan, 1985), with its focus on language in use, has been increasingly extended to analyze non-verbal meanings such as images (Kress & van Leeuwen, 2006 [1996]), sound (van Leeuwen, 1999), and other communicative modes. Now called systemic-functional semiotics (SFS; O’Halloran, 2005), this analytical toolkit proves helpful to describe how semiotic resources are deployed in an exemplar text and in theorizing multimodality.

The toolkit is used here to explore the communicative nature of student-generated science explanations as digital artifacts in the current chapter.

To consider how scientific knowledge is built within the artifacts we use *Legitimation Code Theory* (henceforth: LCT; Maton, 2014). The chapter concludes with a discussion of an emerging theory of multimodality for this interdisciplinary work suggesting that what is needed is a shared language (a ‘metalanguage’) for talking about multimodal forms in discipline-specific contexts in ways that account for the range of semiotic resources in play. Thus, we theorize a toolkit ‘good enough’ (Macken-Horarik et al., 2011) for knowledge building in science fields.

2 Background

A broad goal of the cultural domain of the sciences is to develop explanations for natural phenomena. As such, a discernible focus on the ‘explanation’ as a form of literacy in science education is apparent because learners work with representations to make meaning while generating an explanation (see, e.g., Prain & Tytler, 2012; Tang et al., 2014; Yeo & Gilbert, 2017). Science is a representation-rich field. Explanations in science typically involve the use of images and labelled figures along with extended sections of writing. In order to explain something, one must have necessarily undergone a process of distilling and integrating complex material. Thus, explanation is a key genre for building and demonstrating knowledge.

Enculturation into the practices of science necessarily involves working with standard forms of representation, such as diagrams, charts, symbols, formulas, and simulations, all of which represent disciplinary knowledge in specific ways. Learners also generate many of these representations in lab reports and presentations as part of this enculturation. In studying a contemporary example of a student-generated representational form, the interdisciplinary research we report here brings together science educators and semioticians to explore the research context where students create multimodal digital artifacts that explain science content to a specified audience.

Approaching these multimodal artifacts from the perspective of SFS (Bate-man, 2008; Halliday, 1978, 1994; O’Halloran, 2005) allows us to investigate how learners deploy semiotic resources and what this might represent about their understanding of the content. Consistent with O’Halloran (2008), we define semiotic resources as meaning-making systems that can be combined in multimodal ensembles for a communicative purpose. Semiotic resources offer ‘systems of choices’ (van Leeuwen, 2005) and include image, language, and music, whose meanings are

realized as learners employ them in text creation. Digital explanation as text typically includes the semiotic resources of image (moving or still), written language, spoken language, and music (Hoban et al., 2016a). Exploring how these resources work together within and across texts is vital if we are to deepen knowledge about the deployment of multimodality in tasks like digital explanation.

3 Research Context

In many university science learning contexts, digital explanations have replaced ‘stand-and-deliver’ presentations or summaries of written reports as assessment tasks (Hoban et al., 2016b). These stand-alone products integrate a range of semiotic resources, both self-generated and ready-made, and run for three to five minutes. Depending on the instructor’s purposes, the task specifications guide students to create a multimodal digital product that explains science content to a non-specialist audience.

Our data set includes sample texts produced as assessment tasks in a range of science discipline and science teacher education contexts. Participants voluntarily contributed their texts to the research project. Across the data set, the assessment weighting of such tasks varies from 5 to 40% of the total subject assessment. Characterizing these texts is challenging because not all are instantly identifiable as explanations and they are highly variable in terms of which resources are chosen. Students generally also make their own decisions about choice of software program to use, any of which has particularities and affordances for producing a digital explanation. Furthermore, science texts are commonly multimodal (Lemke, 1999) and the multiplication of resources made possible by digital technology to create them means that meaning potential is expanded. Descriptions help to stretch our understandings of how texts ‘grow’ through combining and nesting smaller elements (Martin & Rose, 2012), and genre theory supports this analysis.

The data set includes digital artifacts that explain environmental issues, pharmacokinetic or biological processes, simple science content, nutritional requirements in a particular cultural context, or other scientific information. For the purposes of this chapter, we focus on one example from pharmacology, entitled ‘Malaria’, produced by a student in response to a task requiring a recontextualization of a 2,500-word technical literature review to answer the question ‘Is ferroquine an ingenious anti-malarial?’. The product is a brief digital presentation (5 min 28 secs) for his non-specialist peers that was assembled in web-based presentation software using a wide range of visual and audio resources. The key purpose of the assessment is to communicate an answer to a complex question

for a non-expert audience. The artifact received full marks, thus we consider it demonstrating what is valued in the task and offering a worthy exemplar of the challenge facing production of effective multimodal explanations.

4 Theoretical Perspectives

A driving force for our work is the social semiotic conception of language that continues to inform research into multimodality. We begin with the semiotic aspect of our interdisciplinary project and then move to the sociological aspect, focusing on recent work within a social realist theory of knowledge building—LCT (Maton, 2014). Systems have evolved (and continue to evolve) according to the meaning-making work that they do in social contexts (Halliday, 1999). Importantly for both analyst and educator, Halliday argued that there is a systematic relationship between the social environment and the functional organization of text (Halliday & Hasan, 1985).

This systematic relationship is captured in Field, Tenor, and Mode—variables evident in any situational context where meanings are exchanged. Field, according to Halliday (1999), refers to the on-going social activity, Tenor to the communicative roles of the participants, and Mode to the channel of communication (written or spoken). The patterns of meaning in a text (*its register*), in the Hallidayan model, are shaped by the configuration of the Field, Tenor, and Mode variables. Macken-Horarik & Adoniou (2007) subsequently extend Mode to include visual and multimodal meanings. The expansion of Mode to include visual, audio, and gestural resources is a more recent development, but entirely central to our endeavor as we indicate later in the chapter. Thus, the digital explanations comprising our data set can be described in terms of their register: the field of scientific knowledge represented, the relationship with the viewer established, and the modes of communication involved.

The register variables map onto three metafunctions, or bundles of semiotic choices: ideational, interpersonal, and textual. The ideational metafunction, in the realization of field-related meanings, serves to construe our experience of the world and make logical connections; the interpersonal enacts Tenor values—the social experiences and relationships relevant to the situation; and the textual metafunction organizes the message into a coherent text in response to the mode/s in play. These metafunctions are realized through networks of systems of choices. Multimodal theorists have extended this metafunctional perspective in considering other semiotic modes and resources (Kress & van Leeuwen, 2006 [1996]; Lemke, 1999; O'Halloran, 2005; O'Toole, 1994; van Leeuwen, 1999). The strength of

systemic-functional theory to multimodality lies in these metafunctional principles offering an integrating platform to consider the interaction of different semiotic resources to create meaning (O'Halloran, 2008).

In other developments of the functional model, Martin (1993) and Rothery (1996) argued that texts also differ according to the configurations of meanings resulting from their broad social purposes; that is, to argue, to discuss, or to explain in different discourse communities. Within a science discourse community, combinations of multimodal elements are common, making the texts and their purposes complex (Martin & Rose, 2012). Genres, in the Sydney School approach to genre theory which we draw on here, are defined as 'staged, goal-oriented processes': goal-oriented because a text unfolds to meet a particular purpose, staged because it usually takes more than one step to do so (Martin & Rose, 2012).

From an initial emphasis on identifying and describing elemental genres (that is, those with a single purpose), genre theory proponents have described how texts expand as their purposes become more complex in specialized contexts (Martin & Rose, 2012). Elemental genres can expand by repeating some stages. For example, a report about an endangered animal will often have a number of stages describing aspects such as diet, habitat, threats, etc. Other genres can also be embedded as stages, for example, an explanation of reproduction could be embedded within the report. Combinations of genres are common, as we will see in the sections ahead detailing the Malaria artifact, and often take the form of macrogenres; that is, a series of texts (instances of single or elemental genres) adjusted to fit together such as often found in textbooks and webpages. Texts that are expanded through such embedding and combining are frequently found in advanced disciplinary contexts (Szenes, 2017).

Context in systemic-functional theory refers to both the immediate context of situation or register and the broad context of culture (Halliday, 1999). Halliday models the relations between context and language, as depicted in Figure 1, arguing that

the context for an instance of language or text is an instance of the culture (situation). And the context for the system that lies beyond each text (language) is the system which lies behind each situation—namely the culture (Halliday, 1999, 275).

Any learner's task is thus to build up the systems available in language (as a whole) through countless instances of listening, speaking, reading, and writing texts. At the same time, the learner is learning the language and participating in situations that enable him or her to 'make sense' of the culture. In the case of scientific learning, this means coming to understand the knowledge practices of science and participate in specific situations and the instances of discourse

these generate. In the current study, learners draw on their repertoires of meaning-making resources together with their disciplinary understandings aligned with specialist communicative practices to construct a multimodal artifact that meets the requirements of the assessment task to explain science content to a non-expert.

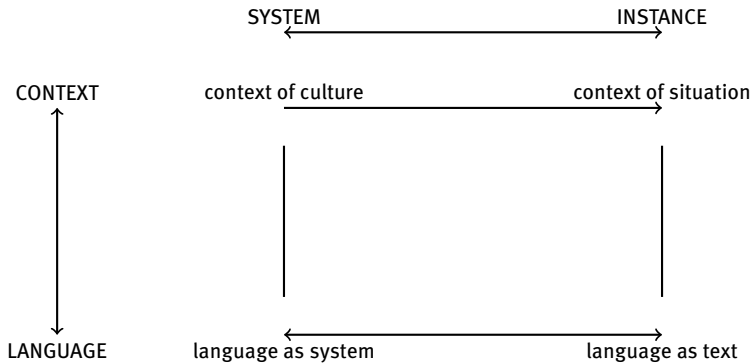


Fig. 1: Language and context, system, and instance (adapted from Halliday (1999)).

In order to develop a picture of how learners responded to the demands of the situational context of generating a digital explanation as an assessment task, we set to work mapping genre and register in a cross-section of the data corpus. Identifying the genre of an artifact provided insights into how the student perceives the broad purpose of the task, while describing the register helps to confirm the genre description. An instance of a genre is recognizable in the patterning of meanings as the text unfolds in order to fulfill its rhetorical purpose. For example, an explanation of the lifecycle of an endangered animal in an upper primary textbook usually foregrounds activity sequences involving a generic entity (Field), and features an expert-to-novice (but largely neutral) ‘tone’ (Tenor), and bundles of information are organized according to time, perhaps with an accompanying diagram (Mode). In contrast, an exposition arguing the need for protecting the habitat of the animal is likely to feature persuasive meanings, foregrounding Tenor with meanings related to Field and Mode less prominent. Thus, mapping genre and register across a digital explanation offers insight to how learners navigate the situated context of this multimodal landscape.

The semiotic categories allow us to sort examples from the larger data set in relation to text type (genre), approach to subject matter (Field), composer-audience relationship (Tenor), and communicative resources deployed (Mode). Because genre and register are considered abstractions above language, it is possible to pro-

duce a broad brush description of a number of the multimodal artifacts in the data set—the first step in the research project that informed the subsequent selection and analysis of a smaller subset of artifacts. With respect to genre, we consider each of the sample artifacts in terms of its broad social purpose (for example, mapping back to the task specifications to see if the student had recognized and recruited the relevant genre patterns). We also identify how the text unfolded as stages or waves over time (phasing). Shunting between verbal (Halliday, 1994; Martin & Rose, 2012) and visual meanings (Chan & Unsworth, 2011; Painter et al., 2013; Unsworth, 2006), and mindful of intersemiotic work, we undertake more granular analyses that adapt linguistically-oriented tools to provide useful descriptions of variations in the dimensions of register. In the context of the current research, each of these is elaborated below:

The term *Field* refers to the social activity of science learning. In a more delicate application to science learning, we consider such matters as whether entities and images were everyday/concrete or technical/abstract and how these were organized into taxonomies, and how ideas were sequenced and connected into activity and/or implication sequences through relations of cause, time, and logic.

The term *Tenor* refers to the communicative roles taken up by the student creator, who must read the task in terms of the needs and interests of the viewer of the artifact. Here, we considered aspects including assumptions about the viewer's previous experience and knowledge of the topic; whether meanings were intensified through heightened evaluative language or extreme images; the use of outside sources; the style of narration (narrator as scientist/teacher/peer); whether or not humor was used; and, whether devices such as images, metaphors, or analogies were used.

The term *Mode* refers to the communicative channel in use and students need to be able to organize the multiple visual and verbal resources into a coherent whole. Here we looked for elements such as whether some aspects are more prominent; types of signposting (questions to the audience, continuatives, labels, timescales, etc.); consistency in visual elements; use of transitions; and the degree of craftedness evident. One particularly relevant resource for integrating meanings in factual texts is periodicity, which in SFL theory describes how the text is organized into waves of information and announced at the text level by macro-Themes which are in turn bundled into small 'chunks' of meaning with hyper-Themes (Martin & Rose, 2012). For example, if an explanation begins with an overview of the phenomenon or concept to be explained, this functions as a prediction of what is to come and suggests that the student has a good idea of the scope of the explanation.

The resulting overview of genres represented in the corpus revealed that few artifacts were prototypical explanations. Several artifacts represented macrogenres that combined elemental (or singular) genres; for example, some artifacts from

primary science education comprised an explanation of an environmental issue followed by an exposition arguing the case for social action. Others represented elemental genres where stages sometimes comprised complete genres of their own; for example, one artifact from Pharmacology discussing the pros and cons of a particular drug therapy (a discussion genre) included an embedded explanation for how the drug works. Such differences were closely related to the nature of the assessment task; that is, whether the purpose was to explain a scientific concept, to persuade the viewer of the usefulness of a particular drug or to report on the dietary needs of a particular group of aged care clients.

With respect to register variation, the analysis allowed us to identify an instance of a genre (or ‘macrogenre’) with some confidence at this early stage. It also enabled us to describe similarities and differences among the artifacts in further detail. In particular, it allowed us to consider whether concepts were explained at an appropriate level of complexity (Field), the assumptions made about putative viewers and degrees of evaluation (Tenor), and how multiple semiotic resources were integrated in the final product (Mode). In other words, these tools enable us to understand something of how students construe the requirements of the assessment situation and how they recruit resources from their semiotic repertoires to respond to the task.

5 Working Interdisciplinarily

A digital explanation as an artifact of learning could be explored semiotically from the point of view of genre and register requirements. But its role in building knowledge in the domain of scientific study also needs to be understood. To this end, the project is also informed by *Legitimation Code Theory* (LCT; Maton, 2014), which is a social realist approach that extends the *Code Theory* by Bernstein (1977) to explore how knowledge practices work in different fields such as sociology, education, or linguistics.

Where systemic-functional theory considers semiotic (particularly linguistic) resources that characterize various contexts, such as school science, LCT considers the sociological issue of ‘uncommonsense’ knowledge and how this manifests in different practices (Martin & Maton, 2017). Focusing on the register variable of Field, LCT offers a way to conceptualize the organizing principles of the field (e.g., knowledge practices) across linguistic and visual resources. Thus LCT enables us to better understand how the student makers must negotiate science knowledge and practices to represent their learning in the situated multimodal context of creating a digital explanation.

In the current research, we explore the interplay between systemic-functional theories and knowledge-building in science to build a semiotic toolkit ‘good enough’ (Macken-Horarik et al., 2011) for the educational practices involved when students create a digital explanation. Knowledge ‘that counts’ in a cultural domain (for example, canonical knowledge in science and ways to communicate this knowledge that are recognized in the community) puts pressure on meaning-making choices (verbal, visual, and multimodal) and influences the choices in a particular situation (register) and text structure (genre). Texts produced in this situation will bear some resemblance to one another and carry evidence of student learning.

Halliday’s (1999) social semiotic model situates learning within a sub-cultural domain like physics or biology and resources instantiate this disciplinary specificity in that situation. Extending Figure 1, the interplay between systemics and knowledge-building is modeled in Figure 2. At the corners of the model are Halliday’s reference points for the relations between the system of resources and the instantiation as evidenced in particular artifacts. There is a space of some middle ground, e.g. register and text types, but we argue that these also need to be recognizable in other domains, too.

The questions in Figure 2 elaborate each of Halliday’s labels in terms of what the learner/creator has to do: process and produce text (instantiate a text type); construe the situation that lies behind the text (situation type); build up the linguistic system upon which the text draws (register); and construe the culture that gives meaning to language (cultural domain).

Inherent in the four questions in Figure 2 is a particularization of both systemics and disciplinary science, following Halliday (1999). Macken-Horarik et al. (2018) construed these questions as compass points, which help to build a picture of a range of disciplinary challenges in the current work. The concept of *relevance* deals with knowledge practices in the cultural domain of science and how multimodality is put to practical use in scientific study or preservice teacher education as learners design and produce texts. Developing these texts is part of disciplinary explorations and a learner’s *repertoire* makes it possible to explain increasingly complex phenomena. Producing a coherent text requires the learner to understand which semiotic *resources* are crucial to the explanation, including taxonomies, temporal and causal connectives, appositive visual models, etc. Genre and register features elucidate *rhetoric* in an effective macroexposition, including knowing how to produce the text for a particular situation type, in this case, a non-expert audience. These four elaborations frame a heuristic to examine how meaning is inherent in a text and further guides text construction in this complex situation. With a backdrop of Halliday’s (1999) reference points, upcoming sections illustrate aspects of our analytic processes through work with one exemplar artifact.

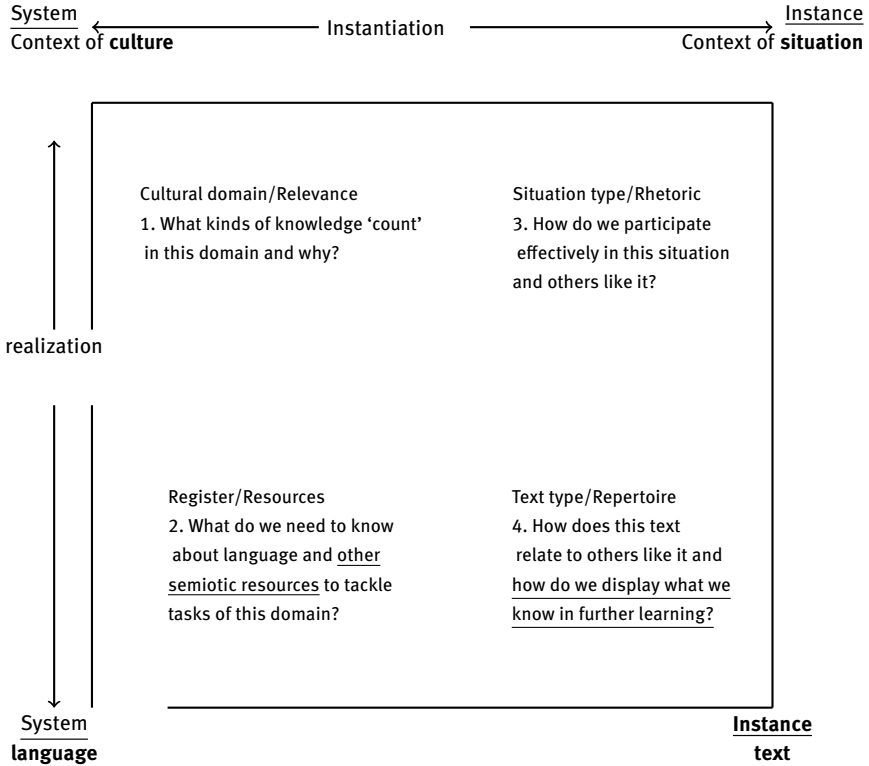


Fig. 2: The interplay between systemics and knowledge building.

6 The Rhetorical Demands of the Task

In this section, we illustrate the genre and register analysis with reference to the Malaria digital presentation as an exemplar artifact, demonstrating the first step in our analytical process.

Genre. The exemplar artifact in Figure 3 uses written and spoken language, audio, still image, and animation assembled via web-based presentation software, and played as a stand-alone artifact. Notably, the presentation opens each central segment in turn, each of which includes additional images, text and narrations. Although rated as highly successful, the exemplar, like many others, is not an

Tab. 1: Genre description of Malaria artifact

00:00-00:47	Ferroquine could be best weapon against malaria (Statement of Position)
00:47-03:19	New treatments needed for malaria (exposition 1) causes of malaria ([[causal explanation]]) — current treatments and their history ([[historical account]]) — need for safer, more effective drug treatments (Argument)
03:19-05:19	Case for ferroquine (exposition 2) How does ferroquine (+ quinine and chloroquine) work? ([[consequential explanation]]) — evidence for ferroquine’s effectiveness (Argument) — trial results (Argument)
05:19-05:28	More clinical trials needed as ferroquine could save many lives (Restatement of Thesis)

instance of an explanation genre. Rather it is a macro-exposition¹ comprising two expositions: one arguing the need for new anti-malarial treatments and the other arguing the case for ferroquine as a novel anti-malarial. Both of the arguments contain embedded explanations. The macrogenre has its own Statement of Position and Thesis Reiteration, which ‘bookend’ the two expositions. Such minor adaptations like this are necessary as “genres flexibly adapt themselves to co-textual, intermodal and contextual environments as needs arise” (Martin & Rose, 2012, 15–16).

The first exposition develops the field, explaining the causes of the disease of malaria through narrated embedded video and animations and describes developments in anti-malarial treatments before identifying problems with other treatments. The second exposition builds momentum, introducing and explaining how ferroquine works, emphasizing its anti-resistance qualities, using evidence from the literature to demonstrate its potential, and arguing for further trials. Embedded in these two expositions are explanations of how malaria occurs, how ferroquine works, and a historical account of anti-malarial treatments. The artifact is visually complex and the genre description in Table 1 shows these elemental explanations embedded as stages alongside other argument stages within the two expositions. Such embedding helped confirm the artifact as a macrogenre.

Register. A striking aspect of this text is how the creator draws on the affordances of the Prezi software to organize the semiotic resources into coherent bundles of information to propel the arguments forward. Macro-Themes (The Problem, The Cause, Existing Treatment, etc.) are made prominent through choices of layout,

¹ Following labelling by (Martin & Rose, 2012) for a text comprising several instances of reports (and other genres) orientating to an overall purpose of classifying Australian desert environments as a macro-report, we have labelled this artifact as an instance of a macro-exposition.

