DE GRUYTER
MOUTON

REFERENCE

CURRENT APPROACHES TO SYNTAX

A COMPARATIVE HANDBOOK

Edited by András Kertész, Edith Moravcsik, Csilla Rákosi

COMPARATIVE HANDBOOKS OF LINGUISTICS

András Kertész, Edith Moravcsik, and Csilla Rákosi (Eds.)
Current Approaches to Syntax

Comparative Handbooks of Linguistics

Edited by
Andrej Malchukov and Edith Moravcsik

Volume 3

Current Approaches to Syntax

A Comparative Handbook

Edited by
András Kertész
Edith Moravcsik
Csilla Rákosi

DE GRUYTER
MOUTON

ISBN 978-3-11-053821-2
e-ISBN (PDF) 978-3-11-054025-3
e-ISBN (EPUB) 978-3-11-053837-3
ISSN 2364-4354

Library of Congress Control Number: 2018963679

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.

Cover image: Jupiterimages/PHOTOS.com/thinkstock
Typesetting: VTeX UAB, Lithuania
Printing and binding: CPI books GmbH, Leck

www.degruyter.com

Acknowledgment

We wish to express our gratitude to the editorial and technical staff of de Gruyter. We much appreciated the efficient handling of the proofs by Sabina Dabrowski at Content Conversion Services and by Project Manager Ina Talandienė and her colleagues at VTeX. We thank Marianna F. Balogh for her thorough work on the indices.

On the editorial side, we are very grateful to Editorial Director Birgit Sievert and In-House Editor of the CHL Series Barbara Karlson for their assistance in compiling and organizing the volume. Our foremost gratitude goes to Barbara. Her expertise, attention to detail, keen eyes to spot problems, sound advice, great patience, and personal touch greatly eased the burden of our editorial responsibilities and fundamentally affected the quality of this volume.

The editors

https://doi.org/10.1515/9783110540253-201

Contents

List of Contributing Authors

Jenny Audring
j.audring@hum.leidenuniv.nl

Cristiano Broccias
c.broccias@unige.it

Philip Carr
philcarrb@yahoo.com

Rui P. Chaves
rchaves@buffalo.edu

Peter W. Culicover
culicover.1@osu.edu

Mary Dalrymple
mary.dalrymple@ling-phil.ox.ac.uk

Sam Featherston
sam.featherston@uni-tuebingen.de

Jamie Y. Findlay
jamie.findlay@ling-phil.ox.ac.uk

Pius ten Hacken
pius.ten-hacken@uibk.ac.at

Norbert Hornstein
nhornste@umd.edu
norbert.hornstein@gmail.com

Esa Itkonen
eitkonen@utu.fi

Ray Jackendoff
ray.jackendoff@tufts.edu

András Kertész
andraskertesz@yahoo.de

Stephan Kornmesser
ste.kornmesser@uni-oldenburg.de

Ritva Laury
ritva.laury@helsinki.fi

Géraldine Legendre
legendre@jhu.edu

Peter Ludlow
peterjludlow@gmail.com

Antonio Machicao y Priemer
mapriema@hu-berlin.de

J. Lachlan Mackenzie
lachlan_mackenzie@hotmail.com

Edith Moravcsik
edith@uwm.edu

Stefan Müller
st.mueller@hu-berlin.de

Tsuyoshi Ono
tono@ualberta.ca

Timothy Osborne
tjo3ya@outlook.com
tjo3ya@yahoo.com

Csilla Rákosi
rakosics@gmail.com

Mark Steedman
steedman@inf.ed.ac.uk

Biographical Sketches

Jenny Audring is Assistant Professor at the University of Leiden. She specializes in morphology and has written on grammatical gender, linguistic complexity, Canonical Typology, and Construction Morphology. Together with Ray Jackendoff she is developing an integrated theory of lexical representations and relations in the Parallel Architecture. A monograph (Jackendoff & Audring The Texture of the Lexicon) is forthcoming with Oxford University Press.

Cristiano Broccias is Professor of English Language and Linguistics at the University of Genoa, Italy. His research interests include synchronic and diachronic English syntax, cognitive theories of language and, more recently, figurative language. He has published a book on change constructions (The English Change Network, 2003) and various papers on cognitive approaches to grammar, resultative constructions, adverbs, *as*-constructions and metonymy.

Philip Carr graduated in 1981 with First Class Honours in Linguistics from Edinburgh University. He then completed a PhD in the philosophy of linguistics at the same university, under the supervision of Roger Lass. He taught linguistics in the English Department at The University of Newcastle-Upon-Tyne, from 1983 to 1999. He then became the Professor of Linguistics in the English Department of the University of Montpellier in France, from 1999 to 2017, where he is now Emeritus Professor. He has also taught linguistics at The University of Khartoum, The University of Texas at Austin, and The University of Canterbury at Christchurch, in New Zealand, where he was a Visiting Fellow. His first book, Linguistic Realities, on the philosophy of linguistics, was published by Cambridge University Press in 1990. He has since published books and articles on phonology, general linguistics and the philosophy of linguistics. He was joint editor of three volumes on phonology and the nature of phonological knowledge.

Rui P. Chaves is an Associate Professor at the Linguistics Department of the University at Buffalo, the State University of New York. He specializes in lexicalist and construction-based models of grammar, experimental syntax, and computational modeling, with particular focus on the syntax-semantics interface, ellipsis, coordination, and long-distance dependencies.

Peter Culicover is Distinguished University Professor of Linguistics at The Ohio State University and a Fellow of the American Academy of Arts and Sciences, the American Association for the Advancement of Science and the Linguistic Society of America. He received his Ph.D. in Linguistics from MIT in 1971. His research has been concerned with explaining why grammars are the way they are. Culicover has worked in recent years on grammar and complexity, the theory of constructions ("syntactic nuts"), the

https://doi.org/10.1515/9783110540253-202

history of the core constructions of English, and ellipsis. Recent major publications include Grammar and Complexity (Oxford, 2013) and Simpler Syntax, with Ray Jackendoff (Oxford, 2005).

Mary Dalrymple is Professor of Syntax at the University of Oxford. Her main interests are in syntax, semantics, and the interface between them. She completed her PhD in linguistics at Stanford University in 1990, and was subsequently a member of the research staff of the Xerox Palo Alto Research Center and a senior lecturer at King's College London before coming to Oxford. Her most recent book is the Oxford Reference Guide to Lexical Functional Grammar, coauthored with John Lowe and Louise Mycock (2019, Oxford University Press). She is a Fellow of the British Academy and a Member of Academia Europaea.

Sam Featherston is originally from London and studied German and Social Anthropology before becoming a teacher. While working in school he became increasingly interested in the linguistic and psychological aspects of language acquisition, so that he completed a Masters degree in Applied Linguistics and PhD in Psycholinguistics at the University of Essex. He works at the boundary of syntax and sentence processing, using data gathering methods and analytical tools from psycholinguistics and applying them to grammar research. The result has been that he has developed an individual approach to modelling the restrictions of the grammar, which blends the abstract generalizations of the generative approach with an increased attention to and respect for the empirical basis of grammar building.