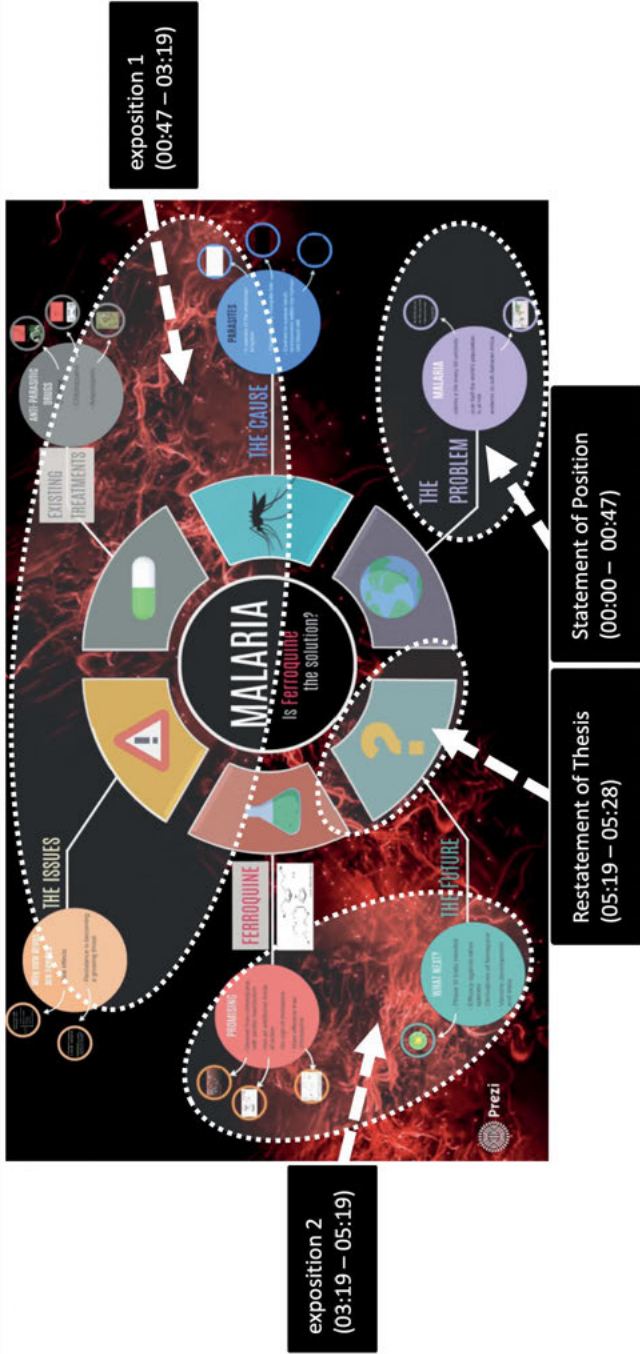


Fig. 3: Malaria: a macro-exposition, assembled in web-based presentation software.

color and labels synchronized with narration and the dynamic transitions enabled by the software (see Figure 4), suggesting the student is highly conscious of the field (perhaps through the pre-requisite task of the literature review). This high level understanding of the field is revealed in the creator's attention to periodicity. The macro-Themes are further elaborated through hyper-Themes.

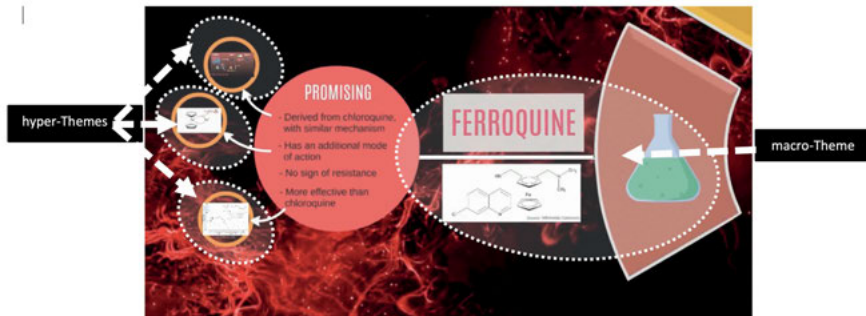


Fig. 4: Close-up showing visual instantiation of periodicity.

Structuring the viewer's attention is not left entirely to periodicity resources or the dynamic affordances of the particular software; the student draws on other resources related to *Mode* and *Tenor*. The narration features continuatives (in bold, in example) and softened commands (underlined) to indicate a shift in the content (e.g. 'But **first** let's take a look at the problem; **Now** let's look at the cause; **Now** let's talk about the issues with current treatments'). These features illustrate the creator's awareness of the need to guide the viewer through the complexities of the artifact, suggesting a heightened awareness of the *Tenor* demands of the task, that is, an explanation for a non-specialist audience about how ferroquine works.

While space prevents a detailed discussion of *Tenor*, our analysis revealed that there is a good deal of evaluation in the narrative as the student 'hooks' the viewer with a military metaphor at key stages of the macro-exposition, positioning malaria and ferroquine as warring elements. This overtly evaluative aspect of *Tenor* confirmed our description of the macro-exposition, as such choices are rarely found in more factual, scientific texts where explanations favor a more 'objective' *Tenor*. The narration introduces malaria as a major disease threat to the world in the Statement of Position ('A disease that has plagued humanity for millennia') and the sense of urgency is intensified in the opening stages of the first argument by graded lexis ('a severe infectious disease, immense human suffering and it kills on average half a million people around the world each year').

In contrast, ferroquine is presented positively (albeit measuredly) as a potential solution referenced briefly in the Statement of Position (‘ferroquine is the strongest candidate we’ve found and it may well be the solution’), appreciated positively in exposition 2 (‘a very promising candidate to reach clinical practice, superior against many different strains of *Falciparum malaria*’) and in the Reiteration of Thesis (‘ferroquine may be exactly what the world needs right now to fight malaria as it is a drug with a lot of potential to save lives’). These subjectively oriented evaluations are bolstered by evidence from the technical literature review (‘a summary of 22 studies’) and depicted graphically. Such persuasive elements contrast with the more ‘authoritative’ tone of the explanation genres embedded as stages in the two expositions (‘ferroquine, quinine and chloroquine all work by inhibiting haemozoin formation’).

The Field of the Malaria text in exposition 1 is built up by verbal and visual resources with the work distributed between them. The major classificatory taxonomies are antiparasitic drugs, species of malaria-causing plasmodia, and side effects of existing drugs, as shown in Figure 5. Although these are not deep taxonomies, they indicated to us on the first pass over the data that the artifact shifts between everyday (side effects of malaria) and discipline-specific (anti-parasitic drugs, species of plasmodium) fields. However, the student not only identifies these classes of side effects, antiparasitic drugs and species of plasmodium, but goes on to build meanings about these using verbal and visual resources. The student is also able to expand on points about the pathogen, building knowledge progressively through messages linked by relevant conjunctions (‘The pathogen reproduces in large numbers **and then** migrates **and** moves on to its next target’). Such implication sequences are always dynamic and are a significant way in which knowledge is built across the artifact (see Table 2). This point is taken up further in the LCT analysis below (see Section 7).

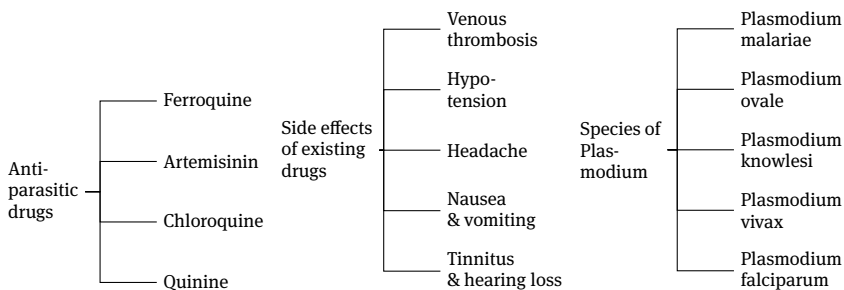


Fig. 5: Malaria: Classificatory taxonomies from exposition 1.

Tab. 2: Malaria implication sequences

01:10-02:08	Plasmodium parasite <i>enters</i> blood stream via female mosquito saliva — <i>travels</i> to liver and <i>reproduces</i> — <i>enters</i> red blood cell — parasite's digestive vacuole <i>converts</i> toxic haematin (derived from haemoglobin) safely to haemozoin — <i>feeds</i> and <i>continues to reproduce</i> — <i>causes</i> malaria
03:27-03:41	Anti-malarials (ferroquine, quinine and chloroquine) <i>inhibit</i> haemozoin formation in parasite's food vacuole — haematin <i>accumulates</i> — <i>kills</i> the parasite
03:45-04:02	Ferroquine iron action <i>generates</i> free radicals — <i>damages</i> parasites — cycle is <i>restarted</i>

In summary, genre and register analysis of the Malaria artifact indicates considerable control over the semiotic and disciplinary demands of the task. The student maker recognizes the demand to both persuade the viewer and explain scientific concepts. The presentation integrates these dual purposes and guides the viewer through the waves of information presented as image, language, and audio. Along with the student's proficiency in using the software, his conceptual understanding is another critical factor in its success, providing the evidentiary basis for and helping to build the overarching argument.

This section illustrated our initial analytical step in a lengthier, more complex analysis involving more delicate tools. This first stage was necessarily descriptive as it set out to map the data set so that the range of artifacts and the ways in which students responded to the task could be better understood. The next step involved examining the dimension of Field more closely, and again drawing on the Malaria text, shows how the student demonstrates his understanding of the biology of malaria and the pharmacokinetic mechanisms for treating it in order to prosecute the argument for ferroquine as an effective anti-malarial.

7 Investigating Knowledge Practices Using *Legitimation Code Theory*

As indicated in Figure 2, it is important for the creator of an explanation to understand what counts as knowledge in the domain. To consider how disciplinary knowledge is built across the digital artifact, we draw on a set of sociological tools that assume knowledge is 'real' and has characteristics that can have differing effects. LCT includes five dimensions: autonomy, temporality, density, specialization,

and semantics (Maton, 2014). The dimensions are used across a range of objects of study in many fields (see, e.g., Maton et al., 2016a), however, semantics has most commonly been utilized in science learning contexts.

The LCT dimension of semantics is particularly relevant to knowledge-building in science because it conceptualizes abstraction and complexity via two constructs: semantic gravity (SG) and semantic density (SD). Semantic gravity conceptualizes abstraction in practices (e.g., symbols, concepts, expressions, gestures, actions, etc.) and semantic density conceptualizes complexity. Complexity, or ‘condensation’ in LCT, is the degree to which meanings are dense or compact within the specialized practices of a particular field. Also, within the field, the degree of condensation of meaning varies along a continuum from stronger (SD+), where more meanings are condensed, to weaker (SD-), where fewer meanings are condensed.

Table 3 shows how semantic density was used to characterize complexity for language used in the narration of the Malaria sample text. This characterization closely followed a coding scheme from Maton & Doran (2017) for English language discourse, which is rather a ‘translation device’ that outlines how the theoretical construct of semantic density relates to empirical data (English language discourse in this case). The characterization assigns individual words to distinct categories that vary in ‘degree of condensation of meaning’. Theoretically, this characterization can facilitate as fine a division and as many categories as necessary. In the current study, four distinct categories were identified. The coding scheme utilizes bolding and/or capitalization as annotations of the different word types and sub-types.

Although the construct of semantic density has been utilized across various modes, including gestures and dance (see, e.g., Maton et al., 2016b), its operationalization is most developed in language. Nevertheless, semantic density was also used to consider images in the current study, although the four categories identified in the image analysis are not comprehensive and each category is much more diverse and ‘large’ in range when compared to language.

To follow how complexity manifested across narration and image in the Malaria text, values were assigned to the relative strengths of semantic density: technical conglomerate was assigned a value of 4, technical compact 3, everyday consolidated 1, and everyday common 0, and for image, scientific complex was assigned a value of 4, scientific simple 3, everyday real 1, and everyday illustrative 0 (as in Tables 3 and 4). Assigning values to words and phrases and adding them for each clause generates a Semantic Density Profile where points indicate a relatively stronger or weaker for semantic density as a function of the clause.

The purpose of this analysis is to explore how meaning is built through the text and to identify how a student might control complexity for a non-specialist audience. The SD profile in Figure 6 shows two distinct ‘patterns’ or points of

Tab. 3: Coding scheme for semantic density (SD) for language following Maton & Doran (2017).

SD	Type	Subtype	Value	Example
Stronger	<u>Technical</u> : meanings are given by their location within a specialized domain of social practice	CONGLOMERATE Comprise multiple distinct parts that each possess a technical meaning	4	monosaccharides
		Compact Comprise a single part with a technical meaning	3	force
	<u>Everyday</u> : meanings are given by their location in specialized domains but rather through their usage in commonplace practices in contents	CONSOLIDATED Encode happenings as processes or events that are normally realized by verbs or things as elements or items (physical or intangible) that are normally realized by nouns	1	production
		Common Happenings or qualities as qualities	0	person
Weaker				

interest: peaks in the language and plateaus in the image profiles (indicated by ellipses). We begin with the patterns because they act as points of interest. When considered more deeply, we see that there are interesting things happening, which we describe in further analyses here. Importantly, these patterns were further analyzed to understand how image and language work together to build knowledge across the text.

The labelled peaks demonstrate that complexity is mediated through both language and image; preceding and following these points of stronger semantic density are sections of weaker SD. For the first peak indicated in Figure 6, the two clauses include technical compact and technical conglomerate terms (e.g., kingdom, plasmodium). Coinciding with the opening stages of the embedded causal explanation, the origin of malaria is supported visually through text-on-screen (as repetition/highlight of narration) and an everyday illustrative image of

Tab. 4: Coding for SD analysis of images

SD	Image Type	Description	Value
Stronger	Scientific complex	Scientific representations with multiple parts or with one part that reflects a canonical representation in the discipline (e.g., found in textbook)	4
	Scientific simple	Simplified version of typical scientific representation	3
	Everyday real	real images (photoreal)	1
Weaker	Every illustrative	representations of real or imagined objects (e.g., cartoons, stylized images)	0
	text image		
No SD image	blank screen		

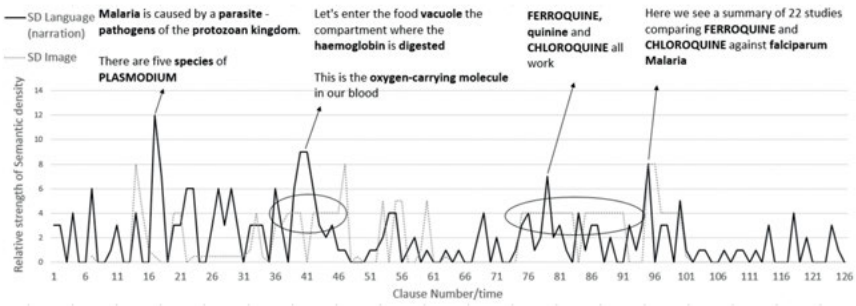


Fig. 6: Semantic density profile for language (—) and image (· ·) noting points of interest.

a mosquito. The clause indicated by this peak is preceded by ‘now let’s look at the cause’ and followed by ‘which can cause it’, clauses that contain only everyday words.

The careful use of complexity in the narration also occurs at other points in the text, as indicated by the other peaks in the profile (rather than as plateaus—sustained sections of complexity—which is otherwise typical of scientific texts). Together with the choice of image, the choice of language acts to ‘unpack’ complexity for a non-expert audience since the use of everyday image acts as a placeholder of meaning (e.g., the mosquito ultimately causes malaria) and represents an effective realization of a combination of resources. In this case, the creator is mediating

complex material by combining resources to help the viewer understand the material.

The ellipses noted in Figure 6 indicate sections where relatively strong semantic density is maintained in the image as the implication sequence ramps up. The image sequence in the first ellipse is depicted, in part, in the image sequence in Table 5. The images are coded as scientific complex that become more so as parts are added to build complexity both visually and textually.

Tab. 5: Increasing SD in image sequence mediated by narration [01:50-02:06]

	<p>Let's enter the food vacuole the compartment where the HAEMOGLOBIN is digested. This is the oxygen-carrying molecule in our blood and provides ample nutrition for the pathogen.</p>
	<p>The digestion of HAEMOGLOBIN produces something called HAEMATIN which is toxic to the parasite.</p>
	<p>The plasmodium negates this hazard by converting it into HAEMOZOIN which is excreted easily and safely.</p>

Rather than arbitrating the complexity, this section is embracing it, which is necessary in order to answer the assessment question but is also typical of explanatory texts in science. In the Malaria example, the two semantic density plateaus are related. One refers to the process of resistance, while the other shows how ferroquine is unique (e.g., because its mechanism of action does not result in resistance). Each SD plateau builds up field knowledge to answer the question about ferroquine as a novel anti-malarial. But only field knowledge related to the question is built. That is, these sections have been carefully chosen to elaborate and explain, whereas in other sections complexity is mediated and given ‘everyday’ placeholders.

Across the Malaria artifact, SD profile analysis illustrates how language and image have been used to build knowledge. In this example, technical and complex resources are used, but they are used in different ways to achieve different purposes. Sometimes, technical terms are used and ‘unpacked’. This manifests as a ‘peak’ in the semantic density values for language, where the unpacking occurs both in the language and inter-modally (e.g., through illustration). Where an explanation is provided, the building of complexity manifests as ‘plateaus’ in the image profile (sustained SD+). These sections indicate stretches across several clauses where the images (specifically, complex images) have been built up to focus on answering the question about ferroquine’s unique properties in treating malaria. Overall, the peaks and plateaus signal a control over complexity, mediating and building knowledge across the artifact.

Further, the peaks and troughs constitute a waving pattern. The waves illustrate how field-related meanings are recontextualized through upward and downward shifts representing ‘cumulative knowledge building’ (Maton, 2013). Maton identifies patterns of ‘semantic waves’ where peaks and troughs shift across the text and indicate the packing and repacking of complexity, effectively recontextualizing information. The patterns are independent of the discipline, according to Maton, because practices within different disciplines have different logics and organizing principles. This also points to ways to analyze how a digital explanation might not work as well. For example, if a peak remains as a sustained plateau, the lack of shifting may indicate that the text remains dense and insufficiently mediated for its intended viewers. If the pattern remains as a flat line low on the SD scale, the presentation may end up being too simple overall and insufficient as a scientific explanation. The waving patterns then possibly help to identify areas where students can be supported to work in this multimodal environment to make meaning for a particular audience.

This includes both what constitutes an explanation and how the affordances of the textual and visual resources are used to condense meanings through their interplay across the explanation. The creator builds complexity across the text by

navigating and mediating the choice of semiotic resources. LCT, and in particular, the dimension of semantic density, thus may support an integrated view of how semiotic resources are deployed to build field-related meanings in a digital explanation.

8 Negotiating Intersemiotic Resources

In investigating how the student-makers communicate an understanding of a specific scientific concept or phenomenon, it is necessary to explore how they have made use of the affordances of a variety of semiotic modes and resources. While genre and register enable us to capture in a broad sense how the maker has construed the task requirements and recruited resources to satisfy these, a fine-grained account is needed for how the affordances of the software are managed and the resources marshaled to realize their choices.

As flagged earlier in this chapter, SFS supports a consideration of the affordances of particular semiotic resources as well as their co-deployment. This deep exploration is warranted as the students' careful orchestration of these meaning-making resources is theorized as reflecting their rhetorical interests, including understanding of the concept to be communicated and the characteristics of the audience (Bezemer & Kress, 2008). By selecting, making, and coordinating multiple semiotic resources, the student-makers build an explanation, construe a complex technical field, establish a relationship with the audience and organize the information into a coherent and cohesive product.

Patterns in the use of specific semiotic resources within the artifacts were identified with the aid of the Multimodal Analysis Video (MMA) software (O'Halloran et al., 2012). This software allows the analyst to map and annotate semiotic resources as configured in a series of metafunctionally-organized syntagmatic choices over time. Using the software, the analyst places time-stamped annotation nodes in system strips which are synchronized with the video, sound, and verbal transcriptions of the artifact. The annotation nodes are color-coded allowing the analyst to visualize combinations of choices, which can be explored further in state transition diagrams in terms of their duration as a proportion of the total artifact time (O'Halloran et al., 2017).

These state transition diagrams can also be used to dynamically visualize shifts in state across the artifact. MMA aids the identification of couplings and the repeated co-patterning of realizations from two or more semiotic systems (Painter et al., 2013) that can be realized both intra- and inter-modally across the metafunctions. For example, within Malaria there are numerous clusters of

ideational meaning where a key entity, the ‘Malaria parasite’, is stated within the narration, depicted in an imported animation and labelled synonymously in written text. There are also regular couplings of ideational and interpersonal meanings in the artifact, particularly in phases where the need for new vaccines is established. For example, the spread of malaria is shown in a dynamic map and stated in the narration (an ideational meaning) alongside explicit negative appreciation of disease impacts (an interpersonal meaning) shown in the use of the color red in the map and stated in the following: “**immense** human suffering”.

A closer examination of how the creator manipulates a range of semiotic modes illuminates the meaning-making practices involved. To illustrate this process in action a key moment of complexity identified within the LCT analysis is explored further here. In the sequence presented in Table 5, a range of semiotic resources are clustered to represent the digestion of haemoglobin in the food vacuole. Using Kress & van Leeuwen’s (2006 [1996]) visual grammar we could identify what was contained in the implication sequence, however, such is the multi-faceted functionality of images that this sequence also acts analytically (to a lesser extent) identifying the entities involved in the digestion. These entities (Haemoglobin, Haemozoin, and Haematin) are depicted in a simple scientific graphic style to flag that this is a generalized process. It is accompanied by vectors (arrows) used as narrative processes to construct the implication sequence, with the direction of the arrows showing causality.

To mediate this process for the non-expert viewer the maker also labels these elements in recognition of the viewers’ likely unfamiliarity with such representations. Interestingly the pathogen itself is left unlabeled with its identification only possible by drawing upon the accompanying narration and the spatial proximity of the entities with the protozoan nudging the haemoglobin (Table 5, middle and final images). The creator draws upon the affordances of the visual mode to explain a process that is not accessible for the unmediated human eye and takes advantage of the dynamism of the medium to stagger the appearance of each entity in time with the spoken narration. Thus, although the complexity in both image and verbal resources increases in this section, they function together to mediate the complexity of the scientific process.

To create a multimodal text that achieves its rhetorical purpose depends not only on an understanding of the meaning-making potential in the cultural domain that includes particular resources but also upon the understanding of the medium used and the communicative situation (Bateman et al., 2017). Following Bateman et al., we consider the medium to be a canvas or material substrate whose material properties can be used to make distinctions that a community of users would recognize as meaningful. The meaning-making affordances of the digital medium are considered using a classificatory system to categorize the kinds of meanings

supported in that medium in terms of their material properties: temporality, spatial dimensionality, transience, participant/observer roles, and the reader's contribution to text co-construction (see Bateman et al. 2017 for an extended discussion of this classificatory system).

Classifying the medium in this manner offers analytic clarity and enables disciplined comparisons to other media. This is an important step for the analyst exploring such texts but also for the teaching of such multimodal phenomena because the classification offers great explanatory power. The Malaria artifact and others taking similar form bear similarities to traditional, static slideshow presentations, but, their dynamic and stand-alone nature means that they have different affordances. Traditional slideshow presentations are created to be delivered by a presenter who mediates the viewer experience of the slideshow by, for example, speech and gesture, and controls the temporal unfolding of the presentation. In contrast, the artifacts in this study need to function independently as stand-alone presentations and as such, logical organization, salience, and audience appeal are foregrounded. The creator of the Malaria artifact demonstrates an understanding of the affordances of the medium by taking advantage of the spatial and temporal dimensions of the medium to layer and compose meanings to build understanding of the pharmacokinetic action of ferroquine and its potential therapeutic benefits.

The presentation software employed by the student-maker also acts to mediate semiosis within the artifact. The software offers a range of tools and templates that can be used to construct a semiotic product; however, its use is regulated by the design of the software interface, which promotes and constrains particular ways of making meaning (Zhao et al., 2014). To illustrate the impact of the semiotic technology, the maker of Malaria takes advantage of the presentation software's spatial capacity to organize the stages of the explanation graphically. As we have seen, the Malaria presentation is organized around the infographic shown in Figure 3, which is comprised of a central segmented ring (e.g., macro-Themes) and six orbital satellites (e.g., hyper-Themes) that radiate from each of the segments.

As the explanation unfolds temporally, the viewer is moved from a central segment out to one of the satellites where more specific information is displayed and from where the creator can zoom in and out to focus on details before returning to the original central segment. The software's graphic display allows the maker to use placement and scale to show the structure of the presentation and to flag connections between elements. By making use of the software's capacity to smoothly expand and compress details, the maker manages the flow of information and constructs hierarchical relations between the information given. The software thus enables the creator to negotiate the complexity of the explanation for the viewer. Such analysis is also possible with other video production or presentation tools.

9 Discussion: Multimodality for Educational Practices

Thus far in the chapter, we have presented our analytic approaches to working with the multimodal artifacts we call ‘digital explanation’. This work, informed by systemic-functional semiotics and LCT, takes place in the interdisciplinary space where learners in university science contexts create these texts in response to assessment tasks requiring them to recontextualize and explain science knowledge for non-specialist audiences. While we presented detailed analyses for a single artifact, the range of issues raised is pertinent to any student-generated digital explanation.

The chapter has thus highlighted a number of issues for analysis in this complex realm. Issues include working at a sufficiently fine-grained level with lenses appropriate to the task of characterizing the artifacts. This work is necessarily adaptive since the tools were originally developed for use in other contexts. Here is where we draw on a semiotics ‘good enough’ (Macken-Horarik et al., 2011) to articulate the relationship between resources deployed and knowledge developed, to advance the field of multimodality in educational practice.

In the current section, we integrate insights from these analytic approaches to build a dialogue between semiotics and disciplinary knowledge practices. To ground this discussion, we consider multimodality a phenomenon of broad contemporary interest that is essential to disciplinary learning (Kress, 2010). Multimodality is also a field of study in its own right where attention is given to material affordances of semiotic resources and how these are put to work in different fields of practice. The field of multimodality also includes practical tasks within various disciplines, which is relevant for how ‘users’ develop their repertoires for work in the area. In reaching into analysis of multimodal artifacts, systemic-functional semiotics moves beyond a focus on language to incorporate images, typography and layout, amongst other resources (Bateman, 2008; Painter et al., 2013). Because making a digital explanation involves exploration and integration of different resources, we need to consider the possibilities for meaning and learning inherent in working multimodally.

There are two key challenges to this work: analysts need to comprehensively account for how semiosis operates in a particular digital explanation, while creators need to understand the separate affordances of semiotic resources and the effects of integrating them as facets to their text construction. Analysis must also allow for an appropriate level of delicacy for the meanings within each of the semiotic resources, which means framing the analysis in both macro terms (e.g., genre and register categories) and micro terms (semiosis of messages as the text unfolds).

In our work with the range of student-created digital artifacts in the current study, we used genre and register analysis to map the multimodal context. Firstly, students must recognize the demands of the task in order to construct a text that is often a combination of genres, organized either sequentially or as embedded forms. They must also demonstrate their control of the relevant field knowledge as well as navigate complex Tenor demands to translate the knowledge for a non-specialist audience. When resources like sound, moving image, figures, and zooming in and out are integrated with the dynamic affordances of presentation software (e.g., Prezi), analysis must be alive to the meaning potentials of each resource and their resultant interplay on the screen.

From a social realist knowledge perspective, knowledge practices need to be understood semiotically and sociologically. The LCT concept of semantic density allows us to see how meanings are condensed visually and verbally: the more meanings are condensed, the higher the level of semantic density, and thus the more complex the text. One could argue that specialized knowledge of the biology of malaria and its potential pharmacological cures will involve higher degrees of semantic density, especially when the audience is assumed to share some knowledge of the field.

However, a text that aims to explain technical information for a non-expert audience must unpack technical and specialist terms and concepts. This can be done through combining resources that are image or language-based as illustrated in the SD profiles. SD profiles then graphically illustrate shifts in the relative strength of semantic density across the text and show how meanings can be ‘built up’ or ‘unpacked’ through narration and image. Points of semantic density capture the interplay of image and verbiage and show how semiotic resources are deployed to negotiate complexity across the digital explanation.

On the science side of the current dialogue, science needs a discipline-specific account of semiosis, including what counts, where, and why. This account must also reflect how individual students participate in science learning activity that may include generating a digital explanation as an assessment task in a disciplinary science context. Language and other semiotic resources must also be translated into metalanguages with which scientists and science teachers can work. Finally, disciplinary science fields must support students to build science literacy repertoires over time and with further learning, which builds on the concept of ‘cumulative semiotic progression’ (Hoban & Nielsen, 2013). Hoban and Nielsen argue that building a sequence of representations, as in a digital explanation, builds understanding because the creator must choose resources and resequence them to progressively build meaning across the artifact. Our analysis begins to account for how meanings are accumulated and integrated in the macrogenre of digital explanation.

10 Conclusion

Constructing a digital explanation is an opportunity to build knowledge through translating science content multimodally, but such tasks impose significant technical and conceptual demands on the creator. Expanding, technicalizing, and unpacking knowledge and integrating semiotic resources are processes that a student must master in order to produce an effective artifact. A multimodal theory that captures the relevant semiotic and social processes so as to advance scientific communication and education in the complexity of the modern digital world is needed.

A discipline of multimodality needs tools, techniques and objects of study to build theories of the important relationship between disciplinary practices and how learners are enculturated into them during their studies. SFS and LCT are thus theories that offer complementary analytical tools for examining digital explanations as artifacts. Future directions for this research include translating these analytical and theoretical insights into teachers' practices, knowledge that could fruitfully be used to support students to advance their understandings and literacies in producing multimodal artifacts.

In this chapter, we have advanced a model of knowledge building in educational disciplines, for understanding student-generated digital artifacts, which extends to how students develop and communicate complex scientific information across multimodal formats. Building from the model of relations between context and language by Halliday (1999), we shaped four key questions to model analysis of digital explanations as multimodal artifacts:

1. What kinds of knowledge 'count' in this domain and why? (cultural domain/relevance)
2. What do we need to know about language and other semiotic resources to tackle tasks of this domain? (register/resources)
3. How do we participate effectively in this situation and others like it? (situation type/rhetoric)
4. How does this text relate to others like it and how do we deploy what we know in further learning? (text type/repertoire)

Looking towards a 'new' discipline of multimodality, creating a digital explanation bridges the interface between literacy practices in a science discipline and how knowledge is demonstrated and negotiated for different audiences. As a discipline, multimodality must articulate what learners need to do at this interface. Our

emerging analyses and modeling of the relations between systemics and knowledge building in science help to shape this dialogue at the intersection of literacy practices and knowledge building. Further, this knowledge could meaningfully translate to science communication and a wider focus on making new science knowledge accessible beyond the domain of experts.

Bibliography

- Bateman, John A. 2008. *Multimodality and Genre: A Foundation for the Systematic Analysis of Multimodal Documents*. Basingstoke: Palgrave Macmillan.
- Bateman, John A., Janina Wildfeuer & Tuomo Hiippala. 2017. *Multimodality – Foundations, Research and Analysis. A Problem-Oriented Introduction*. Berlin: Mouton de Gruyter.
- Bernstein, Basil. 1977. *Class, Codes and Control (Vol. 3): Towards a Theory of Educational Transmissions*. London: Routledge.
- Bezemer, Jeff & Gunther Kress. 2008. Writing in Multimodal Texts: A Social Semiotic Account of Designs for Learning. *Written Communication* 25. 166–195.
- Chan, Eveline & Len Unsworth. 2011. Image-Language Interaction in Online Reading Environments: Challenges for Students' Reading Comprehension. *Australian Educational Researcher* 38(2). 181–202.
- Estrada, Fabiola Christina Rodríguez & Lloyd Spencer Davis. 2015. Science Through the Incorporation of Graphic Design Theories and Practices into Science Communication. *Science Communication* 37. 140–148.
- Halliday, Michael A. K. 1978. *Language as Social Semiotic*. London: Edward Arnold.
- Halliday, Michael A. K. 1994. *An Introduction to Functional Grammar. Second Edition*. London: Edward Arnold.
- Halliday, Michael A. K. 1999. The Motion of 'Context' in Language Education. In Mohsen Gladessy (ed.), *Language Development: Interaction and Development*, Amsterdam: John Benjamins.
- Halliday, Michael A. K. & Ruqaiya Hasan. 1985. *Language, Context and Text: A Social Semiotic Perspective*. Geelong, Victoria: Deakin University Press. (Language and Learning Series). Also: Oxford University Press, London, 1989.
- Hoban, Garry & Wendy Nielsen. 2013. Learning Science through Creating a 'Slowmation': A Case Study of Preservice Primary Teachers. *International Journal of Science Education* 35. 119–146.
- Hoban, Garry, Wendy Nielsen & Chris Hyland. 2016a. Blended Media: Student-Generated Mash-ups to Promote Engagement with Science Content. *International Journal of Mobile and Blended Learning* 8(3). 35–48.
- Hoban, Garry, Wendy Nielsen & Alyce Shepherd. 2016b. *Student-Generated Digital Media in Science Education*. London: Routledge.
- Kress, Gunther. 2010. *Multimodality: A Social Semiotic Approach to Contemporary Communication*. London: Routledge.
- Kress, Gunther & Theo van Leeuwen. 2006 [1996]. *Reading Images: The Grammar of Visual Design*. London and New York: Routledge.
- van Leeuwen, Theo. 2005. *Introducing Social Semiotics*. London: Routledge.

- Lemke, Jay. 1999. Typological and Topological Meaning in Diagnostic Discourse. *Discourse Processes* 27. 173–185.
- Macken-Horarik, Mary & Misty Adoniou. 2007. Genre and Register in Multiliteracies. In Bernard Spolsky & Francis Hult (eds.), *Handbook of Educational Linguistics*, Blackwell: Oxford University.
- Macken-Horarik, Mary, Christina Love, Carmel Sandford & Len Unsworth (eds.). 2018. *Functional Grammaticals: Re-conceptualizing Knowledge about Language and Image for School English*. London: Routledge.
- Macken-Horarik, Mary, Christina Love & Len Unsworth. 2011. A Grammaticals ‘Good Enough’ for School English in the 21st Century: Four Challenges in Realizing the Potential. *Australian Journal of Language and Literacy* 34. 9–23.
- Martin, James & Karl Maton. 2017. Systemic Functional Linguistics and Legitimation Code Theory on Education: Rethinking Field and Knowledge Structure. *Onomazein* 12–45. <https://doi.org/10.7764/onomazein.sfl.02>.
- Martin, James & David Rose. 2012. Genres and Texts: Living in the Real World. *Indonesian Journal of SFL* 1. 1–21.
- Martin, James R. 1993. Technicality and Abstraction: Language for the Creation of Specialised Texts. In M. A. K. Halliday & J. R. Martin (eds.), *Writing Science: Literacy and Discursive Power* Critical Perspectives on Literacy and Education, 203–220. London: Falmer.
- Maton, Karl. 2013. *Knowledge and Knowers: Towards a Realist Sociology of Education*. London and New York: Routledge.
- Maton, Karl. 2014. *Knowledge and Knowers: Towards a Realist Sociology of Education*. London and New York: Routledge.
- Maton, Karl & Yaegan Doran. 2017. Semantic Density: A Translation Device for Revealing Complexity of Knowledge Practices in Discourse. Part 1 – Wording. *Onomazein* 1. 46–76.
- Maton, Karl, Susan Hood & Suellen Shay (eds.). 2016a. *Knowledge-Building: Educational Studies in Legitimation Code Theory*. Abingdon: Routledge.
- Maton, Karl, James Martin & Erika Matruglio. 2016b. LCT and Systemic Functional Linguistics: Enacting Complementary Theories for Explanatory Power. In Karl Maton, Susan Hood & Suellen Shay (eds.), *Knowledge-Building: Educational Studies in Legitimation Code Theory*, 93–113. Abingdon: Routledge.
- Nielsen, Wendy, Helen Georgiou, Pauline Jones & Annette Turney. 2018. Digital Explanation as Assessment in University Science. *Research in Science Education* 1–28. <https://doi.org/10.1007/s11165-018-9785-9>.
- O’Halloran, Kay, Sabine Tan & Marissa K. L. E. 2017. Multimodal Analysis for Critical Thinking. *Learning, Media and Technology* 42. 147–170. <https://doi.org/10.1080/17439884.2016.1101003>.
- O’Halloran, Kay L. 2005. *Mathematical Discourse: Language, Symbolism and Visual Images*. London and New York: Continuum.
- O’Halloran, Kay L. 2008. Systemic Functional-Multimodal Discourse Analysis (SF-MDA): Constructing Ideational Meaning using Language and Visual Imagery. *Visual Communication* 7(4). 443–475.
- O’Halloran, Kay L., Alexey Podlasov, Alvin Chua & Marissa K.L. E. 2012. Interactive Software for Multimodal Analysis. *Visual Communication* 11(3). 363–381. <https://doi.org/10.1177/1470357212446414>.
- O’Toole, Michael. 1994. *The Language of Displayed Art*. London: Leicester University Press (Pinter).

- Painter, Claire, James R. Martin & Len Unsworth. 2013. *Reading Visual Narratives: Image Analysis of Children's Picture Books*. London: Equinox.
- Prain, Vaugh & Russell Tytler. 2012. Learning Through Constructing Representations in Science: A Framework of Representational Construction Affordances. *International Journal of Science Education* 34. 2751–2773.
- Rigutto, Christina. 2017. The Landscape of Online Visual Communication of Science. *Journal of Science Communication* 16(2). <https://doi.org/10.22323/2.16020306>.
- Rothery, Jan. 1996. Making Changes: Developing an Educational Linguistics. In Ruqaiya Hasan & Geoffrey Williams (eds.), *Literacy in Society*, 86–123. Essex: Addison Wesley Longman.
- Szenes, Eszter. 2017. *The Linguistic Construction of Business Reasoning: Towards a Language-Based Model of Decision-Making in Undergraduate Business*. Sydney: University of Sydney dissertation.
- Tang, Kok-Sing, Cesar Delgado & Elizabeth R. Moje. 2014. An Integrative Framework for the Analysis of Multiple and Multimodal Representations for Meaning-Making in Science Education. *Science Education* 98. 305–326.
- Unsworth, Len. 2006. Towards a Metalanguage for Multiliteracies Education: Describing the Meaning-Making Resources of Language-Image Interaction. *English Teaching: Practice and Critique* 5(1). 55–76.
- van Leeuwen, Theo. 1999. *Speech, Music, Sound*. London: Macmillan.
- Warden, Rachel. 2010. The Internet and Science Communication: Blurring the Boundaries. *Ecancermedicalscience* 4(1). 1–8.
- Yeo, Jennifer & John Gilbert. 2017. The Role of Representations in Students' Explanations of Phenomena in Physics. In David Treagust, Reinders Duit & Hans E. Fisher (eds.), *Multiple Representations in Physics Education*, 255–287. Amsterdam: Springer International.
- Zhao, Sumin, Emilia Djonov & Theo van Leeuwen. 2014. Semiotic Technology and Practice: A Multimodal Social Semiotic Approach to PowerPoint. *Text and Talk* 34. 349–375.

Dušan Stamenković and Milan Jačević

Video Games and Multimodality: Exploring Interfaces and Analyzing Video Screens Using the *GeM* Model

Abstract: With the purpose of showing that video game studies should become part of the emerging discipline of multimodality, the present chapter introduces the basics of the study of gameworlds and uses the multimodal document approach to analyze document-like screens coming from two video games: *Football Manager 2018* and *Europa Universalis IV*. These document-like screens are analyzed using the tools coming from the GeM model, which treats these pages as multi-layered semiotic artifacts. Within this approach, all four layers are covered: the base layer, the layout layer, the rhetorical layer, and the navigation layer. Our analysis proposal tries to pinpoint the semiotic specificities of the different layers and test whether the GeM model needs to be adapted for the purpose of approaching these screens. At the same time, the gameworld environment is viewed as an important mediator between the player and the digital game system. We hope that such integrated approach can be beneficial to both multimodality and video game studies and expand the directions of future research endeavors.

Keywords: multimodality, video games, documents, gameworlds, screens, GeM model

1 Introduction: A Multimodal Approach to Video Games

The present chapter was inspired by a recent comment saying that “computer and video games present an elusive but by no means unattainable target for multimodal analysis” (Bateman et al., 2017, 378). Its main purpose is thus to attempt to bring the new discipline of multimodality closer to attaining this target.

On the one hand, we are facing the fact that video games indeed represent a purely multimodal artifact, whereas, on the other hand, we have so far witnessed the state of affairs in which video games have rarely been studied within the existing multimodal approaches. We hope that the chapter will initiate further research into both main targets of our approach—game interfaces and multimodal documents within video games—so as to enrich the present set of phenomena investigated within the multimodal research community.

<https://doi.org/10.1515/9783110608694-011>

On the other hand, video game developers can benefit from such multimodal analyses and from multimodality as a discipline, as the results may show the ways in which interfaces and screens may be altered to reach more effective and more intuitive forms of communicating with the end user. The potentials of investigating video games within the domain of multimodal studies have already been stressed. Bateman et al. note that

[...] approaching games from a clear analytical perspective—as, for example, the discourse analytical perspective or any other perspective for that matter—will benefit from being able to handle their multimodality. This may be considered equally beneficial for game studies, in which multimodality may be considered from the viewpoint of human-computer interaction (Bateman et al., 2017, 366).

What we intend to do in this chapter represents just two of the many ways in which multimodality and video game studies may interact. At the same time, the chapter at least partly follows the line of research we started with applying a multimodal discourse analytical approach to *Metal Gear Solid* (Stamenković et al., 2017). We hope to encourage more research related to video games, which might start with exploring the ways in which the current multimodal apparatus might treat video gaming content.

As stated above, the aims of our chapter are two-fold: one is to focus on video game interfaces, the other on specific video game screens that could be approached as multimodal documents. When it comes to integrating game interface studies into the realm of multimodality, our approach is complementary to that of Kristine Jørgensen, who conceptualizes the gameworld itself as an interface, as “an informational and interactive environment” (Jørgensen, 2013, 4), which metaphorically represents the underlying computational system of the game (Jørgensen, 2013, 75) and thus has an essentially communicative role.

Our analysis proposal will try to pinpoint the semiotic specificities of the different layers and elements which comprise game interfaces, taking into account not only the WIMP (window, icon, menu, pointer) features, but the gameworld environment as well, since the gameworld is an important mediator between the player and the digital game system (Jørgensen, 2013, 20). Secondly, we will use a set of tools stemming from the multimodal document approach (Bateman et al., 2002; Bateman, 2008, 2013, 2014) to analyze screens coming from two popular managerial and strategy video games: *Football Manager 2018* (Sports Interactive, 2017) and *Europa Universalis IV* (Paradox Development Studio, 2013a). The document-like screens are viewed as potential objects of a multimodal analysis performed using the *GeM* model (Bateman, 2008, 2013, 2014; Bateman et al., 2002; Hiippala, 2015, 2017; Thomas, 2007, 2009, 2014), which treats these pages as multi-layered semiotic artifacts. We will focus on those aspects of the model which might reveal

the greatest differences between document-like video game screens and other multimodal documents.

In the following section, we view video game interfaces and gameworlds from the perspective of multimodality, while the third section represents an attempt to use the existing framework coming from the multimodal document approach and analyze document-like video game screens. In the concluding section, we try to show the benefits of our approach in regard to multimodality, the GeM model (as one of its tools), and the video game industry, as well as to present some future research directions.

2 Video Game Interfaces and Multimodality

The concept of interface has been defined and utilized in different ways within the field of game studies, necessitating a brief overview of the understandings of the term within the context of this particular study. For Jesse Schell, the interface in relation to video games can generally be understood as every element of mediation between the player and the gameworld (Schell, 2008, 223). Schell points out the basic distinction between the physical and virtual interfaces in relation to video games: in the case of the former, the term is used to refer to the physical input and output components which facilitate interaction (Schell, 2008, 223–224), while in the case of the latter, the term broadly designates a “conceptual layer that exists between the physical input/output and the game world” (Schell, 2008, 224), comprising both input elements such as menus and output elements such as score displays. Authors such as Mia Consalvo and Nathan Dutton, in turn, define the concept of interface solely in relation to the software environment, as

[...] any on-screen information that provides the player with information concerning the life, health, location or status of the character(s), as well as battle or action menus, nested menus that control options such as advancement grids or weapon selections, or additional screens that give the player more control over manipulating elements of gameplay (Consalvo & Dutton, 2006, para. 18).

Much like that of Consalvo and Dutton, various other definitions of interface in relation to video games stress visuality as a key property of the concept, often at the expense of other semiotic modes; in addition, it is often unclear what constitutes a part of the interface proper, and what does not. In their review of terms used in describing player-game interaction, Loïc Caroux and colleagues discuss the interface with a focus on information output, finding that said output has generally been described in relation to visual and auditory information, with the former

being more discussed and studied than the latter in relation to player experience (Caroux et al., 2015, 370). For Caroux and colleagues, visual interfaces in video games consist of “a main action scene containing objects with which the player can interact (e.g. avatars, enemies, or targets) and a complex, moving background (e.g., interiors, landscapes)” (Caroux et al., 2015, 370), with the authors differentiating these from head-up displays (HUDs), which are superimposed on the action scene and provide contextual information.

While Caroux and colleagues discuss the HUD as distinct from other elements of the visual interface in video games, for Kristine Jørgensen, the video game interface is not limited to the HUD and traditional WIMP features, but also includes other signifying elements found in the gameworld, such as “color schemes that signal the importance of specific objects, or animations that provide information about the current state of objects or characters” (Jørgensen, 2012, 146). In Jørgensen’s understanding of the term, the gameworld is “a world representation that communicates the game-system information that the player needs to know in order to play the game meaningfully” (Jørgensen, 2013, 4). As such, the concept is closely linked to the concept of interface.

According to Jørgensen, a gameworld is both an interactive and informational environment, representing the formal game system and facilitating player-game interaction, while also encompassing superimposed features such as menus and maps, which are functionally coupled to the events which take place in the gameworld (Jørgensen, 2013, 4). Gameworlds should, therefore, be understood both with reference to their properties as interfaces, due to the fact that they represent the formal game system, and as “world constructs suggesting ecological environments” (Jørgensen, 2013, 55) and characterized by worldness (Klastrup, 2003, 2010), a trait which designates these environments as discrete, unique worlds distinguishable from other worlds (Jørgensen, 2013, 55–56).