Jamie Y. Findlay is a DPhil candidate in the Faculty of Linguistics, Philology and Phonetics at the University of Oxford. His current research centres on multiword expressions and their representation in the lexicon, and he is more broadly interested in syntax, semantics, and especially apparent mismatches in the mapping between them.

Pius ten Hacken is a professor of translation studies at Innsbruck University. He studied French and general linguistics in Utrecht and completed his PhD (English linguistics) and Habilitation (general linguistics) in Basel. He has worked for the machine translation project Eurotra and at universities in Basel (Computer Science, General Linguistics), Swansea (French, Translation Studies), and Innsbruck (Translation Studies with specializations in terminology and English). His current main research interests are terminology, word formation, lexicography, and the nature of language as an object of linguistic study. His monographs include Defining Morphology (Olms, 1994) and Chomskyan Linguistics and its Competitors (Equinox, 2007), and his most recent edited volumes are Word Formation and Transparency in Medical English (with Renáta Panocová, Cambridge Scholars Publishing, 2015) and The Semantics of Compounding (Cambridge University Press, 2016).

Norbert Hornstein teaches linguistics at the University of Maryland College Park. He has written extensively on minimalist themes since 1995. He has four book-length discussions including Logical Form: from GB to Minimalism, Move!: a theory of construal, and A Theory of Syntax. He also manages a blog that discusses Generative and Minimalist themes called The Faculty of Language.

Esa Itkonen (born in 1944), now retired, was 1982–2012 Professor of General Linguistics at the University of Turku (Finland) and 1988–2012 Dozent of Philosophy at the University of Jyväskylä (Finland). His research interests are philosophy of linguistics, history of linguistics, and linguistic typology. His book publications include Grammatical Theory and Metascience (1978), Causality in Linguistic Theory (1983), Universal History of Linguistics: India, China, Arabia, Europe (1991), What Is Language? A Study in the Philosophy of Linguistics (2003), Analogy as Structure and Process: Approaches to Linguistics, Cognitive Psychology, and Philosophy of Science (2005), Ten Non-European Languages: An Aid to The Typologist (2005), The Diversity and the Unity of the World's Languages, 3rd edn, in 3 vols (in Finnish), 2008–2010, Papers on Typological Linguistics (2011). He is a co-editor of The Shared Mind: Perspectives on Intersubjectivity (2008).

Ray Jackendoff is Seth Merrin Professor of Philosophy Emeritus and former Co-Director of the Center for Cognitive Studies at Tufts University. He is known for his research on Conceptual Semantics, the Parallel Architecture, and Simpler Syntax, as well as for groundbreaking work on musical cognition. He was the 2003 recipient of the Jean Nicod Prize in Cognitive Philosophy and the 2014 recipient of the Rumelhart Prize, and he is past President of both the Linguistic Society of America and the Society for Philosophy and Psychology. Among his books are Semantics and Cognition (1983), A Generative Theory of Tonal Music (1983, with Fred Lerdahl), Consciousness and the Computational Mind (1987), Foundations of Language (2002), Simpler Syntax (2005, with Peter Culicover), Language, Consciousness, Culture (2007), and A User's Guide to Thought and Meaning (2012).

András Kertész is Professor of German Linguistics at the University of Debrecen (Hungary). In 2001 he was elected as a Member of the Hungarian Academy of Sciences and in 2014 of Academia Europaea (London). His main fields of research are the philosophy of linguistics and theoretical linguistics. He published 12 monographs, 9 collections of papers and about 150 articles. His monographs include Cognitive Semantics and Scientific Knowledge (2004), Philosophie der Linguistik (2004), Data and Evidence in Linguistics (2012; co-author: Csilla Rákosi) and The Historiography of Generative Linguistics (2017).

Stephan Kornmesser is a postdoc at the Department of Philosophy at the Carl von Ossietzky University of Oldenburg and at the Department of German Studies at the

Leibniz University Hannover. He works on the philosophy of science, philosophy of linguistics and philosophy of language.

Ritva Laury is professor emerita of Finnish at the University of Helsinki and professor emerita of Linguistics at the California State University. Her research has focused on the emergence of grammar from interaction, and has dealt with issues of reference, indexicality, grammaticalization, clause combining in everyday conversation, and most recently, the crosslinguistic applicability of traditional linguistic categories.

Géraldine Legendre is Professor of Linguistics and Cognitive Science at Johns Hopkins University. She co-developed, with Paul Smolensky, the soft constraint-based precursor to Optimality Theory known as Harmonic Grammar, and has played a major role in the development of OT in syntax since the early 1990's, focusing particularly on comparative studies of syntactic phenomena and early child syntax. She is the author of numerous papers on OT, the co-author of The Harmonic Mind (2006; MIT Press), and a co-editor of Optimality-theoretic Syntax (2001; MIT Press) and Optimality-theoretic Syntax, Semantics, and Pragmatics (2016; Oxford University Press).

Peter Ludlow is author of numerous publications in the philosophy of language and the philosophy of linguistics, including The Philosophy of Generative Linguistics (Oxford University Press, 2011) and Interperspectival Content (Oxford University Press, 2019). He also publishes on topics ranging from blockchain technology to hacktivism and conceptual issues in virtual worlds.

Antonio Machicao y Priemer studied Germanic Linguistics and Spanish Philology at the Humboldt-Universität zu Berlin. He worked as a research assistant at the Universität Potsdam, in the collaborative research center (CRC 632) "Information structure" at the Freie Universität Berlin, at the Eberhard Karls Universität Tübingen, and at the Humboldt-Universität zu Berlin. He wrote his dissertation about Syntax and Semantics of Noun Phrases in Spanish and German in the framework of Head-Driven Phrase Structure Grammar, and is currently working as a postdoctoral researcher at the Syntax Lab at the Institute for German Language and Linguistics (Humboldt-Universität).

J. Lachlan Mackenzie is Emeritus Professor of Functional Linguistics at VU Amsterdam, having previously been Full Professor of English Language there. With a PhD from the University of Edinburgh (1978), his career was in The Netherlands, working closely with Simon Dik, Kees Hengeveld and many others on the development of Functional Grammar and Functional Discourse Grammar. He is an Editor of the journal Functions of Language and his research interests range from functional linguistics to pragmatics and discourse analysis. Key publications: Functional Discourse Grammar (OUP 2008) and Pragmatics: Cognition, Context and Culture (McGraw Hill 2016).

Edith Moravcsik's main research area has been language typology; she has also worked on syntax, morphology (infixing, reduplication, nominal inflection), language contact, and conflict resolution in grammatical description. In addition to papers and edited volumes, she has published two textbooks on syntax (An introduction to syntax, and An introduction to syntactic theory, both 2006: Continuum) and one on language typology (Introduction to language typology, 2013: Cambridge University Press). For 33 years, she taught linguistics at the University of Wisconsin-Milwaukee; she retired in 2009.

Stefan Müller studied Computer Science, Computational Linguistics and Linguistics at the Humboldt University at Berlin and in Edinburgh. He did his dissertation and habilation in Saarbrücken and had professorships for theoretical linguistics and computational linguistics in Bremen, for German and General Linguistics at the Freie Universität Berlin and is now professor for German language with specialization in syntax at the Humboldt Universität zu Berlin. His main research topic is German grammar but he also works on other languages. He works both empirically and theoretically. Topics of interest are morphology, syntax, semantics, and information structure. He works in the framework of Head-Driven Phrase Structure Grammar (HPSG) and the theoretical analyses are implemented in computer-processable grammar fragments.

Tsuyoshi Ono is professor of Japanese language and linguistics at the University of Alberta, Canada, where he directs the Spoken Discourse Research Studio. His main area of research lies in the study of grammatical structure based on the examination of conversation data in which he and his collaborators have published widely. He is currently engaged in two large-scale collaborative projects: formulaic language in everyday spoken Japanese and documentation of the Ikema dialect of Miyako, an endangered language spoken on remote Japanese islands near Taiwan.

Timothy Osborne completed his studies at Colorado State University and at Pennsylvania State University in German with specialization in linguistics. During his studies, he spent several years in Regensburg, Flensburg, and Kiel in Germany. At Penn State, he worked for a number of years as an adjunct professor. He is currently an associate professor of linguistics at Zhejiang University in Hangzhou, China. His research efforts have been in theoretical linguistics focusing on coordination, comparatives, ellipsis, scope, predicate-argument structure, and other topics. In collaboration with Sylvain Kahane, he translated Lucien Tesnière's monumental work of 1959 Éléments de syntaxe structurale (2015, John Benjamins). Throughout his career, he has been a consistent advocate for dependency grammar.

Csilla Rákosi studied mathematics, philosophy and German language and literature at the University of Debrecen. She is senior research fellow at the MTA-DE-SZTE Research Group for Theoretical Linguistics of the Hungarian Academy of Sciences.

Her current research areas are the philosophy of linguistics, argumentation theory, pragmatics, metaphor processing, psycholinguistic experiments, and statistical meta-analysis. Her books include Data and Evidence in Linguistics (2012, with András Kertész) and The Evidential Basis of Linguistic Argumentation (2014, ed. with András Kertész).

Mark Steedman is Professor of Cognitive Science in the School of Informatics at the University of Edinburgh, to which he moved in 1998 from the University of Pennsylvania, where he taught as Professor in the Department of Computer and Information Science. He is a Fellow of the British Academy, the Royal Society of Edinburgh, the American Association for Artificial Intelligence (AAAI), the Association for Computational Linguistics (ACL), and the Cognitive Science Society (CSS), and a Member of the European Academy. In 2018, he was the recipient of the ACL Lifetime Achievement Award. His research covers a wide range of problems in computational linguistics, natural language processing, artificial intelligence, and cognitive science, including syntactic and semantic theory, and parsing and interpretation of natural language text and discourse, including spoken intonation, by humans and by machine. Much of his current research uses Combinatory Categorial Grammar (CCG) as a formalism to address problems in wide-coverage parsing for robust semantic interpretation and natural language inference, and the problem of inducing and generalizing semantic parsers from data, and in child language acquisition. Some of his research concerns the analysis of music using similar grammars and statistical parsing models.

Edith Moravcsik

1 Introduction

1 Comparing syntactic theories: why and how?

This volume presents descriptions of a selection of contemporary approaches to syntax. In this respect, it is not unique: there have been a number of comprehensive surveys of syntactic theories published recently (Heine and Narrog (eds.) 2010; Borsley and Börjars (eds.) 2011; Luraghi and Parodi (eds.) 2013; Hagemann and Staffeldt (eds.) 2014; Kiss and Alexiadou (eds.) 2015; Kertész, 2017; Müller, 2010, 2016; Bond et al. (eds.) 2016).[1] The multitude and diversity of current syntactic approaches are stunningly illustrated by Jörg Hagemann and Sven Staffeldt's introduction to their edited volume (2014), where ten alternative analyses of the same German sentence are presented, and by Stefan Müller's monumental compendiums of syntactic approaches (2010, 2016, volume I), where differing accounts of a selection of grammatical constructions (the passive, long-distance dependencies, and others) are discussed across nine contemporary approaches.

We intend to contribute to the existing literature in two ways. First, the volume facilitates systematic comparison of the theories in that each approach is described in terms of a small set of parameters that serve as anchor points for comparative surveys. Rather than addressing these points implicitly, the authors discuss them point-by-point to invite easy at-a-glance overview. For example, if the reader wants to know how the approaches differ in their goals, he/she can find "Goals" in each chapter either as a section heading or highlighted in the running text. The chapters also provide full or partial analyses of the same sample sentence or of a somewhat altered version of it.

Second, in addition to the thirteen chapters on syntactic theories, the book contains six studies on their metatheoretical foundations. For the relationship between the two goals, see the Conclusions chapter of this volume.

There are other possible goals that we have not adopted. First, the roster of current syntactic theories showcased in the volume is limited: we have not striven for com-

1 For earlier surveys, see Moravcsik and Wirth (eds.) (1980), Horrocks (1987), Baltin and Kroch (eds.) (1989), Borsley (1991), Droste and Joseph (eds.) (1991), Edmondson and Burquest (1992), Sells (1985), Brown and Miller (eds.) (1996), Darnell et al. (eds.) (1998), Baltin and Collins (2001), Moravcsik (2006). For a cross-theoretic comparison of accounts of agreement, see Bond et al. (2016). A systematic survey of contemporary morphological theories is provided in Stewart (2016).

Acknowledgement: I am grateful to András Kertész, Stefan Müller, Csilla Rákosi, and Pius ten Hacken for their detailed and constructive comments; and I wish to express my special gratitude to András and Csilla, whose extensive knowledge of syntactic theories and metatheories and their unceasing willingness to discuss the relevant issues with me over the years have provided the foundation of this paper. Section 3 *Parameters of syntactic metatheories* was written by Csilla Rákosi.

https://doi.org/10.1515/9783110540253-001

plete coverage of the near-contemporary scene and thus a number of approaches have remained unrepresented. These include Autolexical Syntax (Sadock, 1991), Semiotic Grammar (McGregor, 1997), Functional Grammar (Halliday and Matthiessen, 2004), Syntactic Carpentry (O'Grady, 2005), Role and Reference Grammar (Van Valin, 2010), and the cartographic approach to syntax (Shlonsky, 2010). Second, we have not taken on the task of evaluating the alternative approaches against each other. On the benefits of the diversity of approaches, see Ludlow (this volume).