According to Jørgensen, due to the fact that gameworlds are artificially designed and constructed spaces whose primary purpose is to facilitate ludic activity, all aspects of said worlds can be said to contain relevant gameplay information and help to support the players interaction with the world, including the “traditional WIMP interface features and additional overlay information” (Jørgensen, 2013, 57). In relation to the gameworld, interface features may be implemented on several levels. In Jørgensen’s view, these features may be:

- iconic, sharing a correspondence with real-world features (e.g., a specific weapon utilized by a character in a game, signifying relative power but otherwise not being specifically accentuated);
- emphasized, highlighted or enhanced in some way (e.g., a strikingly-colored item, signifying that it has been selected by the player);

- integrated in the gameworld with no correspondence to real-world features (e.g., a quest marker above a character, signifying the potential to begin a quest);
- overlaid onto the gameworld (e.g., a menu screen popping up when in a shop, signifying the ability to purchase and sell goods); or
- metaphorical, appearing external to the gameworld but still conveying some information in relation to it (e.g., a sound cue, signifying the start of a combat encounter) (Jørgensen, 2012, 147).

This view of gameworlds positions them as information systems of multimodal signification, as part of which traditional interface features, such as WIMPs and HUDs, are experienced and interpreted by the players in relation to the broader context of the ludic activity; therefore, they are understood as integrated and coherent in relation to said activity (Jørgensen, 2013, 61–62). In their discussion of computer game interfaces, and drawing on qualitative studies conducted with players and game developers, Jørgensen points out the fact that interface features seem to be generally accepted within the context of the ludic activity, so long as they provide players with pertinent information necessary for the continuation of said activity (Jørgensen, 2012, 152). In addition, interface features seem to also be accepted due to the fact that they have, throughout the years, become normalized in relation to video games, representing “a convention and a part of the communicative toolset” (Jørgensen, 2012, 153) of these kinds of games.

From the perspective of multimodality, Jørgensen’s description of gameworlds as virtual environments with several layers of signification has important implications. First of all, the framing of the player-game system relationship as mediated by gameworlds sets the stage for a more comprehensive analysis and interpretation of the semiosis of a particular game, one in which the focus is extended from isolated traditional interface features, common to the domain of human-computer interaction, to encompass all those elements contained within the constructed activity space of the game which signify pertinent information to the player.

When approaching a video game with the aim of conducting a semiotic analysis, researchers would therefore do well to not disregard what Jørgensen describes as iconic and metaphorical interface features in particular—such as, for example, the different outfits worn by player avatars in *Fortnite* (Epic Games/People Can Fly 2017), some of which are only available after prolonged periods of play and may indicate a high-level player, or the sound cue in a game like *Final Fantasy VIII* (Square, 1999), which signifies the commencement of a random battle encounter.

Consequently, conceptualizing the gameworld itself as an interface also leads to a broader sensitivity towards the different modes of signification which may be

utilized within a game, and which semiotic analyses need to take into account. In addition to aural, visual, textual, and spatial elements, which may be overlaid onto the gameworld in the form of WIMPs or implemented in the ludic environment in a more integrated manner, one particularly interesting topic of analysis could be the implementation of the haptic modality, most often found in the form of vibrational feedback for certain actions performed by the player in the gameworld.

This form of feedback, utilized in video games since the mid-1970s (Wolf, 2008, 39), has been used to convey various types of information to the player, from the proximity of enemy soldiers in *Metal Gear Solid 3: Snake Eater* (Konami Computer Entertainment Japan, 2004) to the sensation of a vehicle crashing into a solid object in games like *Uncharted 4: A Thief's End* (Paradox Development Studio, 2013b). An exclusive focus on traditional interface features and textual and visual systems of signification then risks sidelining or even neglecting other semiotic modes, consistently implemented in video games but considered more unorthodox in relation to other digital artifacts.

3 Analyzing Video Game Screens as Multimodal Documents

Given the fact that Bateman et al. (2017, 366–378) have already approached typical game interfaces and canvases from a multimodal perspective, touching upon several levels of signification (but not using Jørgensen's terms), we wanted to investigate screens that represent a rather special case in the world of gaming. We will refer to them using the term document-like video game screens, which we can frequently find in simulation and strategy video games, though they are not limited to these genres, as they can appear in role-playing, adventure and sports video games as well.

This approach is intended to complement the multimodal approach proposed by Bateman, Wildfeuer and Hiippala, as their 2017 proposal only briefly focused on canvases that depart from “the normal view” (Bateman et al., 2017, 371) and employ written language, 3D illustrations, diagrams, pictograms, etc., which they illustrated using the technology tree view from Sid Meier's *Civilization V* (Firaxis Games, 2010). Moreover, there are video games where these screens are, in fact, ‘the normal view’, so we believe that they deserved to be treated in a way which will allow us to explore their facets, and possibly improve such screens using multimodal analytical tools.

The screens we want to address belong to the realm of video game materiality and resemble documents in formal and functional terms—their appearance

combines text, images, and other types of graphics in a manner typical of multimodal documents; at the same time, within the gameworld, they frequently have functions similar to those of real-world documents, e.g., they can inform, ask for information, or change the game reality in the way in which real-world documents can change our reality, which may involve virtually sending offers, signing contracts or treaties, declaring wars, etc. (all similar to document acts described in Smith 2014).

Given the wide range of successful applications of the Genre and Multimodality approach (abbreviated *GeM*; for a recent review of its applications see Hiippala 2017), largely used to scrutinize page-based documents and similar items, the logical choice for us was to try to apply it to document-like video game screens. Our approach is then mostly based on Bateman (2008, 2014) and Hiippala (2015). Although video games have so far not been in *GeM*'s focus, there is a range of video games containing similar to the artifacts already analyzed using *GeM*. Along with exploring such screens using the *GeM* model, we also wanted to investigate whether it would (or would not) be necessary to adapt *GeM* for such an analysis. For this purpose, we will use two document-like screens from two popular video games, *Football Manager 2018* (Sports Interactive, 2017) and *Europa Universalis IV* (Paradox Development Studio, 2013a).

Our exploratory analysis focuses on step 1 and step 2 within the methodological process described in Hiippala (2017, 3), and therefore deals with analyzing the four usual layers within the *GeM* model (see Bateman, 2008):

- the base layer, which identifies the elements which can serve as the common denominator for interpretative, textual, and layout elements in any analysis of a document;
- the layout layer, which focuses on those elements which are prominent when a reader encounters a page: it includes the hierarchical and spatial organization of the page, along with its typographic and graphic characteristics;
- the rhetorical layer, which uses Rhetorical Structure Theory (RST, Taboada & Mann, 2006) to establish discourse relations between content elements;
- the navigation layer, which refers to the elements that help the user interact with the document.

In our case, step 1, which focuses on the base layer, has the goal of dividing the screen contents into pre-defined *Recognized Base Units* or *RBUs*, which are then further fed to step 2 for description. *Recognized Base Units* feature elements that should cover most of what one could find in a multimodal document: arrows, captions (of photos, drawings, diagrams), tables, connecting lines, delimiting lines,

diagrams, drawings, emphasized text, figures, floating text, footnotes, footnote labels, headings, headlines, icons, list items, list labels, maps, menu items, page numbers, photos, running heads, sentences, superimposed text list, table cells, text (in photos, drawings, diagrams), titles, etc. (this list is based on Bateman (2008) and Hiippala (2015), although it may also be updated).

Our first task was to see whether the established set of *RBU*s will be enough to treat video game screens that have document-like features. The main screen in our analysis comes from *Football Manager 2018* (Sports Interactive 2017), the latest installment in the long-running *Football Manager* series (previously called *Championship Manager*, developed by Sports Interactive since the mid-1990s), which itself has a unique story of multimodality. The series seems to be a perfect candidate for a diachronic multimodal analysis of very many elements within itself, as its design has evolved from almost entirely textual to highly multimodal (with different intermediary stages in between). Along with being multimodal, it is also interactive in a way which, in video games, differs from the use of the term in analyses of printed materials, and this element will be incorporated in the proposed analysis. Therefore, the game's development has covered a long multimodal path, which makes it a good focus for the current approach.

Being a managerial simulator, after over 20 years of development, and in spite of the substantial visualization, the materiality of *Football Manager* has retained considerable textual content, which still makes many of its screens look like interactive multimodal documents. Of course, we are aware of the fact that these documents do not entirely (or sometimes at all) pertain to reality—they all belong to a virtual world which very much resembles our reality at the very beginning, but departs from it as the game progresses, simply because it is not only a simulator, but also a game.

Part of being a good *Football Manager* game player is knowing how to read these screens. They cover a whole range of details necessary to find your way in the world of football—competition tables, statistics, records and histories, playing and non-playing staff general information, characteristics and career details, club information, squad lists, shortlists, transfer data, etc. The game itself includes perhaps the biggest football database in the world—in the process of conversion a great deal of this database is seen through screens that look like multimodal documents and mostly combine text and images or other graphic elements. For the purpose of identifying *RBU*s in a video game screen, we have selected the club overview screen of the Portsmouth football team as an exemplar, and Figure 1 shows our annotation and basic labelling of each element on the screen. We should note that the screen in question (like almost every screen in *Football Manager*) is outlined by a side-bar and a top-bar, which represent the main components of the user interface overlaid onto the gameworld.



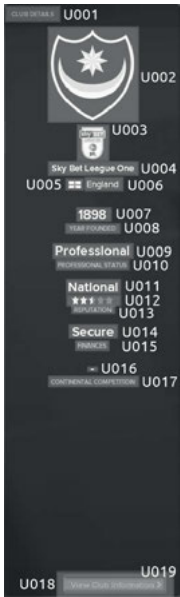
Fig. 1: Identifying RBUs in an FM screen; marking up the elements; club overview in *Football Manager 2018*, Copyright ©2017 Sports Interactive.

As we can see in Figure 1, our bottom-up annotation process (the only option we had in the current approach) resulted in very many components identified on the screen. We were aware of the possibility of embedding one RBU within another (Bateman, 2008, 113), but for the present analysis wanted to explore a more detailed approach, which is why we decided to mark all details on the screen. After identifying the elements, we attempted to classify them into one of the *RBU* categories. In Table 1 we present one part of our description, which should provide enough material to illustrate the whole process and allow us to comment on potential specificities or potential difficulties of a video game screen analysis as compared to the analysis of other multimodal documents.

The categorization and description process revealed that the differences existing between annotating a document-like video game screen and other multimodal documents do not seem to be big at all. The greatest difference is perhaps the fact that document-like video game screens feature interactivity, which is embodied in several different ways. In the presented part of the screen, interactivity seems to be embodied by hyperlinks ‘hidden’ behind boxes and text that compose what we would call buttons—these items represent ‘clickable’ elements that can take the player to another screen.

It is very likely that computer vision and optical character recognition techniques could detect only the surface form, whereas hypertext and hyperlinks will need to be analyzed on another level of description. Along with hyperlinks, in Figure 1, we find another item that can be classified as interactive: U043 is an example of a drop-down menu, which seems to be interactive in a different way—here, it

Tab. 1: Identifying RBUs in a document-like screen; categorizing and describing the elements in one part of the club overview screen in *Football Manager 2018*, Copyright ©2017 Sports Interactive.

	<p>U001: text Club details (heading)</p> <p>U002: figure Portsmouth Logo +hyperlink</p> <p>U003: figure Sky Bet League One Logo +hyperlink</p> <p>U004: text Sky Bet League One Logo (caption) +hyperlink</p> <p>U005: figure The Flag of England +hyperlink</p> <p>U006: text England +hyperlink</p> <p>U007: text 1898</p> <p>U008: text YEAR FOUNDED</p> <p>U009: text Professional</p> <p>U010: text PROFESSIONAL STATUS</p> <p>U011: text National</p> <p>U012: figure quality stars, 2.5 out of 5.0 filled</p> <p>U013: text REPUTATION</p> <p>U014: text Secure</p> <p>U015: text FINANCES</p> <p>U016: text -</p> <p>U017: text CONTINENTAL COMPETITION</p> <p>U018: two-d-element Box</p> <p>U019: text View Club Information > +hyperlink = button</p>
-----------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

allows one to select a team to be included in the graph alongside Portsmouth (and thus enables graph comparison). There are other forms of interaction in screens similar to the one we have presented; these could include sliders, arrows, checkboxes, drag-and-drop items, and other elements or sub-menus that might be more complex in both form and content. An issue related to interactivity that can be particularly problematic when it comes to RBU-layer analyses are elements that change form when a user moves the pointer over them.

In order to show this, we have taken a screenshot from *Europa Universalis IV* (Paradox Development Studio 2013), the latest installment in the *Europa Universalis* grand strategy video game series. The screen in Figure 2 represents the Estates tab in the country view menu. Hovering the mouse over one of the items (types of population) on the list elicits a box containing a piece of text—in fact a hypertext which temporarily appears on the screen (marked by a large oval shape). This hypertext feature allows the game to provide more information on certain issues only when needed, and without taking up a large portion of the screen. Otherwise, it remains completely invisible—the main text itself (‘Nobility’) does not reveal that there is anything hidden in it. The second element, marked by a small square,

is the interactions button, which allows the player to interact with the item on the list, i.e., ask for contributions, demand diplomatic support, grant monopoly charters, etc., all of which are ludic actions, i.e., not something one would find in real life.



Fig. 2: Game-specific elements in document-like screens; estates in the country view menu in *Europa Universalis IV*, Copyright ©2013 Paradox Interactive AB.

Another difference is the vector nature of the screens we find in *Football Manager*, so some document-like screens will look different based on the resolution used, a factor to which computer vision and recognition software packages are already able to adapt. The third peculiarity related to document-like video game screens is related to the contents rather than the form. It can be seen in what we could call ‘fictional elements’ derived from the mutable ludic nature of video games as a medium and which we are far less likely to encounter in other multimodal documents. The content of these elements, in fact, make video games less factual and reminds us of the fact that such screens have been extracted from a game.

The initial state of *Football Manager* overlaps substantially with reality and the football world as it is, but with every ludic step it gets more distant from the real state of things. Even in the screenshot in Figure 1 we can see a fictional manager and randomized fixtures, which illustrates this third difference.

Given the space constraints of the current chapter, in our discussion of step 2 we focus on the layout structure from the layout layer and on relations belonging to the rhetorical layer. We attempt to combine both in the graph presented in Figure 3; in the manner of the GeM model, layout elements related to the club overview screen presented in Figure 1 are represented as boxes and internal links as lines, and rhetorical relations are given in a text box for the sake of clarity. We have expanded only the L.1.1 section, containing the elements zoomed in on in Table 1, since including layout elements from all sections in Table 1 would require more space. However, these elements are sufficient to illustrate a typical layout of a document-like screen in *Football Manager*.

In Figure 2 we can see that the overall layout structure is not too different from what we usually encounter in multimodal documents coming from other sources (Hiippala, 2017). Similarly to the case of the base layer, the biggest alteration is related to the hyper-elements we have identified in the present video game screen. Hyperlinks are usually bound to enablement, as they help the game user find out more about different pieces of information tied to the club in question or move to screens that will allow them to make changes in the virtual world. Hyper-elements are also likely to play a crucial role in the navigation layer, which we will now discuss very briefly. When it comes to video games, the navigation layer is likely to be linked to the screen's intuitiveness; the navigation layer presented in the screen marked up in Figure 1 would probably require a period of adaptation when encountered by a new game player, although to regular game users it already seems intuitive enough. The screen itself has been changed over several previous game installments, which might indicate that its optimal detail level and overall outlook are yet to be reached, i.e., the navigation layer has to get in line with the users' expectations. This is where combining a GeM-based analysis with other approaches could be beneficial, which we will touch upon in the conclusions.

4 Conclusions and Future Directions

On the whole, before we come to our own conclusions, we have to state that we strongly believe that the current perspectives used in multimodality are likely to benefit from handling video games' multimodality. There are at least two reasons why this should happen. First, as we could see, video games can combine and em-

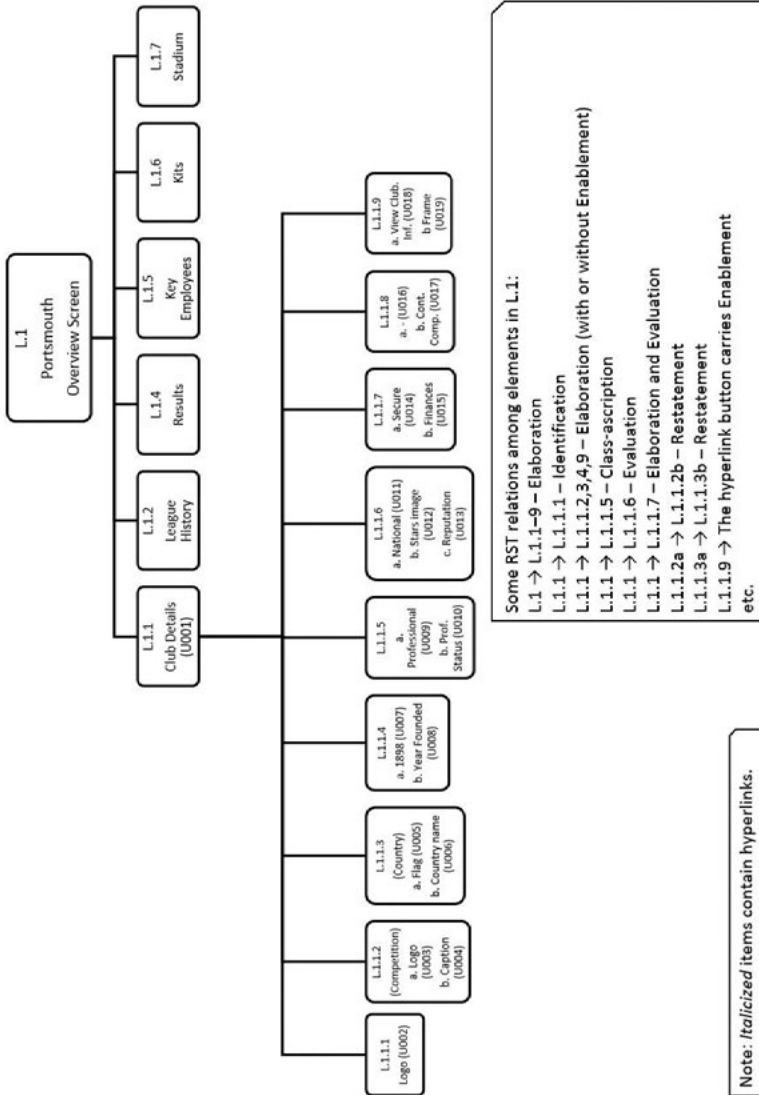


Fig. 3: The layout structure of the club overview screen in *Football Manager 2018* (presented in Table 1).

ploy modalities, and communicate with their users in specific (and usually diverse) ways; along with this, we have also seen that in them multimodality intersects with interactivity, and this intersection establishes a new research frontier.

The second reason lies in the very importance of video games. This can be seen not only from the fact that the video game industry overtook Hollywood nearly two decades ago, but also with regard to the circumstances in which video games go beyond performing an entertainment function. Given their popularity and omnipresence in the modern world, we can see how their power to affect public opinion expands (Coomes, 2004; Greitemeyer & Osswald, 2010). Moreover, they can also be used as training and propaganda tools (see Machin & van Leeuwen, 2007, 74–104), and multimodality needs to embrace their growing relevance and account for the aspects of video games it can address. The present chapter seems to be in line with the claim that this process can be “equally beneficial for game studies, in which multimodality may be considered from the viewpoint of human-computer interaction” (Bateman et al., 2017, 366), but the benefits extend in several different directions, which we now present in the closing paragraphs.

If we are to follow Kristine Jørgensen’s understanding of gameworlds as discrete virtual environments for ludic activity, with several layers and modes of signification, it would seem that interface elements in video games offer interesting challenges for the discipline of multimodality. Given that they incorporate both traditional interface elements found in other digital artifacts, as well as integrated, iconic, and metaphorical interface features, video games require special care when it comes to semiotic analyses, but conversely they also represent a potentially fruitful domain of exploration for multimodality.

Multimodal analyses of video games, such as the one presented here and utilizing the GeM model, help explain how different channels of signification operate in discrete instances of a particular game, such as, for example, in a single screen of a simulation game like *Football Manager*. In turn, this would facilitate discussions on the rhetorical potential of gameworlds and the signification elements therein across games belonging to the same genre or type, helping to explain the conventionalization of gameworld interface features discussed by Jørgensen.

With that in mind, it would seem particularly useful to move beyond the sole focus on interface features as understood and analyzed within the broader domain of human-computer interaction, and to instead include examinations of other semiotic modes regularly utilized in video games, such as haptic feedback. When it comes to analyzing document-like video game screens, the required modifications of the original Bateman’s design and Hiippala’s additions seem to be minimal, which goes along the lines of the requirement for the analytic schemas to operate without specific commitments drawn from individual examples (Tseng & Bateman, 2012).

Of course, the genre itself does bring some additions, but one can assume that these would be applied to most video games that contain document-like screens (they are class-specific rather than case-specific). Although the process might be time-consuming (even if one part of it is automated), the GeM model can be used for the analysis of virtual documents, but its potential is yet to be confirmed by further applications involving a wider range of document-like screens. Yet, as far as the screens addressed in the present chapter are concerned, we have seen that the GeM model in its present form can cover a large number of elements we found in a document-like video game screen, but we could also see what possible extensions of the GeM model could be, given the context of video gaming documents.

Making the model more applicable in this domain would probably lead to including a range of dynamic and/or interactive elements, such as different forms of clickable items in a video game's materiality (e.g., drop-down menus, sliders, drag-and-drop objects), as well as elements that change form when a user interacts with them (such as the elicited hypertext boxes mentioned in the chapter). Other possible additions might emerge from applying the model to different video game genres that contain document-like screens, but it is likely that they would be adaptable to the existing framework. Most (or even all) of them belong to the world of human-computer interaction and can be imported into the GeM model.

The application of the model leads to detailed results, which give us a comprehensive insight into the structure of the screen. In order to make such data useful, we suggest combining them with other sets of data. For instance, if we build a corpus consisting of the most relevant screens from several consecutive video games (from the same series), and correlate the corpus-based data with the bug report counts and complaint statistics (or even with the sales figures), we could find out more about how different screen designs affect the quality of different releases. If we opt for a diachronic approach of this kind, it could also allow developers to track different screens across installments and perhaps measure their efficiency, acceptability, and intuitiveness. Furthermore, trying to combine a corpus-based approach (be it synchronic or diachronic) with the studies that would involve gamers as respondents could also lead us towards improving the structure of the key document-like screens in video games similar to those described in this chapter.

Moreover, applying an analytical approach is also likely to reveal potential flaws in the current screen setups, which sometimes and in some cases can appear counterintuitive. If we combine such analyses with ideas coming from the domain of document design (e.g., Black et al., 2017; Carliner et al., 2006; Schriver, 1997), the possibilities seem to expand. If we take into consideration the GeM model's potentials in comparing genres, another use would be comparing different kinds (genres) of video games and their presentational strategies (e.g., we could see

similarities and differences between simulation games and strategies, or different subtypes of simulation games, such as sports managerial simulation games, construction simulation games, or life simulations).

Such comparisons could lead to a sort of ‘communication’ among genres and perhaps to ‘transfers’ of efficient screen designs from one genre into another. Altogether, the process could be of interest to game developers, as it could lead towards improving game screens in general. Finally, going beyond the scope of video games, we propose that a similar approach might be beneficial to studying real-life documents which are also, in this era of digitization, increasingly interactive, thus making the general approach even more important.