In view of the volume's focus on comparison, two questions arise, one about the possibility of the systematic comparison of the various frameworks and the other about its desirability. Let us consider each in turn.

First, how to compare? To identify differences between two objects, some shared properties must be identified as points of reference. There are two sources of necessarily shared features of syntactic theories. One is their subject matter: they are all about syntax; the other is their conceptual status: they are theories.[2] If we assume that syntax is, minimally, about the way words are selected and linearly ordered in sentences, all syntactic theories must account for the selection and ordering of words. And, given that all scientific explorations strive, by definition, to discover overall patterns behind individual instances, the general nature of statements is also a required characteristic of syntactic theories. Other common features dictated by a scientific approach will be discussed in Sections 2.3.1 and 2.4.1 below.

In addition to necessarily shared properties, the parameters of differences among the frameworks may also be uniformly defined. Syntactic theorizing is construed here as a goal-directed activity relative to a set of data with a set of conceptual tools employed for achieving the goals and with criteria stipulated for assessing success. This concept yields the following basic parameters of variation: Goals, Data, Tools, and Evaluation. These variables are logically independent but, as pointed out to me by Andrej Malchukov, there are implicational relationships among them. Most crucially, the goals adopted are likely to affect the choice of data, conceptual tools, and evaluative criteria. (For discussion, see the Conclusions chapter, Section 5.)

While the comparison of different approaches is therefore possible, a second question looms: is it needed at all? The most important benefit is clarification. Given the bewildering variety of syntactic approaches, it is difficult to see how the various theories relate to each other – that is, in exactly what ways they are different and if they differ at all. One possibility is that two theories are alike in all ways – in the choice of goals, in the data considered relevant, in the choice of descriptive tools, and in the criteria for evaluation – and they differ only in terminology. In this case, they are notational variants: intertranslatable versions of the same approach. A second possible scenario is where the frameworks share their goals, domains of data, and evaluative

2 For lists of ideas that all linguists may necessarily agree on, see Hudson (1981) and Riemsdijk (1984). A critical discussion of Riemsdijk's paper is provided in Kiss and Alexiadou (2015).

criteria but they opt for different conceptual tools to achieve the shared goals. In this case, the theories are different and incompatible: they are in direct competition due to conflicting hypotheses and conflicting predictions. A third possibility is that, while the conceptual tools may be similar, the theories differ in goals, and/or in choice of data, and/or in criteria for evaluation. If so, the theories are again different but complementary: they shed light on different facets of the same object under study. More on this in ten Hacken (this volume) and the Conclusions chapter, Section 5.

Identifying the types of differences among theories should provide a clear view of the land. It should also serve to break down barriers, foster communication and cooperation among researchers, and focus discussion on substantive issues across theories. In the words of Kiss and Alexiadou, it is useful "to observe what the other camp is doing" and "to attack problems from different vantage points" (2015, 14). Understanding the relationships among the frameworks directs research and helps to decide what we should teach to our students. The same considerations hold for the usefulness of comparing different metatheories.

A few notes on terminology. In what follows, the terms "theory", "approach" and "framework" will be used interchangeably. By "metatheory", we mean approaches that take syntactic theories as their objects of study – rather than linguistic phenomena as syntactic theories do – and they do this generally in the context of philosophy of science. Thus, a metatheory may analyze the structure of syntactic theories, the methodological tools applied, the data-handling techniques, and the often implicit assumptions that underlie them.

So far, it was suggested that the comparison of syntactic theories and metatheories, which this volume strives to facilitate, is both possible and useful. The remainder of this introduction details the framework for systematic comparison that has been adopted for the volume (Sections 2 and 3) followed by sample comparisons (Section 4) and conclusions (Section 5).

2 Parameters of syntactic theories

2.1 Goals

2.1.1 Necessary congruities

As noted above, I assume that any syntactic description has the minimal goal of describing the selection and the linear ordering of words within sentences – structures often labeled "surface syntax" – and to achieve this through statements of some generality.[3] For example, given the sentence *These cherries are sweet*, the grammar should provide for the proper selection of the Demonstrative, the Noun, the Verb, and

3 On the problems of defining "word" crosslinguistically, see Dixon and Aikhenvald (2003) and Haspelmath (2011).

the Adjective including the choice of the particular forms of the Demonstrative and the Verb as required by agreement, and the linear order of these components. This notion is widely but not universally accepted. In Minimalism, linear order is viewed as part of phonetics and Dependency Grammar is only marginally concerned with patterns of linearization. Nonetheless, it will serve as the jumping-off point for this introductory paper.

2.1.2 Possible differences

Beyond this minimal goal, a theoretician may opt to study sentence structure within a broader context. There are at least five possible directions for the expansion, which open avenues not only for enriched descriptions but also for explanations.

1) Interface relations
   While syntactic structure can be described in isolation from semantics and phonetics, it may also be viewed as mediating between meaning and phonetic form; that is, in reference to interface relations.
2) Discourse and situational context
   A sentence is generally used as part of a sequence of sentences: a discourse. Discourses in turn take place in given situations accompanied by body language. The description of syntactic structure may thus be placed in discourse and situational context.
3) Crosslinguistic context
   A sentence may be analyzed within a single language variety, such as a register or a dialect or a language as a whole. It may also be viewed in comparison with other varieties of the language and with other languages.
4) Cognition
   Syntax may be viewed in isolation from other cognitive processes or in its relationship to general cognition.
5) Processes
   Sentences originate from somewhere. There are three processes that shape their grammar. One is usage: a syntactic pattern may be altered by how it is used. Another is acquisition: a syntactic pattern is what it is because of how it is acquired in first or second language. A third process is diachronic change: how a pattern has evolved in the history of the language through the processes of acquisition and usage. Syntacticians may opt for analyzing sentence structure in the context of usage, acquisition, and historical change and thus aim at causal explanations for why syntax is the way it is.

In addition to these five kinds of contextualization, syntactic theories may aim at practical applications, such as solving problems of language pedagogy, language pathology, language planning, automatic parsing, or machine translation.

2.2 Data

2.2.1 Necessary congruities

The goals of a theory necessarily delimit the domain of data relevant to it. (For a discussion about the relationship of goals and data, see Ludlow (this volume).) Given that syntax is, minimally, about the selection and linear order of words in sentences, the primary data relevant to syntactic theorizing must include information about (grades of) the well-formedness and ill-formedness of sentences of the language under study.

2.2.2 Possible differences

Two main variables are how well-formedness is defined and what sources of data are considered relevant.

First, how is well-formedness determined? Is well-formedness construed as categorical or scalar?[4] Is the grammar supposed to account only for grammatical sentences or, more generally, for acceptable sentences some of which may be ungrammatical? Are the data "cleaned up" – that is, freed of performance factors – or are they taken "as they come"? An extensive analysis of the gradience of well-formedness and the factors that affect it is offered in Featherston (this volume).

Second, the data chosen for a theory may be obtained in different ways. Possible sources are introspection, elicitation, corpora, experimentation, and existing descriptions. Some approaches allow for multiple sources – whether by principle or tacitly in actual practice – while others rule out some of them.

2.3 Tools

2.3.1 Necessary congruities

Given the content of syntax and the scientific nature of theorizing about it, certain things follow for the conceptual tools employed in syntactic descriptions. Chomsky refers to these as conceptual necessities (e. g. 1995, 169). Below, they are construed as pertaining to the terms used (1/ below), the relationships posited among the terms (2/ below), and the nature of the statements that descriptions consist of (3/ below).

1) Terms
   No syntactic description can simply list sentences. On the one hand, it is impossible to enumerate the possibly infinite number of sentences that a language consists of and, on the other hand, no generalizations would emerge from such an

4 For arguments against the very notion of grammaticality, see Sampson and Babarczy (2014).

attempt. To insure generalizability, an indispensable step in syntactic description is segmentation: cutting sentences into parts resulting in what is known as constituent structure. Minimally, the chunks are words. Furthermore, syntactic statements cannot refer to individual words such as *dog* or *kindly*: the constituent parts need to be categorized, such as nouns or adverbs, to capture similarities among them. These two conceptual operations – **segmentation** (positing part-whole relations) and **categorization** (positing token-type relations) – appear to be universally employed and necessary tools of all scientific analyses.[5] They are term-creating steps yielding the vocabulary of generalizations.

2) Relations among terms
In addition to partonomic (meronomic) and taxonomic relations, there are two other relations that necessarily figure in syntactic descriptions. One is **selectional dependency**. Given that a syntactic account must describe the choice of words in well-formed sentences, it must specify which words depend on the choice of which other words. Second, assuming that a syntactic account describes not only what words can be selected to make a sentence but also in what order they need to be in, **linear precedence** is another relation necessarily to be dealt with.

3) Statements
As noted before, it is by definition that statements of a syntactic theory must be generalizations.

2.3.2 Possible differences

Besides the required kinds of terms, relations among the terms, and general statements, there is room for variation in each of these three areas.

1) *Terms*
Different conceptions of the terms of a syntactic account may have to do (A) with constraints on segmentation, (B) with constraints on categorization, and (C) with choice of modality.
   (A) Segmentation
   The simplest partonomic (constituent) structure is two-level, binary, with each part belonging to only one directly superordinate part, and the parts being equal. Deviations may be as follows:
      (a) More than two levels
      A partonomy may be single-level or multi-level, with subparts analyzed as having further subparts, such as a sentence split into clauses, clauses into phrases, and phrases into words.

5 For a comprehensive encyclopedia of the role of part-whole relations in all areas of science, the arts, and social and individual thought, see Burkhardt et al. (eds.) (2017).

(b) More than two parts
Splits may be required to be binary or wholes may be allowed to consist of more than two directly subordinate parts, such as if a coordinate structure linking three conjuncts is analyzed as a tripartite split.

(c) More than one directly superordinate whole
Parts may be analyzed as belonging to one or to more than one directly superordinate whole, as in the raising sentence *I expect him to return*, where *him* may be viewed as belonging to two clauses.

(d) Parts unequal
The immediate components of a phrase generally differ categorially, such as adjective and noun. Beyond the categorial differences, cross-categorial asymmetries may also be posited with phrases having heads and dependents, such as a verb phrase having the noun as the head and the adjective as a dependent.

(B) Categorization
Parallel options hold in taxonomies. The simplest taxonomy is two-level, binary, with each item assigned only to one directly superordinate type, and the subtypes equal. Here are possible variants.

(a) More than two levels
Categorization may be single-level or multi-level, with sub-categories having further sub-categories (e. g. nouns being common nouns and proper nouns and common nouns being count and mass).

(b) More than two subtypes
The subdivision of categories may be binary or a category may have more than two subtypes (e. g. verbs being intransitive, transitive, and bitransitive).

(c) More than one directly superordinate type
An item may be viewed as belonging to two or more different categories, such as adjectives being both nominal and verbal.

(d) Subtypes unequal
The subtypes of a type generally differ categorially, such as a manner adverb and a time adverb. Beyond the categorial differences, cross-categorial asymmetries may also be posited with a type having a member that is more representative of the category (prototypical, or unmarked) and another member that is less so (peripheral, or marked).

(C) Modality
Syntactic terms may also vary in whether they have both meaning and phonetic form assigned to them. They are bimodal if they are represented with both meaning and phonetic form; they are monomodal if they only have one of the two. A framework may allow for “empty” categories that have phonetic form but lack meaning and for “zero” categories that in turn have meaning but no audible form.

2) *Relations among terms*
   In addition to the term-forming relations of partonomy and taxonomy and the relations of selectional dependency and linear order that are necessary parts of any syntactic description, other relations may also be posited. Examples are grammatical relations such as subject and object, and information-structural relations such as topic and focus. Additional relations may also be defined with reference to constituent structure, such as c-command, government, binding, or chains. Categories may be derived from other categories or may be taken as primitives.
3) *Statements*
   Syntactic statements may differ in whether (A) they posit different levels of syntactic analysis and (B) how they relate to other components of grammar.
   (A) Levels of syntactic analysis
       A conspicuous difference among syntactic theories is the presence or absence of multiple levels. Some analyses propose that syntactic structure can be described on a single level. In case multiple levels are posited, their connections may be directional requiring the transformation of one representation to the other, or they may be non-directed correspondences. In a transformational (derivational) framework, structure is built incrementally while in single-level descriptions, a single structure is given and it stays invariant. In derivational frameworks, there may be constraints on how a structure can be changed (e. g. monotonicity).
   (B) Relation to other components of grammar
       Two variables have to do with the interaction among the components and with similarities among them.
       (a) Interaction among the components
           Everything in grammar contributes to relating meaning and phonetic form but the division of labor among the components may be adjudicated differently. Syntax may be relieved of the responsibility of constraining both the selection and the linear ordering of words if some of the task is transferred or shared by other parts of the grammar. Here are some possibilities.
           – Is syntax semantics-free? Is the choice and order of words independent of meaning?
           – Is syntax phonology-free? Is the choice and order of words independent of phonetic form (segmental and suprasegmental)?
           – Is semantics syntax-free? Is the meaning of a sentence independent of the choice and order of words?
           – Is phonetics syntax-free? Is phonetic form – segmental and suprasegmental – independent of syntax?
           – What is the role of the lexicon in constraining the proper choice and ordering of words?
           – How does morphology relate to syntax?

(b) Similarities across components
Grammatical statements in semantics, syntax, phonology, morphology, and the lexicon may be couched in the same form differing only in the terms and relations referred to. Alternatively, statements of the different components may have their own distinct structural features.