To conclude, we hope in this chapter to have shown why video games need to be taken as lying on multimodality’s path to becoming a full-fledged discipline. Throughout this chapter we have considered multimodality as a discipline for one main reason: in section 3 of the chapter we were able to perform a video game analysis by applying a model that does not belong to game studies at all. We are aware of the fact that this analysis was partial and that it mainly focused on modalities present in one screen, but it nevertheless demonstrated that multimodality has already developed its own tools that are applicable to media for which they have not been primarily designed.

This then shows the strength of the tools themselves. Wherever we encounter multimodal phenomena, the discipline of multimodality should be able to deal with them in a way that will sometimes overlap with other disciplines, but that will be consistent across media. This consistency can be guaranteed by the tools it employs, and the tools should be able to operate in a way that will not be tightly bound to any particular medium, let alone committed to individual examples within different media. More importantly, such tools should lead to results that reveal something more than that achieved in existing studies: that is, the multimodal approach itself supports a deeper understanding of the objects under study.

Our contribution in the present chapter is then to have shown that the GeM model is precisely one of those tools. In our case, the information on different layers and their structure gathered by applying the GeM model was such that it seemed likely that it could be combined with findings coming from other disciplines (game studies included), while also adding to such studies and those studies’ ability to characterize their objects of study. Such joint efforts should definitely improve various aspects of video games. Although multimodality might not be able to do it all alone, it would definitely be one of the partners in the process, alongside other relevant disciplines.

Bibliography

- Bateman, John A. 2008. *Multimodality and Genre: A Foundation for the Systematic Analysis of Multimodal Documents*. Basingstoke: Palgrave Macmillan.
- Bateman, John A. 2013. Multimodal Analysis of Film within the GeM Framework. *Revista Ilha do Desterro: A Journal of English Language, Literature in English and Cultural Studies* 64. 49–84. <https://doi.org/10.5007/2175-8026.2013n64p49>.
- Bateman, John A. 2014. Developing a GeM (Genre and Multimodality) Model. In Sigrid Norris & Carmen Daniela Maier (eds.), *Interactions, Images and Texts: A Reader in Multimodality*, 25–36. Berlin: Mouton de Gruyter.
- Bateman, John A., Judy L. Delin & Renate Henschel. 2002. A Brief Introduction to the GEM Annotation Schema for Complex Document Layout. In Graham Wilcock, Nancy Ide & Laurent Romary (eds.), *Proceedings of the 2nd Workshop on NLP and XML (NLPXML-2002) — Post-Conference Workshop of the 19th International Conference on Computational Linguistics (COLING-2002)*, 13–20. Academica Sinica, Taipei, Taiwan: Association of Computational Linguistics and Chinese Language Processing.
- Bateman, John A., Janina Wildfeuer & Tuomo Hiippala. 2017. *Multimodality – Foundations, Research and Analysis. A Problem-Oriented Introduction*. Berlin: Mouton de Gruyter.
- Black, Alison, Paul Luna, Ole Lund & Sue Walker. 2017. *Information Design: Research and Practice*. London: Routledge.
- Carliner, Saul, Jan Piet Verckens & Cathy de Waele (eds.). 2006. *Information and Document Design: Varieties on Recent Research*. Amsterdam: John Benjamins.
- Caroux, Loïc, Katherine Isbister, Ludovic Le Bigot & Nicolas Vibert. 2015. Player-Video Game Interaction: A Systematic Review of Current Concepts. *Computers in Human Behavior* 48. 366–381.
- Consalvo, Mia & Nathan Dutton. 2006. Game Analysis: Developing a Methodological Toolkit for the Qualitative Study of Games. *Game Studies. International Journal of Computer Game Research* 6(1). http://gamestudies.org/0601/articles/consalvo_dutton. Last accessed: 17 September 2019.
- Coomes, Michael D. 2004. Understanding the Historical and Cultural Influences that Shape Generations. *New Directions for Student Services* 106. 17–31.
- Firaxis Games. 2010. *Sid Meier's Civilization V*. 2K Games.
- Greitemeyer, Tobias & Silvia Osswald. 2010. Effects of Prosocial Video Games on Prosocial Behavior. *Journal of Personality and Social Psychology* 98. 211–221.
- Hiippala, Tuomo. 2015. *The Structure of Multimodal Documents: An Empirical Approach*. London: Routledge.
- Hiippala, Tuomo. 2017. An Overview of Research within the Genre and Multimodality Framework. *Discourse, Context and Media* 20. 276–284. <https://doi.org/10.1016/j.dcm.2017.05.004>.
- Jørgensen, Kristine. 2012. Between the Game System and the Fictional World: A Study of Computer Game Interfaces. *Games and Culture* 7. 142–163.
- Jørgensen, Kristine. 2013. *Gameworld Interfaces*. Cambridge, MA: MIT Press.
- Klastrup, Lisbeth. 2003. *Towards A Poetics of Virtual Worlds – Multi-User Textuality and The Emergence of Story*: IT University of Copenhagen dissertation.
- Klastrup, Lisbeth. 2010. Understanding Online (Game)Worlds. In Jeremy Hunsinger, Lisbeth Klastrup & Matthew Allen (eds.), *The International Handbook of Internet Research*, 287–302. Dordrecht and Heidelberg and London and New York: Springer Verlag.

- Konami Computer Entertainment Japan. 2004. *Metal Gear Solid 3: Snake Eater*. Konami Corporation.
- Machin, David & Theo van Leeuwen. 2007. *Global Media Discourse: A Critical Introduction*. London: Routledge.
- Paradox Development Studio. 2013a. *Europa Universalis IV*. Paradox Interactive.
- Paradox Development Studio. 2013b. *Uncharted 4: A Thief's End*. Sony Computer Entertainment.
- Schell, Jesse. 2008. *The Art of Game Design: A Book of Lenses*. Amsterdam: Elsevier.
- Schriver, Karen A. 1997. *Dynamics in Document Design: Creating Texts for Readers*. New York: John Wiley and Sons.
- Smith, Barry. 2014. Document Acts. In Anita Konzelmann-Ziv & Hans-Bernhard Schmid (eds.), *Institutions, Emotions, and Group Agents: Contributions to Social Ontology*, 19–31. Dordrecht and Heidelberg and London and New York: Springer Verlag.
- Sports Interactive. 2017. *Football Manager 2018*. Sega Europe.
- Square. 1999. *Final Fantasy VIII*. Square.
- Stamenković, D., M. Jačević & Janina Wildfeuer. 2017. The Persuasive Aims of Metal Gear Solid: A Discourse Theoretical Approach to the Study of Argumentation in Video Games. *Discourse, Context and Media* 15. 11–23. <https://doi.org/10.1016/j.dcm.2016.12.002>.
- Taboada, María Teresa & William C. Mann. 2006. Rhetorical Structure Theory: Looking Back and Moving Ahead. *Discourse Studies* 8(3). 423–459.
- Thomas, Martin. 2007. Querying Multimodal Annotation: A Concordancer for GeM. In *Proceedings of the Linguistic Annotation Workshop*, 57–60. Prague: Association for Computational Linguistics.
- Thomas, Martin. 2009. *Localizing Pack Messages: A Framework for Corpus-Based Cross-Cultural Multimodal Analysis*. Leeds: Centre for Translation Studies, University of Leeds PhD Thesis.
- Thomas, Martin. 2014. Evidence and Circularity in Multimodal Discourse Analysis. *Visual Communication* 13(2). 163–189. <https://doi.org/10.1177/1470357213516725>.
- Tseng, Chiao & John A. Bateman. 2012. Multimodal Narrative Construction in Christopher Nolan's *Memento*: A Description of Method. *Journal of Visual Communication* 11(1). 91–119. <https://doi.org/10.1177/1470357211424691>.
- Wolf, Mark J.P. 2008. *The Video Game Explosion: A History from Pong to Playstation and Beyond*. Westport: Greenwood Press.



Part IV: More Disciplinary Thoughts

John A. Bateman

Afterword: Legitimizing Multimodality

Abstract: This afterword draws insights and conclusions from the preceding chapters by critically engaging once again with the disciplinary status of multimodality. It explicates the main points of discussion of the contributions and makes some recommendations concerning disciplinarity and multimodality. The path taken addresses multimodality at a fundamental philosophical and practical level, reflecting some of the requirements that establishing a discipline of multimodality would entail. This presents multimodality and its study increasingly in a particular light of its own and, as a consequence, it will be argued that a strengthening of disciplinary claims for multimodality is at this time not only beneficial but, in certain important respects, crucial for advancing beyond the current, somewhat disparate, state(s) of the art.

Keywords: multimodality, disciplines, horizontal knowledge, hierarchical knowledge, Legitimation Code Theory

1 Disciplinary Re-Considerations

In this volume, we have collected together several contributions from participants at our previous BreMM conference on multimodality and framed these and others with respect to the ‘discipline question’, i.e., whether or not it would be beneficial to consider multimodality as a discipline in its own right. For some readers, this entire discussion concerning disciplinarity and the potential benefits of setting up multimodality along disciplinary lines may appear odd: surely, one might think, interdisciplinarity and transdisciplinary research, focusing on themes and fields, rather than disciplines, is the ‘modern way’. And indeed, as noted by Christie and Maton, there is still (or again) a rather common view in which ‘disciplinarity’ is seen as “reactionary and conservative, while ‘interdisciplinarity’ is viewed as progressive and egalitarian” (Christie & Maton, 2011, 1)—an attitude also reported on by Krishnan, who we cited in the Introduction, thus:

According to this new orthodoxy, scientists should aim to develop fruitful relationships to other disciplines than their own and perhaps even to transcend disciplinary thinking altogether. (Krishnan, 2009, 5)

<https://doi.org/10.1515/9783110608694-013>

As Krishnan notes, however, the basic notion of a ‘discipline’ presumed in much discourse of this kind is often left sufficiently vague as to preempt further serious discussion.

This is important because, working within the sociological framework of Legitimation Code Theory (LCT), Christie and Maton argue that claims concerning the outdatedness of the very notion of disciplines, and consequently of working within disciplines, are not only far from accurate, they are also positively harmful. Indeed, working with such assumptions continues to have negative impacts in several fields, multimodality included. This alone warrants a more careful consideration of the relationships between the field of multimodality and its possible incarnation as a discipline. Whereas in our Introduction to this volume we made comparisons between multimodality as a field and other areas, particularly the Digital Humanities, that have similarly found themselves in the situation of drawing on diverse disciplinary roots to form what are, generally, now considered to be their own disciplines, in this chapter I engage with the notion of disciplinarity ‘as such’ by drawing on several theoretical constructs applied by Christie, Maton and colleagues’ theoretical engagement with disciplinarity within LCT.

To begin, I first briefly situate and motivate the use of LCT for our current purposes and set out those of the theory’s constructs that will be relevant for the meta-theoretical analysis to follow. I then discuss the contributions to the volume in these terms, before turning to characterize some concerns with the current status of multimodality as illuminated from the perspective of LCT more broadly. Finally, I conclude with some observations on the place of multimodality and its future development.

2 Knowledge Structures and the ‘Discursive Gap’

In several approaches to multimodality, particularly but not only those drawing on social semiotic or interactionist perspectives, the construction of ‘knowledge’, i.e., any understanding of the world and our places in it, is seen as essentially socially mediated. Knowledge is consequently treated as a social process anchored in social practices, predominantly *discoursal* practices of communication (cf. Bateman et al., 2017, 64–66). This contrasts with, or rather complements, views of ‘knowledge’ that take a more cognitive or individual perspective; even the acquisition of individual knowledge is considered as arising from social contexts of interaction. Whereas this view can be taken to apply for all kinds of knowledge, significant work by Foucault and others placed emphasis on the role that social *institutions* play in knowledge creation, distribution and maintenance. One prominent social

institution for managing knowledge is, of course, education, and so there are prominent proponents of accounts of the sociology of knowledge, LCT included, that focus particularly on education and its functioning as a social practice.

An orientation of this kind also holds for the entire field of systemic-functional linguistics, out of which several prominent accounts of multimodality have arisen (cf. Jewitt et al. 2016 and several of the contributions to this volume). Interactions between sociology, and particularly the sociology of knowledge, on the one hand, and systemic functional linguistics on the other, consequently go back to the beginnings of the theory. This is most evident in the cooperation between the systemic-functional linguist Ruqaiya Hasan and the sociologist Basil Bernstein in the 1960s; this work addressed the relationship between variations in language and variations in social structure with a clear focus on children's early phases of socialization (e.g., Hasan, 1973). Facets of the assumed tight relationship between the deployment of semiotic resources and social configurations remain in almost all versions of the social semiotic account to date, including its work on multimodality.

A further related concern of Bernstein was the social institution of disciplines, which also clearly play a major role in establishing and distributing 'knowledges' of various kinds, both for education and in society more broadly (Bernstein, 2000). More recently, these issues have been taken up within the developing field of Legitimation Code Theory (Maton, 2014), which combines elements drawn both from Bernstein and from theories of the sociologist and philosopher Pierre Bourdieu to engage more deeply with the dynamics and structures of knowledge, always seen as a social process. As mentioned above, there are now some quite explicit considerations of the nature of 'disciplines' developed within LCT; building on these as 'meta-theorizations' of the distinct kinds of knowledges that disciplines and their practices enact will now allow a closer consideration of just what we should expect a discipline to provide for its practicing participants. It will also permit further beneficial triangulations to be carried out concerning the current state of multimodality. This then takes a functional view on 'disciplinarity': what is it that disciplines actually *do* and could accounts of this also be used to characterize more precisely the current state of the theory and practice of multimodality?

Arguably one of the most essential contributions of practice within a discipline, and so a central challenge facing any discipline, is to serve an enabling function for the *construction of knowledge over time* for that discipline's community, addressing strategies by which this construction can best be achieved and diagnosing problems that might hinder this. This involves two tasks: finding appropriate organizational forms for the knowledge itself, and finding effective ways for statements made within that discipline's theoretical framework(s) to be related to its objects of concern, whatever those may be. That is: it is not enough simply to have a sophisticated theoretical language for disciplinary reflection, that language

must also engage with something ‘outside’ the theoretical descriptions produced, i.e., it must engage with the body of phenomena about which a theory is intended to be a theory. Explicitly considering both of these tasks in relation to the current state of multimodality will be shown below to be highly enlightening.

First, concerning forms of knowledge, Bernstein, and later Maton within LCT, examined in detail distinct disciplinary forms of knowledge and knowledge-building activities with a view to assessing their effectiveness for advancing knowledge. These discipline-based forms of knowledge are characterized as *internal* conceptual languages, of frameworks of knowledge, that the discipline-as-institution uses for its own activities and practices and promotes in its education. Internal languages in this sense can be seen as the forms of knowledge constitutive for a discipline, i.e., the categories, structures, and conceptual relationships with which theories within the discipline are expressed and developed. Such organizations include both the specific terms and relations between terms practiced in a discipline and any more generalized structuring systems (metaphors, paradigms, and so on) that provide a sense-making background for those terms and relations.

Bernstein then proposed distinguishing between two broad kinds of organization that such internal languages and their respective forms of knowledge exhibit. Those knowledges may, on the one hand, be tightly integrated, relational—i.e., ‘hierarchical’—or, on the other hand, be more associative and segmented—i.e., loosely ‘horizontal’. Hierarchically-organized knowledge relies substantially on explicitly interrelated classes and categories and their respective distinguishing properties; horizontally-organized knowledge relies in contrast on more networked associative connections. Maton consequently contrasts these forms of knowledge thus:

Hierarchical knowledge structures, exemplified by the natural sciences, are explicit, coherent, systematically principled and hierarchical organizations of knowledge that develop through the integration and subsumption of existing knowledge. Horizontal knowledge structures, exemplified by the humanities and social sciences, are a series of strongly bounded approaches that develop by adding another approach alongside existing approaches. (Maton, 2011, 63)

The link that Maton draws here to exemplifying disciplines is important and central to what follows below.

Second, moving beyond ‘internal’ languages, Bernstein argued that it was also essential to consider quite explicitly how the conceptual terms and organizations of a discipline are to be related to the situations and phenomena about which knowledge is to be constructed or achieved. These are two, inherently distinct ontological realms and give rise to what Bernstein (2000, 29) came to term the *discursive gap* (cf., e.g., Moore & Muller, 2002):

All research involves [...] a 'discursive gap' between theory and data, but frameworks differ in whether and how this gap is traversed. Most fail to either recognize or overcome this gap—they may have a powerful and persuasive internal conceptual language but reduced powers to provide externally unambiguous descriptions of the capacity for building epistemologically powerful knowledge. (Maton & Chen, 2016, 29)

Bernstein made this a central focus in its own right, observing that within many disciplines the mechanisms by which the two realms are to be related are left implicit or unthematized. Explicit engagement with the discursive gap is, however, an essential precondition for effective knowledge-building. To support this, he consequently drew a further meta-theoretical distinction between a discipline's internal conceptual language (whose discourses may be horizontal or hierarchical as just indicated) and 'external' *languages of description* that serve precisely the purpose of organizing objects of investigation, i.e., data, in ways that make those objects accessible to analysis.

Crucially, these external languages are ways of characterizing data *without already enforcing internal theoretical distinctions on that data*, since this may well turn out on subsequent investigation to be inappropriate or premature. As Maton and Chen continue:

Reinforcing this problem is a tendency [...] to portray explicit means of enacting concepts as imposing theory onto data in a 'cookie-cutter' model which ignores the particularities of objects of study. While such an approach would indeed be deaf to data, so is denial of the discursive gap. Failure to recognize that relations between theory and data are not immediate or unproblematic but rather require an explicit means of translation typically leads to theory becoming deaf to data, for nothing seems to fall outside of the theory. (Maton & Chen, 2016, 29)

As a precursor to the discussion to be taken up below, it should be noted that this is also a rather exact characterization of several critiques that have now been brought against multimodality research (cf., e.g., Forceville, 1999, 2007; Bateman et al., 2004; Ledin & Machin, 2019).

The recognition of external languages of description is intended to address this problem directly, allowing characterizations of data always to be able to go beyond what might be predicted by a theory and so supporting the finding of counter-examples or new phenomena. Only in this way can a theory be stretched and forced to extend to accommodate new empirical results. Simultaneously, however, the internal language of a theory or discipline should also be able to 'generate' combinations and categories independently of the data, i.e., to make predictions about what can occur, not only describing what has already been observed to occur.

Bernstein described this ‘generative’ capability, at both the levels of internal and external languages of description, in terms of ‘grammar’: a ‘strong grammar’ characterized a language of description (whether internal or external) capable of generating configurations, a ‘weak grammar’ characterized languages of description (again whether internal or external) as less able to generate novel configurations. Correlations between this notion of ‘grammar’ and the kinds of knowledges constructed are also readily observed; as, for example, Moore and Muller set out with respect to the sociology of education:

Sociology of education is thus a horizontal knowledge structure with a weak grammar, with a conceptual syntax not capable of generating unambiguously precise empirical descriptions. And because this grammar cannot relate empirical descriptions non-contentiously, empirical description cannot arbitrate conceptual disputes. [...] there is no generally accepted means for clearing out old superannuated theories that begin to clutter the literature. (Moore & Muller, 2002, 630)

Again, comparisons with much of the state of the art in multimodality are striking.

Finally for our current purposes, Maton subsequently develops these ideas further within LCT, expanding the external languages of description across three levels of abstraction: *data instruments* and two kinds of *translation devices*: *mediating languages* (i.e., non-specific external languages, translating between theory and all empirical forms of a phenomenon) and specific *external languages* (translating between theory and empirical data within a specific problem situation). Together, these concepts support explicit engagement with a range of concerns faced by all disciplines and can be used for diagnosing points of methodological difficulty and opportunity with particular clarity.

The function of data instruments is to make explicit the first step towards empirical data in a manner open to the specificities of that data but still oriented reflectively towards the conceptual concerns established by theory; they

provide a methodological guide to a project by delineating how concepts suggest foci for data collection and questions for analysis. They make explicit the movement *from* theory *towards* data. (Maton & Chen, 2016, 30)

Data instruments consequently concern more the process than the product of research and can be related to other methodological steps developed for systematic analysis in other fields, such as the ‘codes’ and ‘concepts’ stages of Grounded Theory (cf., e.g., Glaser & Strauss, 1967; Strauss & Corbin, 1990; Kelle, 2005), or the criteria identified for defining effective coding schemes for content analysis (e.g., Schreier, 2012; Krippendorff, 2004). Thus, even though the external languages do not impose theoretical distinctions directly on data (the “cookie-cutter” syndrome), they by no means assume that data should (or can) be approached as if the analysis

is theory-free; data will always be being viewed from a particular disciplinary perspective. It is then the task of a discipline's translation devices to fill out the methodological and practical steps that negotiate appropriate relations between theory and data—that is, to establish relations that do not simply impose theoretical categories on data but which instead are open to the specificities of data without losing sight of the more general conceptual goals and frameworks of a discipline or theory.

This can be seen as considering the notion of 'applying' a theory at a rather more detailed level. Rather than assuming that theoretical categories may be applied 'directly' to data, the framework makes it clear that several important methodological steps need to intervene. To apply to data at all, categories must be operationalized so that they are reliably recoverable; but premature operationalization may lead to patterns in the data being missed. Relating the respective levels of abstraction involved, Bernstein characterized the internal conceptual language of a discipline as L^1 and external languages of description as L^2 . Maton consequently proposes that mediating languages receive the label $L^{1.5}$, showing their intermediate status on the way between theory and empirical data (Maton & Chen, 2016, 30–31); I will follow this notational scheme below. Distinguishing the three levels of abstraction as proposed so that they can be considered 'independently' of one another significantly increases the chances that investigations of data will be able both to test theoretical claims and to show when those theoretical claims are in need of revision.

Following the directions established by its origins, LCT has to date been applied predominantly to pedagogy and education: two areas that are very much in need not only of methodologies for investigation, but also of general principles that can guide how to conduct research that both engages with specifics and which allows more general conclusions to be drawn. Here, being able to bridge theory and data is crucial. However, this issue is a particularly telling issue for *all* disciplines and research fields that operate in a more 'discursive' manner. The valuable service provided by LCT is then to offer ways of moving forward, of 'traversing' the discursive gap, even for disciplines that are traditionally more wary of 'data' and empirical studies (Maton & Howard, 2016). This will lead us below to several further interesting analogies between LCT's proposals for dealing with such problems and approaches currently being taken to multimodality. Most specifically, we will see a clear parallelism with those brands of multimodality that are also more discursively oriented. In fact, many of the problems that LCT suggests for pedagogy and education will be argued to closely echo critiques that now need to be made of discursive multimodality as well.

3 The Contributions to this Volume Re-Viewed

The framework given in the previous section now allows us to construct a further triangulation point outside of multimodality for the individual contributions to the volume. Almost all of the contributions make reference to data and its empirical analysis and so one way of providing a comparative review is to consider just how they spread their assumptions, activities, and results across the continuum of abstraction from internal conceptual languages (L^1) through to data descriptions (L^2). In addition, all of the contributions also make references to ‘other’ disciplines with which they interact or draw conceptual frameworks from; this then suggests further characterizations of at least the L^1 -components of their accounts. It is interesting to consider just how these diverse disciplinary inputs are intended to combine. In all of the descriptions that follow, readers are referred to the individual chapters for specific references to any of the constructs or authors named.