2.4 Evaluation

2.4.1 Necessary congruities

In addition to generality, all syntactic theories need to be assessed in terms of the cross-scientific criteria of empirical adequacy, consistency, and simplicity. These criteria, however, leave plenty of room for different interpretations.

2.4.2 Possible differences

1) Empirical adequacy
Do all statements in a grammar have to be surface-true – i. e., to correspond directly to observable properties of sentences? Or is it only the entire description that should be in line with the data? Are there structures or rules posited for a language which never surface? What is the difference between something being part of a structure but suppressed, as opposed to it not being there at all? This is the issue of abstractness: different frameworks may allow different degrees of abstract representations.
2) Consistency
What conflicts – inconsistencies – arise in a particular theory and how are they dealt with? Is inconsistency tolerated, and if so, under what conditions and how?
3) Simplicity
Economy is a central motivation in grammar writing. What should be simple? The individual rules? Or the entire account? And how is simplicity measured?

3 Parameters of metatheories (written by Csilla Rákosi)

The two main issues are the relationship between a syntactic metatheory and the philosophy of science and the significance of metatheoretical reflections for object-theoretical linguistic work.

3.1 Necessary congruities

Chomsky and several generative grammarians seem to accept the norms of Popperian falsificationism. This is the most influential and widespread branch of the standard view of the analytic philosophy of science which, however, has not only been questioned but, starting in the 1960-s, has been viewed as obsolete and generally abandoned in favor of a pluralistic view. Therefore, metatheories must reflect on the role and applicability of this approach and check whether diverse syntactic theories have in fact been constructed along the principles that they claim to follow.

A second common feature of metatheories is that they have to involve a descriptive component, i. e. tools for the reconstruction of the structure of theories and/or the process of linguistic theorizing.

3.2 Possible differences

The reaction to Popperian falsificationism may be twofold: rejection or acceptance. Reasons for the first decision may be different. One alternative is advocating methodological autonomy, or as Raffaele Simone coins it, fulfilling "Saussure's dream", according to which one should "provide linguistics with an appropriate method, one not borrowed more or less mechanically from other sciences, but designed to be peculiarly and strictly of its own" (Simone, 2004, 238). A second possibility is the application of some other post-Kuhnian metatheoretical approach to the description of syntactic theories and/or research activities. A third possible method is opened up by contemporary philosophy of science's rejection of the idea of general, uniform norms for scientific theorizing. Instead, a close and fruitful cooperation has begun between philosophers of science and researchers working on different branches of science. As a result, there is a view in current philosophy of science that the philosophy of special sciences has actually become fused with the theoretical sciences themselves (Machamer, 2002).

Views about the object and role of metatheoretical reflection in syntax may also diverge. First, a metatheory may strive not only for description but may also contain a normative component. Thus, a metatheory may

- argue that there are some basic, general methodological rules based on a priori criteria that all empirical sciences should apply;
- try to elaborate methodological rules specific for linguistics that should be applied by all syntactic theories;
- provide tools for the evaluation of theories on the basis of the continuous comparison and conciliation of methodological rules of other branches of science and linguistics; and
- define methodological guidelines on the basis of the study of successful research praxis of linguists.

Second, a metatheory may narrow down its focus to the structure of object-scientific theories, or it may extend its scope to the whole process of syntactic theorizing.

4 Comparison by three parameters

Partial comparisons among syntactic approaches are offered in almost all the chapters in both parts of the volume. The possibility of comprehensive comparisons is in turn facilitated by the fact that the four parameters – Goals, Data, Tools, Evaluation – are highlighted in the individual descriptions throughout the first part of the volume; some of them are also discussed in the metatheoretical papers in the second half. In what follows here, I will take up three parameters and, without attempting full coverage, provide an overview of how they are construed by some of the theories represented in the volume.

Each of the three topics has to do with the place of a smaller domain within a broader one. The first, mentioned under *Tools* above (Section 2.3.2), is the internal organization of a syntactic description. It deals with how various aspects of syntactic structure are accommodated in a description by employing the tool of analytic levels. The second and third are listed under *Goals* above (Section 2.1.2): how syntax fits into the entirety of grammar, and, more broadly, how the knowledge of grammar in turn relates to other cognitive abilities.

4.1 The internal organization of syntax: levels

As mentioned in Section 2.3.2, one of the ways in which syntactic descriptions differ from each other is whether they are monostratal – i. e. they posit a single level of structural representation – or multistratal, positing more than one level. Every framework that involves a derivation necessarily contains levels: information is changed in the course of the derivation; but in some cases, there are dedicated levels that are multiply motivated and labeled.

For example, distinct levels are assumed in earlier versions of Generative Grammar (D-structure and S-structure). For an overview of the evolutionary stages of generative grammar, see Kornmesser (this volume). Lexical-Functional Grammar also has designated levels: c-structure and f-structure. C-structure represents constituent structure and linear order; f-structure is unordered, it is about grammatical functions, long-distance dependencies and other patterns. The relation between the two is not derivation but correspondence: they are independent planes of grammatical organization.

Other approaches presented in the volume also posit representational levels, such as Simpler Syntax's Grammatical Function tier (Culicover and Jackendoff, this volume, Section 3.4. *Grammatical functions; control and raising*) and Functional Discourse Grammar (Mackenzie, this volume). In their paper on Head-Driven Phrase Structure Grammar, Stefan Müller and Antonio Machicao y Priemer explicitly state that levels like Deep Structure and Surface Structure are not part of the analysis but they say the Argument Structure list is similar to Deep Structure (Section 1.3. *Tools*). Some versions of Dependency Grammar have two levels, while others are monostratal (Osborne, this volume). In Usage-Based Grammar, the number of levels is left open for empirical work to decide (Laury and Ono, this volume). This is an example of an interesting parameter of variation among theories: whether something is assumed – i. e. taken for certain and therefore not subject to testing – or whether it is hypothesized: possible but not certain depending on evidence.

The employment of levels of analysis is common among syntactic theories. Why? What is the utility of positing levels of analysis? A revealing answer is suggested by Rákosi. In talking about an early generative analysis of oblique question words in English (such as in *Who did you ask?*) that appear sentence-initially but by syntactic relations belong to a post-verbal position, Rákosi writes: "It is with the purpose of resolving a conflict that Chomsky introduced transformations that separate the two contradictory bits of data: in deep structure, the question word is still after the verb but in surface structure, it is moved to the beginning of the sentence" (Rákosi, 2005, 181; translated). Another example of how levels address contradictory data is given in Dalrymple and Findley (this volume, Section 4.3. *Mismatches*). Take the phrase *these floods*. There is a structure-function mismatch here: two words but a single function. Representing this state of affairs in a single statement with the conflicting information inseparable would be contradictory. However, if structure and function are separately represented each on its own level as done in Lexical Functional Grammar, each representation remains internally consistent. The conflict is now converted into one holding between the two levels of structure and function; but this conflict is multiply documented and thus to be acknowledged as a general feature of language.

These examples demonstrate that resorting to levels of analysis is a conflict-resolving measure in the spirit of paraconsistent logic. More is said about this below in Kertész and Rákosi (this volume, Section 2.2. *On property* (*ii*)*: The tolerance of inconsistency*) and in Section 5 *Conclusions* of this introduction.

If a framework posits no levels, the question is how contradictory information is represented? How is, for instance, contradictory evidence about the position of a constituent described if not by movement? In these cases, other descriptive tools are recruited to take over the task. In the Minimalist Program, Hornstein (this volume) says that the operation of Merge itself generates "structures with the properties of movement" (Section 2. *Tools and particulars*). In Combinatory Categorial Grammar, Steedman (this volume) shows that the lexicon and combinatory rules are the mechanisms employed for this purpose. In Construction Grammar, Chaves (this volume) illustrates how varying positions of an item can be accounted for by the alternatives given in the lexicon (Section 3.2. *Phrasal constructions that contribute meaning*).

4.2 The internal organization of grammar: components

As noted in Section 2.1.2, one of the optional goals of syntactic theories is to explore how syntax is related to other components of grammar. Components – such as syntax, semantics, and phonology – are similar to levels. The difference is that they are necessarily distinct because of the distinctness of the terms that figure in the generalizations of each level. Phonological terms such as syllable are not referred to in syntactic and semantic structure and similarly, argument structure is not relevant to phonology.

Although all grammars will posit at least these three components, there are variations among frameworks in at least three respects. First, the components are delimited differently: there may be blurriness at the boundaries. For example, in the Minimalist Program, linear order is regarded as part of phonology. Frameworks differ also in what is to be considered a syntactic pattern and what is a semantic one. The basic principle of Simpler Syntax is to minimize syntax and thus semantics (as well as phonology) has a larger share in descriptions than in other frameworks. Usage-based Grammar takes an altogether dim view of components: practitioners do not assume their existence pending on robust empirical evidence (Laury and Ono, this volume, Section 4.2. *Components, levels, categories and relations*).

Second, the nature of the rules in the various components may be different. In some frameworks, the structure of the rules is uniform across components, such as in Optimality Theory (Legendre, this volume), Cognitive Grammar (Broccias, this volume), and Functional Discourse Grammar, where the template structure of syntactic and morphological rules is the same (Mackenzie, this volume).

A third difference has to do with the derivational priority of the components. Generative Grammar, including the Minimalist Program, is syntax-centric: the only formational rules of the grammar are in syntax, with semantics and phonology interpretive. In contrast, all three components are taken to have their own formation rules in Simple Syntax and Parallel Architecture.

The status of the lexicon as a component also differs across theories. Lexical rules are taken to be distinct from syntactic ones in several frameworks while in Cognitive Grammar, both types of rules express symbolic equivalences sharing the function of relating form and meaning. This is also a foundational property of Gerald A. Sanders' *Equational Grammar* (1972, 1975), a little-read and soon-forgotten approach, some of whose basic principles came to be independently proposed later and now prominently figure in several contemporary theories. These include the insight that syntactic, lexical, and phonological rules are of the same ilk in that they all express symbolic equivalence relations between meaning and form; that rules of syntactic structure and linear order must be separately formulated; that linear order should be recognized as a feature of phonetic form; that statements about linear order should be surface-true and thus invariant (i. e. not subject to movement); and that the discourse, rather than the sentence, should be the proper domain of grammatical descriptions.

4.3 The internal organization of cognition: modularity

As also noted in Section 2.1.2, apart from identifying the proper relationship between syntax and the other grammatical components, another possible theoretical goal is to determine how grammar as a whole relates to other aspects of human cognition. The question is whether linguistic knowledge is an integrated part of cognition or whether it is a special module with principles distinct from and possibly even in conflict with those of other aspects of cognition. The logical options are these:

(a) Linguistic knowledge is fully domain-specific.
(b) Linguistic knowledge is fully domain-general.
(c) Linguistic knowledge is partly domain-specific and partly domain-general.

In each case, a further question is about specifics: the particular domain-specific and domain-general abilities posited. Let us review the three options and the specific abilities proposed.

(A) Full domain-specificity
The only framework that posits a domain-specific factor is the Minimalist Program. Hornstein (this volume) highlights the analytic significance of the operation Merge that puts two things together to make a third thing. He demonstrates how Merge unifies and thus explains several previously disparate theoretical domains of generative grammar, such as Control and Case. However, while Merge is presented as part of the Faculty of Language, or Universal Grammar, it is quite plausible that Merge-like cognitive operations play a role outside language as well. This issue is independent of the details of syntactic description: were Merge found to be domain-general, this would not invalidate grammatical analyses in terms of Merge. This possibility is not excluded by Hornstein, nor is the role of domain-general abilities altogether rejected (i. e., Chomsky's "third factor", e. g. (Chomsky, 2011, 263)).
In his metatheoretical paper, Carr (this volume) does not accept innate syntactic knowledge. Following Itkonen (this volume) and Michael Tomasello, he says syntax is based on social conventions that are "internalized in individual brains" (Section 5. *On the acquisition of grounded syntactic knowledge*). Thus, acquiring syntax is a domain-general process of skill learning based on understanding other people's intentions. Nonetheless, Carr says, there may be a syntax module emergent from domain-general knowledge.

(B) Full domain-generality
None of the frameworks presented here explicitly exclude the possibility of domain-specific knowledge. The papers discuss some linguistically relevant components of domain-general knowledge without denying the possibility of domain-specificity in other respects. For example, Construction Grammar posits general

cognitive mechanisms without explicitly rejecting the existence of language-specific genetic endowment. Thus, this framework and all the others in the volume may cautiously be taken to subscribe to the third option.

(C) Both domain-specific and domain-general factors
Simpler Syntax is explicit about the issue. Rather than being committed to eliminating domain-specific factors, the goal is to minimize them (Culicover and Jackendoff, this volume, Section 1 *Goals*). Parallel Architecture (a general framework that includes Simpler Syntax) is interested in what is in people's memory about language and how it is used (Jackendoff and Audring, this volume, Section 1. *Goals*). The question of what the mental resources are that belong specifically to the language faculty and what are aspects of more general mental phenomena is taken to be an empirical issue. The scope of the framework includes music, visual cognition, and other non-linguistic mental faculties. Similarly, Optimality-theoretic Syntax focuses on domain-general processes but allows for domain-specific representations as well. Usage-Based Grammar investigates language as embodied action that involves non-verbal behavior such as visual cues (gaze direction). In this framework, there is a basic stance against compartmentalized views of linguistic knowledge and while most usage-based analysis do not assume a domain-specific linguistic component, this possibility is not excluded.