3.1 Building Theories and Bridging with Data

As would be expected, all three chapters of Part II’s ‘Disciplinary Thoughts’ address issues concerned primarily with L^1 - $L^{1.5}$ - L^2 interrelationships. This begins with **Stöckl**’s explicit review of the current state of the L^1 of multimodality, in which he sets out those terms most commonly found in theoretical discussions of multimodality and seeks connections between them. However, despite the fact that many of these core terms reoccur across different approaches, it is by no means straightforward to relate them. Stöckl suggests that one reason for this might be the lack of maturity of the field; connecting terms remains necessary but there is evidently an “insufficiently integrated theory” at work. This can also be seen, therefore, as pointing rather directly to a broadly ‘horizontally’-organized L^1 for multimodality rather than an L^1 that exhibits a more ‘hierarchical’ integrated internal organization. In this particular contribution, Stöckl does not specifically address L^2 issues, although elsewhere he has extensively illustrated the use of corpora for multimodality research as well as drawing attention to the importance of the use of genre (cf., e.g., Stöckl, 2004a). The chapter does, however, explicitly suggest additional sources of L^1 conceptual languages that should be brought into contact with multimodality—namely multimodal rhetoric, cognitive semiotics, and transtextuality, each with their own particular ways of relating to their objects of concern (i.e., $L^{1.5}$ and L^2). How precisely this interdisciplinarity is to be configured remains a central question.

The chapter by **Thomas** moves on to consider issues that arise for the relation between L^1 and, in particular, $L^{1.5}$, if the empirical basis of multimodality is to be broadened. Thomas argues that multimodality needs to engage with larger-scale analyses of data in order to achieve more robust and inclusive descriptions. He does not, however, place any particular restrictions on which L^1 is being assumed for multimodality in such work, although import relations are clearly recommended for corpus linguistics and computational visual processing – both perhaps better situated as $L^{1.5}$ contributions as they impact most strongly on method. At this point in time, Thomas sees the avoidance of potentially premature theoretical categorization as a more important issue and so is quite willing to proceed without prestructuring data according to categories such as semiotic modes, an issue we will return to below.

The final chapter of Part II by **O'Halloran and colleagues** takes a further, quite different perspective on the L^1 of multimodality. Rather than proposing combinations of disciplines, they suggest instead that the conceptual language of multimodality should itself be seen as serving a similar structuring role across disciplines as that played by the conceptual languages of mathematics and linguistics. In other words, they argue at a meta-disciplinary level that there are strong parallels to be drawn between the use made by *other* disciplines of the L^1 – $L^{1.5}$ configurations of mathematics and of linguistics and what could be done with an appropriate L^1 – $L^{1.5}$ configuration for multimodality as well. O'Halloran and colleagues are also quite specific about the L^1 – $L^{1.5}$ they wish to adopt, broadly following the theoretical dimensions of systemic-functional linguistics and including metafunctionally-organized semiotic resources as central. This raises a recurring theme for the discussion below as we consider to what extent approaches assume (explicitly or implicitly) the adoption of an L^1 to have *already* addressed the L^2 issue—something which LCT would question. The relation to L^2 discussed in this contribution is also interesting in its proposal, similar in some respects to that of Thomas, for the application of computational methods for automatic feature extraction in order to avoid the theory losing touch with what it is intended to describe and explain, i.e., in the terms introduced here, failing to adequately address the discursive gap.

The respective contributions in Part III of the book are then more specific in their analytic targets, although the contribution by **Schmidt and Marx** on a particular genre of YouTube videos is still nevertheless the chapter that is most explicit both in its consideration of distinct L^1 and L^2 characterizations and particular $L^{1.5}$ methods for bridging between them. This is in large part due to the anchoring of the account within linguistics and, more specifically, empirical corpus linguistics, where these methodological issues have long received close attention. Schmidt and Marx also usefully extend their access to the object of analysis by importing

Goffman's notion of participation frameworks, which similarly makes contact both to L^1 -notions of social organization and L^2 (or mediating $L^{1.5}$) notions of how to organize data being investigated. One can then characterize the framework as: (i) importing from both corpus linguistics and interaction frameworks giving rise to an L^1 organized in terms of interactional pragmatics, forms of communication (medium) and communicative uses (genres), (ii) employing interactional pragmatics to drive the data instrument that selects what is to be attended to in the collected data, and (iii) generating a mediating $L^{1.5}$ in terms of verbal transcriptions and further corpus annotation schemes for accompanying 'nonverbal' properties of the interactions studied. Many of the components at work here are already strongly hierarchically organized, involving well-developed integrated conceptual frameworks.

O'Hagan's chapter then explicitly calls for interaction at the L^1 -level between multimodality and the field of ethnohistorical studies, including that field's notions of archival research and attention to the fine-grained social contexts of any objects of study. Conceptions of power, class, and identity are clearly strong elements of the imported L^1 , which also then resonate well with O'Hagan's stated aim of providing more evidence for Kress's contention that semiotic resources differ across cultures (and, thus by implication, across times and places as well). This also drives the data instrument that O'Hagan applies: finding varied book inscriptions and characterizing them in terms of their materiality. The bridge between the objects of analysis and the L^1 -categories is formed by an L^2 including identification of colors, image contents, and positioning and a mediating $L^{1.5}$ drawing on van Leeuwen's accounts of color and typography. As O'Hagan notes, however, the bridge is incomplete as the detailed analyses do not yet provide sufficient engagement with Kress's conjecture to deliver robust conclusions. This kind of problem is, in fact, rather common: in the terms introduced here it can be taken as indicative of a disconnect between the various levels of abstraction ranging from L^1 to L^2 , from theory to data, and the way that that data has so far been made accessible to study.

The chapter by **Harnett** on graphic novels also takes an expanded L^1 , presupposing imports of Groensteen's notion of braiding in visual narrative, discourse-oriented interpretative critiques, and accounts of 'agentive readers' and narrative functions. The target of the investigation is to reveal some of the narrative and aesthetic consequences of aligning semiotic modes in various ways. The data instrument adopted is thus the selection of graphic novels that are well recognized as being challenging for interpreters. The L^2 adopted is one of identifying visual properties, color, shapes, texture, visual motifs and similar so that these can be discussed in relation to hypothesized narrative effects. The bridge between these levels of abstraction is managed discursively, indicative of a broadly horizontal L^1 organization.

Taylor's chapter adopts a fairly horizontally-organized L^1 for multimodality as well, importing questions and approaches from the broad field of audiovisual description. The data instrument applied is to seek situations where individuals with sight-related difficulties can be supported by audiovisual descriptions. The data is then discussed with respect to broadening the range of semiotic modes that may usefully be applied to go beyond the verbal.

The chapter by **Larkey** adopts as its L^1 the view of multimodality found in Kress, Jewitt, and others, focusing on modes and the specifics of mode combinations as indicative of social configurations, combined with aspects of narratology. Their data instrument is to seek places where semiotic resources are used in different cultures for narratively comparable situations, leading to a data analysis drawing on an L^2 ranging over verbal language, gestures, postures, and music. These are read in terms of relations of power, gender and family relations, so as to delineate differences in culture. As in the previous contribution, the bridge between levels is again primarily managed discursively, although the incorporation of more corpus-based methods is also suggested.

Nielsen and colleagues' chapter on educational issues already takes LCT as part of its toolset, although applied at a rather different level of abstraction to that being employed here. The L^1 adopted is similar to that assumed in O'Halloran and colleagues' contribution, drawing on systemic-functional notions of systems of choice that build systemic resources of image, language and music, combined additionally with notions of disciplinary knowledge and literacy. It is hypothesized (L^1) that characterizing differences in how semiotic resources express disciplinary knowledge will be beneficial for supporting critical reflection. As a data instrument the authors adopt the production of short multimodal explanation presentations, which are then analyzed both visually and linguistically. One form of analysis applied draws on LCT's notion of semantic density—essentially the relative degree of 'condensation' of knowledge within expressive forms—extended to apply to both images and language, giving one way of bridging between data and theory. However, another form of analysis reported draws on some of the proposals made in multimodal social semiotics for characterizing visuals, such as Kress and van Leeuwen's treatment of 'arrows' and similar visual connectives. This latter form of analysis is less clear concerning its relation to data: either a particular collection of L^1 constructs (e.g., "narrative processes") is also being read as providing $L^{1.5}$ -mediation categories, or mediation categories are being used whose precise relation to the L^1 conceptual framework and ways in which disciplinary knowledge is expressed remains in need of further explication.

Finally, **Stamenković and Jačević's** chapter on game screens as interfaces to gameworlds further suggests connections between the L^1 of multimodality and additional disciplines, most specifically the entire field of games studies. They do

not particularly address an L^1 of multimodality itself, however, focusing more on the organization of data for corresponding corpus studies and the extensions that the broadened materiality of game interfaces makes necessary when attempting to adopt an existing corpus annotation scheme developed for static page-based artifacts. This can be situated in LCT terms most plausibly as a contribution both to the mediating languages of description and the external L^2 . Annotation schemes intended to apply to a range of data sets can quite generally be seen as serving this mediating, $L^{1.5}$, function.

The allocation of various components of the frameworks and the analyses of the individual contributions set out here may not always correspond to how the authors themselves see their accounts—clearly more dialogue would be useful from this meta-theoretical perspective. It should, however, now be clear just how a more explicit orientation to the various levels of description to be bridged in work of this kind may assist in relating contrasting approaches and locating potential sources of commonality or difficulty. For example, and most specifically for the purposes of this volume, we can consider the characterizations offered with respect to just which positions the contributions to the volume adopt, if any, on the ‘discipline question’. The positions articulated vary considerably concerning the potential value of such a move: some approaches appear open to the idea; others are explicitly more skeptical. The next subsection explicitly relates this variation to differences in the filling out of the L^1 - $L^{1.5}$ - L^2 bridge.

3.2 Positions For or Against a Discipline of Multimodality

Independently of the specific contributions described above, there are several potential grounds for finding different evaluations of the ‘discipline’ question voiced with respect to multimodality. One simple reason for not considering a further, distinct discipline of multimodality would be that one is quite happy where one is and the overhead of participating in an additional discipline appears to be just that, i.e., pure overhead. Alternatively, rather more negative grounds for rejection might be a certain ‘fear of the other’ or a sometimes quite justified suspicion of disciplinary ‘imperialism’, whereby one discipline or field of activity is seen as encroaching on the territory of another, thus leading to problems of funding, of intellectual boundary transgressions, and the like. Muller (2011) relevantly distinguishes two forms of interdisciplinarity thus:

creative or productive interdisciplinarity occurs when it is driven by a *bona fide* intellectual problem which draws together a group of creative scholars who pool their disciplinary energies in one way or another. If this proves to be a stable basis for a new domain of problem

solving, it will morph into a new region, which if it proves stable and productive over time will finally become a discipline in its own right. (Muller, 2011, 23)

I will only be concerned with situations within which productive interdisciplinarity is taken as a basic condition for interaction; predatory interdisciplinarity will be rejected from the outset. None of the contributions to this volume presents predatory interdisciplinarity as an issue for them.

More considered reasons for not seeing multimodality as a discipline are then that the questions that one wants to ask and, for the most part, the methods for approaching those questions are already sufficient. Thus, if one is addressing primarily linguistic concerns, then a consideration of multimodality might be beneficial when language occurs in richer multimodal contexts of use but the main questions at issue remain linguistic and the primary methods, e.g., corpus methods, remain as well. Similarly, for those addressing painting, literature, film, or information design, similar anchorings in a ‘home’ discipline may be considered sufficient. This will particularly be the case when a home discipline already offers a hierarchical internal language with well developed ways of relating that language to data. Linguistics in most of its forms would be a prime example of such a discipline: not only are there extensive and tightly integrated theoretical frameworks, but there are equally finely developed methodologies for approaching actual instances of data as well.

For those who do consider a distinct discipline of multimodality as a valuable, perhaps necessary, step, therefore, there must then also be questions and, ideally methods, that are *not* already provided by established disciplines—which may then lead to productive interaction in the manner described by Muller above. This can be the case even if some of the subject matter or materials overlaps with subject matters and materials that have come within the purview of other disciplines. For horizontally-oriented disciplines, this can occur in a straightforward fashion as it does not require integration of already tightly interwoven knowledge structures—although, even though in principle ‘easier’, there are then fewer grounds for actually explicitly engaging with *integration* because the boundaries between the discourses involved are in any case relatively weak. Approaches couched within horizontally-organized L^1 languages might then see little benefit in an explicit adoption of multimodality as a discipline, since it is easier to place whatever is assumed for multimodality ‘alongside’ what is assumed from other disciplinary sources simply by re-using the terms. The fact that disciplinary boundaries might be less clearly defined cannot, however, be seen exclusively as a positive property. For a discipline to be effective, it must provide effective ways of building and accumulating knowledge but, as Maton sets out, not all styles of discourse are equal in this regard. Horizontally-organized discourses face considerably more challenges

when attempting to build knowledge by combining approaches and frameworks precisely due to their lack of tight definitions.

This returns again to the issues raised in Stöckl's contribution. Unclear boundaries raise problems for understanding what the terms employed might actually mean. Consequently, extending Stöckl's list of terms with terms drawn from other similarly loosely organized lexical fields from other disciplines would by no means be an immediately convincing research strategy as this just adds further (horizontal) associations: the 'vocabulary' of the resulting multi-discipline is extended, but without necessarily showing how the imported meanings combine. Finding genuine points of commonality and difference that might be resolved by reference to empirical studies is then difficult: terms are placed alongside one another rather than being forced to interact, which precludes deeper integration.

A considerable proportion of multimodality research now relies on horizontally-organized L^1 knowledge of this kind, even though earlier pioneering work in multimodality from the 1980s and 1990s took significant steps towards articulating detailed frameworks. Many of these, however, have since become increasingly divorced from empirical validation, leading to horizontal extension with terms standing 'side by side' rather than in hierarchical relations of depth and classification. Work where empirical validation does retain a strong driving force is, for example, where the connection to linguistic methodologies has remained strong, as illustrated in the contribution of Schmidt and Marx, or in the growing body of work conducted within psychology and the brain sciences on issues centrally related to multimodality (e.g., Raz & Hendler, 2014; Loschky et al., 2015; Cohn, 2016), as well as in the general appeals for stronger empirical foundations made in this volume by O'Halloran and Thomas.

One might then consider whether this is an appropriate place to stop development—in which case, multimodality might remain a field of inquiry as it is currently often configured. The consequences of it being considered a discipline or not would remain primarily institutional in nature, as set out in the introduction to the volume. The next subsection considers this issue rather more critically, however, again applying the notions from LCT of bridging theory and data via appropriate forms of disciplinary knowledge and suitably well articulated mediating languages of description. I will suggest that the current state of affairs is, in fact, *not* a suitable resting place.

4 Return to a Discipline of Multimodality

Explicit accounts within the sociology of knowledge, and how disciplines function within this (cf., e.g., Maton, 2014, 2016), offer a more useful place to bring the considerations offered in this volume concerning multimodality to a close, providing a good additional perspective on several of the issues and problems raised in the introduction and above. In this section, therefore, I consider in particular Maton's development and theoretical articulation of Bernstein's earlier notion of distinct kinds of organizations of knowledge, including disciplinary knowledge, explicitly with respect to the general state of affairs in multimodality. As suggested above, knowledge-building is clearly a core functionality to be supported by a discipline, but it is just this cumulative knowledge-building that is often currently constrained, or even derailed, among and between the diverse activities constituting multimodality. Applying the concepts introduced above from LCT will now allow these symptoms to be diagnosed more precisely.

First, let us consider again the *consequences* of the different kinds of internal languages, i.e., the forms of L^1 adopted. Martin characterizes this distinctly thus:

Hierarchical knowledge structures [...] test theories against data; horizontal knowledge structures use theory to interpret texts. (Martin, 2011, 42)

The current state of multimodality is distributed around these possibilities. Although often proclaiming to offer analyses, both the internal languages of description of the theoretical constructs and the mechanisms for relating those constructs to data are left underdeveloped. As noted above, this does not preclude such activities from being 'disciplines' as there are already many disciplines that this situation would characterize. It does, however, suggest that there may be problems in developing more systematic conceptual frameworks of knowledge to advance the field on the basis of empirical studies because the relationship to data is unclear. This is generally the case in the forms of multimodality that developed away from linguistics to constitute social semiotics (cf., e.g., Kress & van Leeuwen, 2001), where terms increasingly show properties of the horizontal disciplines of cultural studies.

Much of this legacy of horizontally-oriented discourses can, for example, be illustrated in the continued attempts to define, or to avoid defining, the core multimodality notion of *semiotic mode*: definitions here have changed little over the past 20 years, as have the accompanying statements that modes are difficult to define. Most of the initial tensions in considering multimodality at all are still to be observed at work. One strong symptom that there are indeed problems within the L^1 -conceptual organization revolving around terms such as 'semiotic mode' is

the prevalent reliance on examples rather than definitions. The definitions that are offered can lie at such a distance from the examples that it is difficult to relate the two in a productive fashion—that is, the L^1 cannot be used *generatively* (see above).

Taking, on the one hand, semiotic modes as the “work of culture in shaping material into resources for representation” and “regularised organised set of resources for meaning-making” (Jewitt & Kress, 2003, 1-2) no doubt describes what is occurring, but does little to provide workable identity criteria; the gap is then filled by examples, such as “image, writing, gesture, gaze, speech, posture” (Jewitt, 2014, 1) and so on. Listing examples can be useful as an indication of what kind of entities are intended, but it can equally well create problems when the examples taken have very different properties or theoretical statuses that make it far from obvious just what criteria are being employed to warrant inclusion.

This is particularly evident in the following introductory paragraph from an otherwise interesting and useful collection concerning TV commercials:

we can identify three main modes apart from the coded verbal language. Probably the most important, given the attention it gets in scholarly circles, is the visual mode made up of still and moving images. Another set of meanings reach us through our ears: music, diegetic and extra-diegetic sound, paralinguistic features of voice. The third is made up of the very structure of the ad, which subsumes or informs all other levels, denotes and connotes meaning, that is, lecture-type ads, montage, mini-dramas. (Pennock-Speck & del Saz-Rubio, 2013, 13–14)

Lists of this kind reflect the usual tension found between ‘perceptual’ and ‘semiotic’ perspectives when discussing semiotic modes, where different approaches draw their divisions in different ways. It is not possible from lists of this kind to deduce just what kind of thing a semiotic mode is and what kind of thing it is not: ‘visual’ and ‘aural’ are sensory channels, ‘verbal language’ is clearly semiotic, and ‘mini-dramas’ are, presumably, generic forms. Maintaining perceptual-semiotic splits of this kind commonly leads to the kinds of cross-classification problems discussed in detail by Stöckl (2004b).

Several authors have attempted to organize the terminological field at issue here more explicitly by acknowledging this distinction. Caple (2018), for example, follows O’Halloran in offering ‘multimodal’ as a term relating explicitly to sensory channels (“visual, aural”), and reserving ‘semiotic resource’ for the semiotic area. ‘Multimodality’, however, is then admitted in addition as a term covering both ‘multisemiotic’ artifacts/performances and ‘multimodal’ artifacts/performances. Fricke (2012, 47–48) analogously talks of multimodality ‘in a narrow sense’ (multiple sensory channels) and multimodality ‘in a broad sense’ (multiple semiotic systems, also potentially including only a single perceptual channel). It is then the

presence of *combinations* of modes that is most commonly taken as an indicator that phenomena relevant for ‘multimodality’ are at hand; Kress (2010, 28, 162) consequently added further terms, such as ‘modal ensembles’, that emphasize this. However, given that we still do not know what ‘semiotic systems’ there are, the situation remains relatively unclear: semiotic modes are ‘meaning-making systems’ appearing in combination, but particular properties and consequences do not appear to follow from this apart from, probably most prominent and widespread, the ‘cookie-cutter’ attribution of a still largely empirically unvalidated metafunctional organization (cf. Bateman, 2019),

There are several common responses to this state of affairs. Many authors simply proceed directly to analysis relying on what are, in the last resort, actually still ‘pre-technical’ categories such as ‘text’, ‘image’, and so on—which creates problems whenever the objects of analysis do not fit within the boundaries that such informal categories usually invite. Other authors also proceed directly to analysis, but more or less explicitly reject the utility of the entire conception of ‘semiotic modes’ as well (cf., e.g., Moya Guijarro 2014, 59–60; Green 2014, 9–10; Ledin & Machin 2019). Others may place the main locus of definition elsewhere, as in Norris’ productive focus on sites of interactive engagement (Norris, 2016, 122) which may draw on any semiotic resources, objects or environmental configurations as required, again without hypostatizing a notion of ‘semiotic mode’ at all.

These reoccurring difficulties with ‘semiotic mode’ and its definitions reflect rather precisely the difference between an internal conceptual language with ‘stronger grammar’ and one with ‘weaker grammar’ in Bernstein’s sense. The latter does not support generation of new descriptions; items are connected loosely with one another and hence do not support reasoning with the terms invoked. In contrast, a ‘stronger grammar’ provides precisely this generative capability of reasoning—that is, further propositions follow from the attribution of properties. The distinction is in many respects analogous to Kress and van Leeuwen’s earlier proposal of the application of the grammatical-lexical cline to semiotic modes (Kress & van Leeuwen, 2001, 113); the grammatical end of the continuum supports reasoning, while the lexical end of the continuum only supports labeling. Placing a definition of something as, for example, a semiotic mode within a stronger system, then has consequences: there are other properties and configurations of theoretical elements that will apply as well as associated $L^{1.5}$ methods for validation; within a weaker system, nothing particularly follows from such an attribution apart from the fact of labeling itself and discursive connections to other horizontally associated terms, such as ‘meaning-making’, ‘material’, ‘metafunction’, etc.

Nevertheless, it is also equally important here not to lose sight of the fact that horizontal discourse has made, and continues to make, crucial contributions

to multimodality and its development as a field – a situation we have positively evaluated in this collection in terms of diversity. Indeed, as Maton emphasizes, not only do *both* ‘modes’ of knowledge building—i.e., the cumulative knowledge building of hierarchical discourse and the segmented knowledge building of horizontal discourse—produce knowledge, the latter can even be particularly “prolific in providing exemplary analyses for others to study” (Maton, 2011, 80), as well as creating “concepts of sufficient versatility to be flexible enough for any research” (Maton, 2011, 76). Maton is talking specifically of Bourdieu in this case but the point holds far more generally; Doran (2019), for example, discusses at length a humanities text of central importance in the establishment of ‘ethnopoetics’, analyzing how the text argues that its

way of seeing the world can be applied to any number of semiotic practices, from language to ritual to poetry to dance to the dream and onwards. In this way the text cultivates a particular disposition, one that can appreciate an ever-wider range of phenomena through a nuanced interpretative gaze. The ideal knower being built here is one that can perceptively interpret and appreciate a range of potentially new situations with a particular kind of principled judgement. [...] Many disciplines do not focus on precisely describing and accurately predicting the world. They cultivate refined ways of interpreting the world [...] (Doran, 2019)

Producing such forms of ‘knowers’ is also an important task for any discipline, but particularly for those involved in horizontal knowledge building, where the cultivated knower’s gaze must play a central role.

The extension of this analysis to the case of multimodality is straightforward. To see the value and importance of horizontal discourse of this kind for multimodality it is sufficient to consider many of the more theoretical discussions given, for example, by Gunther Kress. Kress developed richly associative webs of terms which not only describe complex semiotic situations, even those which are quite everyday and ‘banal’, as he characterizes it, but which also make those situations and their ways of working accessible as objects of analysis. By being able to read situations through webs of associated terms, those situations are revealed at a level of detail demanding closer attention. One web of such associations for reading sense into communicative events can stand here as an example for many:

‘What are the means for making these meanings as *signs*, as *syntagms*, as *texts*, as *arrangements*?’ and ‘How are these means (to be) used in making apt *arrangements*?’. By and large I will focus on three of these means for making meanings material: on *mode* as the material stuff, the socially shaped material means; on *text* – or equivalent semiotic entity – as the largest level unit in communication; and on *syntagms* as *arrangements* of many kinds. (Kress, 2010, 146; original emphasis)

The italics are important in this quotation, as it is precisely the way that the italicized terms are woven into the discourse that serves as their ‘definitions’. This web of associations is then the language provided for engaging with multimodality and communication; it provides the particular ‘nuanced interpretative gaze’ expected of the multimodal researcher.

By these means, uses of various ways of making meaning can be linked simultaneously in discursive analysis to their material instantiations, to their enacting and restructuring of social roles of power and dependence and to their selection as appropriate (‘apt’) responses of a concerned communicating agent (always socially embedded). The drive to consider communicative situations (potentially) in their entirety, without prejudging the issue as to what material variations in those situations will be ruled out of court, constitutes the essential momentum out of which multimodality as a field emerged. Without such work of boundary-extension many communicative phenomena would remain invisible and, as a consequence, would not have been made available as potential targets of systematic study. They would thereby remain positioned beyond theoretical understanding and practical investigation.

The use of terms in this way correlates with a fluidity in their use, where terms can readily morph and flow into one another, as when modes become media, resources become modes, genres become media (and *vice versa*), and so on. Descriptions of this kind cannot be ‘falsified’, therefore, and contributing to any such discourse of falsification is not their aim. The fluidity of the terms and their multiple inter-relationships created through the horizontally organized account instead construct discourses that can be engaged with and applied, modified and extended, discussed for their internal consistency and economy, and grown by drawing on yet further sets of terms—Huc-Hepher (2015), for example, makes links from Kress’s terms to other sets of terms from Bourdieu, while Kress himself relates to further terms from ecopsychology (‘affordance’), from sociology and Goffman’s notions of framing, work on Japanese composition (e.g., ‘modular’: Kress, 2010, 147), and many more.

Descriptions produced in this way and their underlying constructions are invaluable *tools for thinking semiotically* and can support useful conjectures, new conceptual arrangements, and are always ready to address new phenomena. But they also complicate any contact to data because they do not stand in the kind of relationship to empirical study necessary for allowing empirical results to feed back into formulation of the model and of the relationships proposed between its terms. In short, on the one hand, there is no L^2 to speak of and, at best, a purely horizontally organized $L^{1.5}$; but, on the other hand, there is a highly open and enriching L^1 capable of bringing attention to previously unthought of areas and forms of communication and framing our engagement with such areas and

forms in a potentially highly beneficial fashion. As set out above, several of the contributions to this volume similarly take on precisely this task of engaging with new areas of multimodal concern.

However, these benefits notwithstanding, although useful when studies are more explorative, the much less constrained contact with data exhibited by this kind of discourse also leads readily to analytic imprecision when tasks are intended to become more empirically targeted. For researchers coming from hierarchically-oriented disciplines, approaches of this kind consequently appear to lose too much with respect to their hold on data and so will, in all likelihood, automatically be found less appealing. Thus, as a discipline develops, for all its potential for expanding the field of interest, remaining solely within the horizontally-organized mode of knowledge building is problematic. But remaining within a purely hierarchically-organized discourse would also be disempowering and self-limiting, particularly for the purposes of a theory of multimodality. Disciplines whose knowledges are hierarchically-organized tend to configure their objects of concern more narrowly—as with, for example, a traditional view of linguistics that explicitly excludes the ‘non-’ or ‘extra’-verbal; this would then readily come to speak against the kind of progressive expansion of subject matters that is almost definitional for the pursuit of multimodality.

Combining the positive functionalities of the two styles of knowledge construction can then be seen as one of the main challenges to be faced. Both styles have their roles to play, but, centrally for the formation of any ‘discipline’ of multimodality, they will need to be made to engage in productive dialogue. Too often, in other fields as well as in multimodality, the co-existence of the two ‘modes’ of discourse can lead more to a split, or a fragmentation, along all too traditional lines such as qualitative-quantitative, empirical-hermeneutic, scientific-humanities: for multimodality to advance and grow, this would be a fatal development.

5 Final Words and Outlook

The points of potential connection and difference between differing facets of performing multimodality research discussed above therefore establish the following burning issue: how to move away from a predominance of horizontal knowledge building to include more integrated and integrative views of multimodality. Such views need both to productively *combine* hierarchical and horizontal knowledge formations *and* to articulate the external languages of description necessary for effectively relating to data. Both goals need to be addressed; either alone is insufficient.

First, any hierarchically-organized L^1 that does not have access to empirical methods is problematic, regardless of how detailed it may become. Consequences of this state of affairs are widespread in the literature. Examples include cases where close analyses of some particular multimodal artifacts or performances are accompanied by claims that no such analysis has been provided previously for such artifacts or performances, even though artifacts and performances that are very similar (considered from the perspective of multimodality) *have* received attention. Here the problem is generally that without an appropriate mediating $L^{1.5}$ it is not clear just what descriptions and results may be generalized and which not. This leads to very similar analyses being recycled without growth, cumulation and generalization of any results that have been achieved. Similar repetitions occur at the methodological level, with more or less commonsense proposals concerning how multimodal analysis should be performed being recycled in the various technical languages of any adopted L^1 .

Second, to the extent that it does not engage with related areas of theory that have become more hierarchically-organized, any horizontally-organized L^1 is intrinsically disempowered with respect to its *reliable applicability*. Its accounts remain discursive and, often, anecdotal. Instead, drawing on areas within an overall L^1 that have become more hierarchical in their organization constructs potential bridges to data (e.g., via a correspondingly defined L^2) that, in turn, may serve to progressively enlarge hierarchically-organized pockets even within a more broadly horizontally-organized L^1 . Note that it is always possible, and indeed likely, that such areas of relatively increased hierarchy will emerge, even within horizontally-organized domains: one example of this is the move seen running through work from Kress & van Leeuwen (2001) onwards concerning the identification of more general semiotic principles, such as framing and arrangements, that apply across semiotic modes and materials. Developing corresponding $L^{1.5}$ and L^2 descriptions for such areas should then be seen as a logical next step, and has, in fact, already been pursued in the finely articulated account of materiality and its methodological implications given in Bateman et al. (2017, 101–110).

This then not only suggests that an explicit and far stronger disciplinary orientation to leverage off such combinations would be beneficial, but also begins to delineate just what kinds of theory-building would ideally advance the state of the art at this time—serving not only to increase the field's internal coherence but also to provide a more attractive partner for interaction with other hierarchically-oriented disciplines. There are, after all, few grounds for going from a discipline which offers effective strategies of integrative knowledge-building to one which, apparently, does not.

To move the field forward, therefore, considerations of achieving disciplinary status and, moreover, the *consequences* of such a move for how multimodality

is enacted, are now urgent: it is time to consider just what such a status might mean for the *practice* of researching multimodality. If a field has claims to be a discipline, there should also be certain goals that it should be able to meet, or at least set within its sights. To assist in this, LCT has been relevant for the present discussion at two levels: first, LCT concerns knowledge-building in general and so can be employed to characterize preferable or potentially more effective styles of discourse within any discipline, including multimodality; second, it has been used to describe some existing forms of engagement and has developed both diagnostics for some of the problems that those accounts are facing and strategies for solving those problems. In the Introduction to this volume, we saw how multimodality as a field is moving to accept larger-scale empirical investigations as an inescapable and necessary step for the field as a whole. This makes it particularly important that more clarity is achieved concerning not only its conceptual L^1 underpinnings but also its practices of operationalization for approaching data, both non-specifically, in terms of forms of annotation and data organization ($L^{1.5}$), and specifically, with regard to the actual materialities being interrogated. Taking those challenges seriously will be of considerable significance for advancing the field.

Achieving tighter L^1 formulations will also be necessary for improving relations to other fields and disciplines, an issue of central importance at this time. Maton et al. (2016, 111) claim, for example and rather rhetorically, that “Interdisciplinarity suffers from a rhetoric-reality gap. Arguments proclaiming its necessity outnumber examples of its actuality.” The contributions to this volume, as well as the large majority of work in multimodality more broadly, document that this is certainly not the case with multimodality. The interdisciplinarity of research questions being addressed from multiple disciplines simultaneously by scientifically heterogeneous teams is very much part of the multimodalist’s everyday reality. However, the greater the distance between the disciplines involved, particularly in terms of the kinds of discourses (horizontal and hierarchical) that they bring to bear both internally and in relation to data, the more challenging such cooperative work becomes. Moving freely between qualitative and quantitative methods and between experimental, corpus-based, and conceptual considerations, will play a, probably *the*, major role in making this work. This will again demand closer attention to the kinds of unifying methodological and conceptual considerations set out here.

Finally, therefore, the establishment of a discipline of multimodality that is firmly anchored on the integrative side, with defined mediating languages for bringing theory together with data effectively, and with structured, relational definitions of its conceptual basis, can only be a beneficial move. Such an entity would be of service not only for understanding multimodality as a phenomenon in its own right—the proper ‘object of study’ of the discipline as such—but also for any

other disciplinary investigations where some object or objects of concern come into contact with multimodal issues. Being able to draw on and apply mechanisms and methods for engaging with multimodal materials alongside any such discipline's own areas of expertise would share both theoretical and practical 'work' more appropriately and in ways that would avoid impoverished reinventions. Connecting to the disciplinary concerns of multimodality would also naturally then help results and questions extend beyond the natural confines of those disciplines, thus furthering knowledge-building on a broader scale as well.

Acknowledgements

Thanks are due to Emilia Djonov (Macquarie) for the initial suggestion to consider the application of LCT, and to Yaegan Doran, Jim Martin and Karl Maton (Sydney) and Jana Pflaeging (Salzburg) for detailed feedback on the result—none of whom should receive blame for surviving inaccuracies or more contentious claims.

Bibliography

- Bateman, John A. 2019. Towards Critical Multimodal Discourse Analysis: A response to Ledin and Machin. *Critical Discourse Studies* 16(5). 531–539. <https://doi.org/10.1080/17405904.2018.1550430>.
- Bateman, John A., Judy L. Delin & Renate Henschel. 2004. Multimodality and Empiricism: Preparing for a Corpus-based Approach to the Study of Multimodal Meaning-Making. In Eija Ventola, Cassily Charles & Martin Kaltenbacher (eds.), *Perspectives on Multimodality*, 65–87. Amsterdam: John Benjamins.
- Bateman, John A., Janina Wildfeuer & Tuomo Hiippala. 2017. *Multimodality – Foundations, Research and Analysis. A Problem-Oriented Introduction*. Berlin: Mouton de Gruyter.
- Bernstein, Basil. 2000. *Pedagogy, Symbolic Control and Identity: Theory, Research, Critique. Second Edition*. Oxford: Rowman & Littlefield.
- Caple, Helen. 2018. Analysing the Multimodal Text. In Charlotte Taylor & Anna Marchi (eds.), *Corpus Approaches to Discourse: A Critical Review*, 85–109. Abingdon and New York: Routledge.
- Christie, Frances & Karl Maton. 2011. Why Disciplinarity? In Frances Christie & Karl Maton (eds.), *Disciplinarity: Functional Linguistic and Disciplinary Perspectives*, 1–12. London: Continuum.
- Cohn, Neil. 2016. A Multimodal Parallel Architecture: A Cognitive Framework for Multimodal Interactions. *Cognition* 146. 304–323.
- Doran, Yaegan. 2019. Cultivating Values: Knower-Building in the Humanities. *Estudios Lingüística Aplicada* In press.
- Forceville, Charles J. 1999. Educating the Eye? Kress and Van Leeuwen's *Reading Images: The Grammar of Visual Design* (1996). *Language and Literature* 8(2). 163–178. <https://doi.org/10.1177/096394709900800204>.

- Forceville, Charles J. 2007. Book Review: *Multimodal Transcription and Text Analysis: A Multimedia Toolkit and Coursebook* by Anthony Baldry and Paul J. Thibault. *Journal of Pragmatics* 39(6). 1235–1238.
- Fricke, Ellen. 2012. *Grammatik multimodal: Wie Wörter und Gesten zusammenwirken*. Berlin and New York: Mouton de Gruyter.
- Glaser, Barney G. & Anselm L. Strauss. 1967. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. New York: Aldine de Gruyter.
- Green, Jennifer D. 2014. *Drawn from the Ground: Sound, Sign and Inscription in Central Australian Sand Stories*. Cambridge: Cambridge University Press.
- Hasan, Ruqaiya. 1973. Code, Register and Social Dialect. In Basil Bernstein (ed.), *Class, Codes and Control: Applied Studies Towards a Sociology of Language*, 253–292. London: Routledge and Kegan Paul.
- Huc-Hepher, Saksia. 2015. Big Web Data, Small Focus: An Ethnosemiotic Approach to Culturally Themed Selective Web Archiving. *Big Data & Society* 2(2). 1–15. <https://doi.org/10.1177/F2053951715595823>.
- Jewitt, Carey. 2014. An Introduction to Multimodality. In *The Routledge Handbook of Multimodal Analysis. Second Edition*, 15–30. London and New York: Routledge.
- Jewitt, Carey, Jeff Bezemer & Kay O'Halloran. 2016. *Introducing Multimodality*. London: Routledge.
- Jewitt, Carey & Gunther Kress. 2003. *Multimodal Literacy (New Literacies and Digital Epistemologies 4)*. Frankfurt a.M. and New York: Peter Lang.
- Kelle, Udo. 2005. "Emergence" vs. "Forcing" of Empirical Data? A Crucial Problem of "Grounded Theory" Reconsidered. *Forum Qualitative Sozialforschung / Forum Qualitative Social Research* 6(2). Art. 27. <https://doi.org/10.17169/fqs-6.2.467>.
- Kress, Gunther. 2010. *Multimodality: A Social Semiotic Approach to Contemporary Communication*. London: Routledge.
- Kress, Gunther & Theo van Leeuwen. 2001. *Multimodal Discourse. The Modes and Media of Contemporary Communication*. London: Arnold.
- Krippendorff, Klaus. 2004. *Content Analysis: An Introduction to its Methodology. Second Edition*. London and Thousand Oaks: Sage.
- Krishnan, Armin. 2009. *What Are Academic Disciplines? Some Observations on the Disciplinarity vs. Interdisciplinarity Debate*. University of Southampton National Centre for Research Methods. Working Paper.
- Ledin, Per & David Machin. 2019. Doing Critical Discourse Studies with Multimodality: From Metafunctions to Materiality. *Critical Discourse Studies* 16(5). 497–513. <http://doi.org/10.1080/17405904.2018.1468789>.
- Loschky, Lester C., Adam M. Larson, Joseph P. Magliano & Tim J. Smith. 2015. What Would Jaws Do? The Tyranny of Film and the Relationship between Gaze and Higher-Level Narrative Film Comprehension. *PLoS One* 10(11). <https://doi.org/10.1371/journal.pone.0142474>.
- Martin, James R. 2011. Bridging Troubled Waters: Interdisciplinarity and What Makes it Stick. In Frances Christie & Karl Maton (eds.), *Disciplinarity: Functional Linguistic and Disciplinary Perspectives*, 35–61. London: Continuum.
- Maton, Karl. 2011. Theories and Things: The Semantics of Disciplinarity. In Frances Christie & Karl Maton (eds.), *Disciplinarity: Functional Linguistic and Disciplinary Perspectives*, 62–86. London: Continuum.
- Maton, Karl. 2014. *Knowledge and Knowers: Towards a Realist Sociology of Education*. London and New York: Routledge.

- Maton, Karl. 2016. Legitimation Code Theory: Building Knowledge about Knowledge-Building. In Karl Maton, Susan Hood & Suellen Shay (eds.), *Knowledge-Building: Educational Studies in Legitimation Code Theory*, 1–24. Abingdon and New York: Routledge.
- Maton, Karl & Rainbow Tsai-Hung Chen. 2016. LCT in Qualitative Research: Creating a Translation Device for Studying Constructivist Pedagogy. In Karl Maton, Susan Hood & Suellen Shay (eds.), *Knowledge-Building: Educational Studies in Legitimation Code Theory*, 27–48. Abingdon and New York: Routledge.
- Maton, Karl & Sarah K. Howard. 2016. LCT in Mixed-Methods Research: Evolving an Instrument for Quantitative Data. In Karl Maton, Susan Hood & Suellen Shay (eds.), *Knowledge-Building: Educational Studies in Legitimation Code Theory*, 49–71. Abingdon and New York: Routledge.
- Maton, Karl, James Martin & Erika Matruglio. 2016. LCT and Systemic Functional Linguistics: Enacting Complementary Theories for Explanatory Power. In Karl Maton, Susan Hood & Suellen Shay (eds.), *Knowledge-Building: Educational Studies in Legitimation Code Theory*, 93–113. Abingdon: Routledge.
- Moore, Rob & Johan Muller. 2002. The Growth of Knowledge and the Discursive Gap. *British Journal of Sociology of Education* 23(4). 627–637.
- Moya Guijarro, Arsenio Jesus. 2014. *A Multimodal Analysis of Picture Books for Children: A Systemic Functional Approach*. London: Equinox.
- Muller, Johan. 2011. Through Others' Eyes: The Fate of Disciplines. In Frances Christie & Karl Maton (eds.), *Disciplinary: Functional Linguistic and Disciplinary Perspectives*, 13–34. London: Continuum.
- Norris, Sigrid. 2016. Multimodal Interaction – Language and Modal Configurations. In Nina-Maria Klug & Hartmut Stöckl (eds.), *Handbuch Sprache im multimodalen Kontext* (Handbooks of Linguistics and Communication Science (HSK) 7), 121–144. Berlin: de Gruyter Mouton.
- Pennock-Speck, Barry & María M. del Saz-Rubio (eds.). 2013. *The Multimodal Analysis of Television Commercials*. Valencia: University of Valencia Press.
- Raz, Gal & Talma Hendler. 2014. Forking Cinematic Paths to the Self: Neurocinematically Informed Model of Empathy in Motion Pictures. *Projections: Journal for Movies and Mind* 8(2). 89–114.
- Schreier, Margrit. 2012. *Qualitative Content Analysis in Practice*. London: Sage.
- Stöckl, Hartmut. 2004a. *Die Sprache im Bild — Das Bild in der Sprache: Zur Verknüpfung von Sprache und Bild im massenmedialen Text. Konzepte – Theorien – Analysemethoden*. Berlin: Walter de Gruyter.
- Stöckl, Hartmut. 2004b. In Between Modes: Language and Image in Printed Media. In Eija Ventola, Cassily Charles & Martin Kaltenbacher (eds.), *Perspectives on Multimodality*, 9–30. Amsterdam: John Benjamins.
- Strauss, Anselm L. & Juliet Corbin. 1990. *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*. Newbury Park, CA: Sage.

List of Contributors

John A. Bateman (Ph.D. in Artificial Intelligence) has been Professor of Applied Linguistics in the Faculty of Linguistics and Literary Sciences at the University of Bremen since 1999. His main fields of research range over computational linguistics (particularly natural language generation, discourse, and dialogue), formal ontology, and the theory and practice of multimodality. He has published widely in all of these areas, including monographs on text generation (1991, Pinter, co-authored with Christian Matthiessen), multimodality and genre (2008, Palgrave), film (2012, Routledge, with Karl-Heinrich Schmidt), text and image (2014, Routledge), and an introduction to multimodality as a whole (2017, de Gruyter, with Janina Wildfeuer and Tuomo Hiippala). Recent work focuses specifically on the semiotic foundations of multimodality and the use of empirical methods for their investigation.

Kevin Chai (Ph.D. in Computer Science) is currently working as the lead data scientist for Curtin's Institute for Computation (CIC) to provide data, computing, analytics, and machine learning/artificial intelligence expertise to collaborate in projects with researchers and industry partners. Kevin completed his doctorate in computer science at Curtin University, worked as Research Fellow at the University of New South Wales developing machine learning and natural language processing models to analyze large medical datasets and as a Data Scientist at Curtin University on developing predictive models. Kevin has research interests and experience in machine learning/AI and its application to solving important problems in domains such as health, medicine, education, science, engineering, the humanities, and many other diverse areas of research.

Helen Georgiou (Ph.D. in Physics) is a Lecturer in Science Education at the University of Wollongong, Australia. She holds an undergraduate degree in physics and a DipEd in Education and was previously a practicing high school physics teacher. Her Ph.D. was awarded from The University of Sydney and focused on Physics Education, specifically, students' understanding of thermodynamics. Her research aims to understand the issues in science education, including students' conceptual difficulties, through studying the literacy practices and social aspects of the discipline. She is currently studying students' multimodal representations of scientific knowledge using Legitimation Code Theory.

John Harnett (Ph.D. in English Literature) is an independent post-doctoral researcher from Mary Immaculate College in Limerick, Ireland, which is an academic partner of the University of Limerick. His research interests include psychoanalysis, post-colonial theory, popular culture studies, adaptation studies and the gothic with a focused interest in the application of narratology, psychoanalysis, and multimodality to the field of comics studies. He is currently analyzing the affordances and constraints of modality within the comics medium with regards to its articulation of such concepts as time, memory and trauma. His publications include several interdisciplinary, literary and bridge-building analyses of comics and graphic novels with Inter-Disciplinary Press, and he has presented on a wide variety of approaches to the field of comics studies.

Milan Jačević, M.Sc., M.A., is a Ph.D. student at the Institute of Visual Design at the Royal Danish Academy of Fine Arts. His Ph.D. project focuses on the examination of synchronic and diachronic effects of digital gaming from the complementary standpoints of practice theory and cognitive

<https://doi.org/10.1515/9783110608694-013>

science. He holds an M.Sc. in Game Design and Theory from the IT University of Copenhagen, an M.A. in Language, Literature and Culture from the University of Belgrade, and a B.A. in English Language and Literature from the University of Niš. In addition to practice theory and cognitive science, his research interests include multimodality, game ontology, ludohermeneutics, and ideology.

Pauline Jones (Ph.D. in English) is an Associate Professor in the School of Education at the University of Wollongong, Australia, where she convenes undergraduate and postgraduate English curriculum and literacy subjects. Her research interests are informed by systemic-functional theory and include educational semiotics, classroom discourse, disciplinary literacies, and teachers' professional learning. She supervises a number of Ph.D. students (including Annette Turney) studying related areas. Her current projects include an Australian Research Council funded project (with Wendy Nielsen) investigating the semiotic character and use of student generated digital explanations, and a Faculty Seed project investigating creativity and innovation in curriculum disciplines (with Helen Georgiou). She has recently led a large-scale investigation of contemporary literacy practices across curriculum fields involving more than 40 teachers and developing over 70 case studies of the experiences of students over three years at different points along the preschool to junior secondary continuum.

Rebecca Lange (Ph.D. in Astronomy) is a Computational Specialist with the Curtin Institute for Computation. Rebecca completed her Ph.D. in astronomy at the International Centre for Radio Astronomy Research at the University of Western Australia. She has worked on the development, testing, and in situ implementation of instruments and software for the Imaging & Sensing for Archaeology, Art History & Conservation group at Nottingham Trent University in the UK. Throughout her studies and research she has gained extensive programming as well as data analytics and visualization experience utilizing various programming languages. Rebecca is particularly interested in applying her technical skills to support and train researchers in the digital humanities.

Edward Larkey (Ph.D. in Cultural Studies from Humboldt University Berlin) is a Professor of German Studies and Intercultural Communication at the University of Maryland, Baltimore County (UMBC), USA. His research projects cover areas of international format adaptations and their localization strategies in Germany, Israel, and other countries. He is currently involved in cross-cultural and transnational comparative research methodologies of local format adaptations incorporating narratological and multimodal approaches. Working with an international team of researchers from a variety of countries, he is currently collaborating in a research project on comparing local adaptations of the Quebec television series *Un Gars, Une Fille*, one of the most popular and successful global formats for scripted genres.

Mary Macken-Horarik is an adjunct Associate Professor in the Institute for Learning Sciences and Teacher Education (ILSTE) at the Australian Catholic University, Australia. Her research interests include educational linguistics, literacy education, knowledge practices in school English, and media discourse analysis. Her recent publication focuses on Functional Grammatics (2018, Routledge).

Konstanze Marx (Ph.D. in Linguistics) is Professor of German Linguistics in Mannheim, Germany. She has previously worked as Assistant Professor at the department of General Linguistics at

Technical University Berlin. Her publications include a first monograph entitled *Die Verarbeitung von Komplex-Anaphern. Neurolinguistische Untersuchungen zur kognitiven Textverstehenstheorie* and a second one entitled *Diskursphänomen Cybermobbing. Ein internetlinguistischer Zugang zu [digitaler] Gewalt*. She is also the co-author of *Internetlinguistik. Ein Lehr- und Arbeitsbuch* (with Georg Weidacher). Her research fields include internet linguistics, cognitive linguistics, discourse linguistics, language and emotion (especially cyberbullying), and research infrastructures for the analysis of interaction in social media.

Wendy Nielsen (Ph.D. in Science Education) is an Associate Professor in the School of Education, University of Wollongong, Australia. She earned a B.Sc. in Biology from California Lutheran College and Master's and Doctoral degrees in Curriculum Studies from the University of British Columbia. Her research focuses on teaching and learning strategies in science, particularly in preservice primary teacher education. In 2016–2018, she led an Australian Research Council Discovery Project examining university science student learning through the creation of digital explanations. She has written in STEM-related areas of science teacher education, nature of science, history of science, and environmental education. Other research interests include doctoral education and supervising teacher knowledge.

Lauren O'Hagan recently completed a Ph.D. in Language and Communication within the School of English, Communication, and Philosophy at Cardiff University, Wales. Her thesis was entitled 'Class, Culture and Conflict in the Edwardian Book Inscription'. It involved an interdisciplinary approach, which blended theory from multimodality, ethnography, and book history, to explore social class in Edwardian society through a dataset of 3,000 inscriptions present in books circulated between 1901–1914 in Great Britain. Lauren currently works as a translator and EFL tutor.

Professor **Kay O'Halloran** is Head of Department and Chair in Communication & Media in the School of the Arts at the University of Liverpool in the United Kingdom. She has a background in mathematics, multimodal discourse analysis, and linguistics. She has extensive experience in establishing and leading interdisciplinary research teams to develop and make widely available new digital tools and techniques for analyzing images, videos, and 360 videos. Kay is currently developing mixed methods approaches that combine multimodal analysis, data mining, and visualization for big data analytics in areas which include online extremism, political rhetoric, maintenance communications in the mining industry, and 360 video.

Jana Pflaeging is a researcher in English and applied linguistics at the Department of English and American Studies at Salzburg University. She currently pursues a binational Ph.D. at Salzburg University, Austria, and Halle-Wittenberg University, Germany. Her research interests are in multimodal genre studies and text/discourse linguistics. Trained in English linguistics and fine arts, she explores the synergies between both fields when creating visualizations of linguistic and multimodal theories, methods, and data.

Axel Schmidt (Ph.D. in Media Studies) is a Research Fellow at the Institute of German Language and Professor of Media Studies at the University of Mannheim, Germany. He has previously worked as a research fellow at the Institute of Media Studies at the University of Basel, Switzerland. His publications include the monograph *Doing Peer Group. Die interaktive Konstruktion jugendlicher Gruppenpraxis* as well as *Medien – Interaktion. Zum Zusammenhang von Handeln und Darstellen am Beispiel faktualer Fernsehformate*. His research fields cover multimodal inter-

action analysis, conversation analysis, television studies, ethnography, and media production analysis.

Ognyan Seizov (Ph.D. in Communication Science) is currently a Postdoctoral Research Associate at the Contractual Management Institute at SRH Hochschule Berlin, Germany, working in an international collaborative project on improving the transparency of online consumer contracts. Prior to that, he was a postdoctoral researcher for multimodal political communication at the University of Bremen and a visiting postdoctoral fellow at the department of Comparative Media Studies/Writing at the Massachusetts Institute of Technology (MIT), conducting research on multimodal political expression on the Web. He has published on informative and persuasive communication online, digital media production and reception, and multimodal content analysis.

Dušan Stamenković (Ph.D. in Linguistics) is an Assistant Professor in the Department of English, Faculty of Philosophy, University of Niš, Serbia, a researcher at the Center for Cognitive Sciences, and Head of the Language Cognition Lab. He received his B.A. in English from the University of Niš, M.A. in the English language and literature from the University of Novi Sad, and a doctorate in linguistics from the University of Novi Sad. His research interests include cognitive approaches to linguistics, metaphor comprehension, multimodality, comics and video game studies, as well as contrastive linguistics and translation studies. In 2016/17 he was a Fulbright Visiting Scholar at UCLA's Psychology Department and Reasoning Lab. He is a member of the Association for Researching and Applying Metaphor (RaAM), the Slavic Cognitive Linguistic Association (SCLA), the European Society for the Study of English (ESSE), and the Serbian Association for the Study of English (SASE).

Hartmut Stöckl (Ph.D. in English Linguistics) is Full Professor of English and Applied Linguistics in the Department of English and American Studies at Salzburg University, Austria. His main research areas are in semiotics, media/text linguistics/stylistics, pragmatics, and linguistic multimodality research. He is particularly interested in the linkage of language and image in modern media, typography and an aesthetic appreciation of advertising. A volume on *Shifts towards Image-centricity in Contemporary Multimodal Practices* with Routledge (2019) is his most recent co-edited book.

Sabine Tan (Ph.D. in Linguistics) is a Senior Research Fellow and a member of the Multimodal Analysis Group in the School of Education at Curtin University, Western Australia. She has a background in critical multimodal discourse analysis, social semiotics, and visual communication. She has applied multidisciplinary perspectives for the analysis of institutional discourses involving traditional and new media. She has worked on interdisciplinary projects involving the development of interactive software for the multimodal analysis of images, videos, and 360 videos for research and educational purposes. Sabine's current work involves developing multimodal approaches to big data analytics in the fields of online terrorism, political discourse, and 360 video.

Christopher Taylor is Full Professor (retired) of English Language and Translation at the University of Trieste, Italy. He has worked in the field of film translation for many years with significant publications relating to such issues as dubbing, subtitling, and multimodal transcription, and more recently audiovisual translation for the deaf and audio description for the blind. His numerous publications in this field include, most recently: *The Multimodal Approach in Audiovisual*

Translation in a special issue of 'Target' (2016) and *Reading Images (including moving ones)* in the Routledge Handbook of Systemic Functional Linguistics (2017). He recently coordinated a European Union project ADLAB (Audio description: lifelong access for the blind) which achieved 'Success Story' status.

Martin Thomas (Ph.D. in Linguistics). Following a period spent looking at toothpaste packaging and trying to generate multimodal annotations from the output of optical character recognition engines, Thomas now teaches and researches about multimodality, genre, and translation theory at the University of Leeds, UK. He is still interested in leveraging computational methods to support multimodal text analysis at scale and in exploring the intersections between writing systems, typography, and genre.

Chiao-I Tseng (Ph.D. in Linguistics) is a Researcher affiliated to the University of Bremen, Germany. Her area of research focuses on narrative designs across different media, such as films, graphic and interactive media. Tseng specializes in developing frameworks for analyzing narrative forms and contents, particularly the methods for systematically tracking types of events and actions, character features, narrative space, and motivation and emotion. She has published journal articles and book chapters on applying these methods to addressing a wide variety of socio-cultural and aesthetic issues such as complex/non-linear film narratives, characters' developments and motivations, genre comparison, authorship, screen violence, transmedia adaptation, or persuasive functions in film and video games.

Annette Turney (M.Ed., M.Sc., PGCE) is a Ph.D. student at the University of Wollongong, Australia. Her Ph.D. project focuses on the exploration of student-generated digital media in tertiary science contexts using systemic-functional theory. She received her B.A. (Hons) in European Literature and Art History from the University of East Anglia, UK. She is a former secondary English teacher and holds two M.Ed.'s from the University of Wollongong, one in Language and Literacy, and the other in TESOL, and a M.Sc. in Educational Research Methods from the University of Exeter, UK. In addition to systemic-functional theory and education, her research interests include multimodality, digital literacies, and legitimation code theory.

Rui Wang (Ph.D. in Computer Science) is a specialist data scientist at Rio Tinto and former data scientist in the Curtin Institute for Computation at Curtin University, Western Australia. His research areas include natural language processing, machine learning, and text mining. Rui has also worked in data mining for CRC Water Sensitive Cities at the University of Western Australia. His research involved developing time series data mining algorithms to detect anomalous water consumption. Rui is also involved in a collaboration with the team from University of Western Australia Computer Science and Planning and Transport Research Centre for analyzing large traffic datasets. In addition, Rui has nearly 10 years' commercial software development experience.

Peter Wignell (Ph.D. in Linguistics) is a Senior Research Fellow and a member of the Multimodal Analysis Group in the School of Education at Curtin University, Western Australia. He has an academic background in systemic-functional linguistics, discourse analysis, and multimodal analysis. His research has been both theoretical and applied. For example, his work on the role of language in the construction of specialized knowledge systems has informed literacy theory and pedagogy. Recent research has led to publications in a diverse range of applications such as brand semiotics, performance studies, and translation. Peter's current research focus is on the

application of a systemic-functional multimodal analysis approach to violent extremist discourse and political discourse.

Janina Wildfeuer (Ph.D. in Linguistics) is an Assistant Professor in Language and Social Interaction in the Department of Communication and Information Studies at the University of Groningen, NL. Her main research interests lie in the areas of multimodal linguistics, media studies, discourse analysis, and semiotics. She teaches classes in multimodal, interdisciplinary, and applied linguistics and analyses films, comics, and other multimodal documents in several projects exploring the notion of multimodal discourse. Her publications include several monographs and edited collections as well as contributions and articles on the analysis of multimodal artifacts, mostly focusing on interdisciplinary approaches in the humanities and beyond.

Index

- abstract structures
 - in linguistics 97, 100, 103, 104
 - in mathematics 94–96, 100, 103
- aesthetic proximity 244
- aesthetics 146, 306
 - multimodal 73
 - web 73
- affordance 10, 11, 23, 47, 53, 75, 80
 - modal 175, 179, 180
- annotation 70, 80, 81, 83, 284, 285
 - manual 72
 - XML 80, 81
- archive 78, 147, 151
 - digital 146, 151, 166
- audio description 10, 195–198, 203, 209, 213, 216
 - approaches 199, 201, 206, 211, 216
 - as genre 216
 - definition 197
 - textuality 196
- author 127, 129

- big data 105
- book inscription 145, 147, 150, 151, 153, 164, 166

- canvas 49, 50, 282
- cause-effect chain 225
- class conflict 147
- classification 100, 105, 108
- clause relation 55, 56
- co-contextualization 41
- coherence 52–58, 60, 64, 65
- cohesion 52, 53, 55, 57, 60
- comics 171–173, 181, 182, 185, 187, 190, 192
- communication 93
 - different forms of 116, 118, 123–125, 134
 - human 94, 100, 101, 109, 110
 - non-mediated 116
 - YouTube 118, 119, 124, 125, 128–131, 133, 137
- communication studies 41, 70, 116
- communicative act 116, 129, 133, 136

- communicative action 11
- communicative structure 117
- communicative unit 115, 117, 131
- comparative analysis 220
- computer pattern recognition 82
- computer vision 73, 81, 83, 88
- context 41, 47–50, 52, 53, 55–59, 65, 69, 75–77, 84, 88, 94, 98, 100, 101, 103, 105, 107, 108, 148, 152
 - historical 145, 149
 - social 49, 50, 70, 149, 152, 155
- conversation analysis 244
- cookie-cutter syndrome 301
- corpus 72, 78, 83, 86
 - multimodal 12, 25, 71, 77, 79
- corpus building 71, 83, 115–117, 119, 129–131, 134, 136, 137
- cross-modal realization 174, 175, 178, 179
- cultural analytics 72, 73
- cultural capital 152, 163
- cultural context 14
- cultural semantics 51
- cultural studies 72, 146, 147
- culture 145, 146, 148, 150, 158, 164
 - Victorian commodity 150, 155

- data mining 117
- data sampling 117, 131
- database 78, 116
- dataset 70, 76, 78
- design 45, 49, 59–61, 64, 65, 73, 81, 85
 - document 72, 85, 86
 - graphic 73, 82
 - web 73
- dialog transcription 244
- dialogicity 11, 53
- digital artifact 249, 250, 262, 272, 273
- digital content 81, 83
- digital explanation 14, 248, 250, 251, 253, 255, 256, 267, 268, 271–273
- digital humanities 6, 9, 16, 18–21, 27–29, 31, 33, 72, 146
- digital technology 93, 94, 109, 110
- digitalness 77, 86

<https://doi.org/10.1515/9783110608694-014>

- directionality 80
 discipline
 – as institution 310
 – characteristics of 4–6, 8, 9, 12, 14–19, 21, 28, 30, 41, 42, 53, 93, 94, 96, 109, 110
 – low- vs. high-end 18, 28, 31
 discursal practices of communication 298
 discourse analysis 10, 24, 42, 278, 283
 – multimodal 24, 25
 discourse semantics 47, 48, 50, 52, 57, 59, 62, 64, 75
 discursive gap 298, 300, 301, 303
 document 72, 73, 82–84
 – multimodal 78, 83, 84
 – properties 70, 83
 – two-dimensional 70, 71, 80, 85, 86
- ethnography 148, 149
 ethnohistory 146–149, 151, 164
 external languages of description 301, 302, 316
 eye-tracking 25
- feature enhancement 105, 107
 feature extraction 103, 105
 feature representation 105, 108
 field 251, 253–255, 260–262
- game system 278, 280, 281
 gameworld 278–281, 284
 GeM model 80, 81, 83, 278, 279, 283, 288, 290–292
 genre 41, 42, 44, 45, 48–50, 52, 54–57, 59, 60, 62, 64, 65, 76, 77, 84–86, 88, 148, 152, 248–250, 252–258, 261, 262, 268, 271, 272
 – elemental 252, 254, 255
 – macro- 252, 254, 255, 258, 272
 grammar 47, 55
 – systemic-functional 41, 55
 graphic novel 10, 13, 25, 172, 174, 179, 181, 185
 grid 81, 85
- handwriting 157, 159, 165
 head-up display 280, 281
 hyperlink 136, 285, 288
- interaction 116, 118, 124–126, 128, 129, 134, 136, 137, 279, 280, 286, 287
 – human-computer 278, 281, 290, 291
 – sender-recipient 124, 126, 128, 137
 – situated 118
 interaction analysis 10, 28, 42, 52
 interactional data 131, 134
 interactional pragmatics 118, 130
 interactive visualization 108
 interdisciplinarity 5, 7–9, 18, 70, 308, 318
 interface 277–282, 284, 290
 intermodal harmony 53, 54
 intermodal relations 10, 24, 78, 79
 internal conceptual languages 304
 interpretation 42, 44, 47, 48, 52, 55, 56, 58, 61, 65
 intersemiotic complementarity 53, 64
 intersemiotic relations 53
- knowers 314
 knowledge 145
 – construction 298, 316
 – creation 298
 – hierarchical 300, 306, 311, 316
 – horizontal 300, 302, 310, 311
 – multimodal 244, 245
- language 73, 79, 80, 82, 93, 94, 97, 98, 100, 101, 103, 105, 110
 – markup 82
 – model 75
 – multimodal 110
 – natural 98, 103–105
 language sciences 78, 94, 97, 109
 languages of description 301
 layout structure 288, 289
 Legitimation Code Theory 14, 18, 251, 255, 262, 263, 269, 271, 298–300, 302, 303, 305, 307, 310, 318
 – translation devices 302, 303
 Let's Play 115, 117, 119, 120, 127
 lexicogrammar 47, 97, 101
 linearity 80
 linguistic anthropology 148, 149
 linguistics 9, 22–24, 27, 28, 77, 79, 93, 94, 97, 100, 103, 109, 148
 – computational 97

- corpus 79, 85
- different branches of 97
- historical 149
- systemic-functional 10, 24, 74, 97, 100, 101, 148, 299
- text 41
- literary parallax 173, 174
- literary studies 97
- logocentrism 78, 79, 82

- machine learning 104, 105, 108
- material 11, 23, 29
- materiality 11, 23, 47–50, 55, 56, 69, 76, 77, 146, 149, 153, 165, 282, 284, 291
- mathematics 11, 12, 93–95, 97, 100, 103, 109
 - applied 95
 - pure 95
- meaning making 88, 93, 100
- meaning multiplication 11, 44, 52, 63, 64
- media 41, 44–47, 49, 52, 56, 82
 - page-based 14
 - visual 47
- mediality 47, 49, 50, 61, 62, 64
- mediation 279
- meta-theorization 299
- metadata 100, 105, 107, 117, 131, 134
 - automatically generated 134
 - self-generated 134
 - semi-automatically generated 134
- metafunction 75, 86, 97, 101, 107, 108, 148, 153, 305
- metaphorical imagination 178
- mixed methods 105
- modal competence 13
- modal complexity 52, 53
- modal intensity 52
- modal reach 50, 51, 54, 59, 62
- modality 282
- mode 10, 11, 22–25, 28, 30, 41, 44, 46–49, 52, 54, 57, 59, 60, 62–64, 70, 75–77, 79, 85, 86, 117, 119, 132, 148, 248, 251, 253–255, 260, 263, 268, 269
 - combinations 41, 47, 48, 50–56, 59, 60, 62, 64, 76, 79, 80
 - core vs. peripheral 48, 52, 54, 62
 - family 47
 - linking 41, 57, 59, 64
 - mutual elaboration 41, 51
 - orchestration 60, 65
 - properties 46–53, 64, 75, 76, 79
 - semiotic 45, 47, 49, 50, 53, 59, 61, 76, 78, 80, 82, 84, 86, 88, 305–307, 311–313
- mode linking 53
- multidimensional vantage point 183
- multidisciplinarity 4–6, 30
- multimodal analysis 69–74, 76, 78, 80, 82, 88, 277, 278, 284
- multimodal collocation 85
- multimodal concordance 85
- multimodal corpus 130
- multimodal discourse 42, 45, 60–62, 65
- multimodal document 277–279, 283–285, 287, 288
- multimodal ensemble 41, 60, 64, 65
- multimodal fluency 172, 175, 178, 185, 192
- multimodal reception analysis 25
- multimodal rhetoric 11
- multimodal sciences 93, 97, 100, 109, 110
- multimodal theory 42, 44
- multimodality 97, 116, 147, 148, 164
 - characteristics of a discipline 4, 26, 27, 94, 100, 116, 166, 278, 292
 - cross-cultural 220
 - disciplinary status 3, 6, 8, 12, 16, 17, 30, 31
 - ethnohistorical approach to 146, 147, 149
 - institutionalization 3, 26, 27
 - linguistic 11
 - research object 4, 12, 21
 - social semiotic approach to 146, 147, 166
- music notation 244

- narrative 172–176, 178–186, 188, 190, 192, 221, 222, 224–227, 229, 232, 241, 243, 244
 - cohesion 225, 226, 228
 - dilemma 225, 226, 229
 - multimodal 187
 - omnitemporal 182, 183
 - sequential 185, 187, 192
 - televisual 222, 231
- narratological analysis 225, 227, 240
- navigation 178, 179, 186, 187

- neural network 104, 108
- new literacy studies 145, 149
- omnitemporal reader 181
- orchestration 172, 180, 182, 184
- ordinary writing 145–147, 164, 166
- page layout 173, 176, 179, 187
- participation framework 116–119, 125, 126, 129
- performativity 149
- phases of socialization 299
- portrait direction 163
- pragmatics 41
- recipient 125–128, 131
- recognized base unit 283–286
- register 251–258, 262, 268, 271–273
 - variation 255
- releaser 127–129, 133, 136
- resemiotization 61
- rhetorical strategy 59, 61
- rhetorical structure theory 56
- scale 70–73, 80, 81, 83, 88
- scene duration 221, 226, 227, 237, 239
- science education 248, 249, 255
- scientific information 250, 273
- screen 117, 121, 122, 132, 133, 278, 284, 291, 292
 - document-like 278, 279, 282–287, 290, 291
- segmentation 78, 80–83
- semiotic choice 100, 101, 103
- semiotic resource 11, 14, 22, 23, 29, 47–52, 59, 60, 62, 63, 65, 73, 74, 93, 94, 100, 101, 107–109, 148, 164
- semiotic work 47, 55–57, 62, 64
- semiotics 10, 11, 17, 24, 27, 72, 76, 82
 - cognitive 11, 14
 - social 5, 24, 71, 74, 116
 - systemic-functional 256, 271
- sender 126, 127, 133
- signification 281, 282, 290
- simultaneity and panopticism 187
- social class 147, 149, 151
- social institution 299
- social media 48, 116–118, 128, 137, 138
- social practice 145, 146, 148–150
- sociology 116
- television format adaptation 219, 220
- temporality 149
- tenor 251, 253–255, 260, 272
- text 146, 147, 149, 160, 164, 166
 - multimodal 93, 100, 103, 105, 107, 108
- text-image relations 45, 52, 55
- texture 153
- trans-textuality 42, 45, 61, 62, 65
- transcription 51, 77–79, 81, 83, 117, 132
 - multimodal 77, 80, 132
- transdisciplinarity 30
- translation studies 70
- turn 115, 117, 118, 129
- vector 105, 108, 287
- video editing 244
- video game 278–282
- video game studies 278
- video sequence 244
- visual navigator 188
- visual repetition 188, 190
- visuality 279