The particular domain-general principles cited in the papers include the following:

- The Same-Except principle (Simpler Syntax)
- Association, Automatization, Construal, and Categorization (Cognitive Grammar)
- Unification, stored schemata, categorization, imitation, social interaction (Parallel Architecture)
- Constraints interaction, conflict resolution, and optimization (Optimality-theoretic Syntax)
- interactive skills, imitation, social interaction, social conventions (Cognitive Grammar, Usage-Based Grammar).

In sum, in most frameworks, there is a healthy uncertainty as to the existence of a genetic language-specific component of cognition. It is taken to be an empirical issue that is independent of the validity of individual linguistic analyses. With the exception of The Minimalist Program, the primary thrust of research is the identification of the language-relevant portions of domain-general endowment.

5 Conclusions

This paper articulated the goals of the volume: a systematic presentation of a set of contemporary syntactic theories by defining parameters of comparison, applying these parameters consistently and conspicuously across the chapters, and probing into their metatheoretical foundations. Both the desirability and the possibility of comparison were argued for.

In what follows, I would like to highlight an additional, global parameter of theory comparison: conflict resolution.

It is revealing to view grammatical theorizing from the point of view of how conflicts are resolved. In his discussion of English tense marked on the verb but pertaining to the entire proposition, Ray Jackendoff remarked: "Much dispute in modern syntax has been over these sorts of mismatch and how to deal with them. (I don't think most linguists have viewed it this way, though.)" (Jackendoff, 2002, 15). Even before 2002 (e. g. the resolution rules of agreement in Corbett, 1983) but particularly since then, the topic of mismatches has been explicitly addressed in morphosyntactic research; e. g. Francis and Michaelis (eds.) (2003), Corbett (2006), and MacWhinney et al. (eds.) (2014). Of the theories presented in this volume, conflict resolution is discussed most centrally in Optimality-theoretic Syntax (Legendre, this volume), Rákosi (2014) and in the P-model (Kertész and Rákosi, this volume).[6]

Conflicts that arise in grammatical analysis are of several kinds: a constituent may seem to be both present and absent, belonging both to one category and to another, and being both in one position and in a different one. There are several possible ways of accommodating them.

For resolving conflicts in general, the most obvious logical possibilities are few in number: given A and B that are in conflict, either one trumps the other, or a compromise is reached through some kind of a hybrid solution, or both are given up.

Override: A prevails over B
Compromise: A and B both prevail in part
Deadlock: neither A nor B prevails

There is nonetheless a fourth option as well. Consider a dispute over child custody. Override means one parents gets the child; Compromise involves some kind of jointly negotiated decisions-making with one or both demands softened; and if neither parent gets the child, there is a Deadlock. But there is a further possibility: each parent's rights are assigned to a separate domain. These domains may be defined partonomically: each parent has the child for part of the time. They may also be defined taxonomically, with one parent taking care of educational issues and the other of health care for the child. This solution, which may be labeled Separation, is legitimized by paraconsistent logic, according to which a conflict within a single domain can be resolved if each of the conflicting factors is relegated to a separate domain. Separation has the flavor of Compromise; yet it is distinct from it since it does not involve chang-

6 On conflict resolution in grammar, see also Rákosi (2005), Kertész and Rákosi (2006, 2012, 2013), and Moravcsik (1993, 2006, 2010).

ing the content of one or both of the conflicting entities; instead, it means restricting their scope of validity. It provides for separate and peaceful co-existence of the conflicting factors. Separation as a conflict-resolving strategy highlights the importance of the two conceptual tools: partonomy and taxonomy.[7] The domain within which a conflict obtains is split into two either by dividing it into the subparts or by sorting it into subtypes.

The various types of resolution crop up repeatedly across the theories of this volume. Optimality-theoretic Syntax documents all three of the basic logical possibilities. This framework is founded on the recognition – also evident in countless debates in individual, social, political and ideological debates – that when conflicts arise, it is not always because the relevant factors differ but, rather, due to the varying importance attributed to the same factors. While the theory focuses on Override scenarios, the two other logical possibilities are also illustrated: Compromise – when of two constraints, both have some say on the outcome – and, according to some analyses, Deadlock, labeled No Parse, or Ineffability (Legendre, this volume, Section 3. *Tools* (*with sample analyses*)).

The fourth option – Separation – is widely resorted to in syntactic theories (cf. Moravcsik, 2010). As shown above in Sections 4.1 and 4.2, the very assumption of levels and components is based on paraconsistent logic: a self-contradictory statement may be eliminated if the conflicting factors are separated into distinct domains.[8]

The conflict-resolving tools in grammatical analysis are widely paralleled by strategies invoked in science and everyday life. When confronted with a conflict, choosing one option, striking a compromise between two options, declaring each as belonging to a different domain, or simply walking away from the problem are ubiquitous phenomena in all walks of life.[9] Tools of conflict resolution are therefore prime candidates for being part of domain-general cognitive endowment.

The analysis of the different grammatical metalanguages demonstrated in this volume is akin to the study of natural languages. Speakers of a language face the challenge of how to express a thought – a process fraught with conflicting desiderata. Grammarians in turn face the challenge of how to analyze how people express thoughts. Different languages can thus be viewed as varying records of how speakers have come to terms with conflicts attendant to linguistic expression. Analogously, grammatical descriptions are records of how linguists are dealing with conflicting

7 For a comprehensive encyclopedia documenting the role of whole-part relations in science, art, and social and individual discourse, see Burkhardt et al. (2017). For part-whole relations in syntax, see Moravcsik (2009).

8 For more on paraconsistent logic, see Kertész and Rákosi (this volume) especially Section 2.2. *On property* (*ii*)*: The tolerance of inconsistency* and literature cited there, and Kertész and Rákosi (2013).

9 For an accessible overview of coping with conflicting data in physics, see Crease and Goldhaber (2014).

data about languages. In this sense, the chapters of this book offer primary data for a relatively new field of research studying the cognitive armamentarium of scientific argumentation, known as the cognitive science of science (Rouse, 1998; Kertész, 2004, 29–32).

References

Baltin, Mark & Anthony S. Kroch (eds.). 1989. *Alternative conceptions of phrase structure*. Chicago, London: The University of Chicago Press.

Baltin, Mark & Chris Collins (eds.) 2001. *The handbook of contemporary syntactic theory*. Oxford: Blackwell.

Bond, Oliver, Greville G. Corbett, Marina Chumakina & Dunstan Brown (eds.). 2016. *Archi. Complexities of agreement in cross-theoretical perspective*. Oxford University Press.

Borsley, Robert D. 1991. *Syntactic theory: a unified approach*. London: Edward Arnold.

Borsley, Robert D. & Kersti Börjars (eds.). 2011. *Non-transformational syntax: formal and explicit models of grammar*. Chichester: Wiley-Blackwell.

Brown, Keith & Jim Miller (eds.). 1996. *Concise encyclopedia of syntactic theories*. New York, Oxford: Pergamon.

Burkhardt, Hans, Johanna Seibt, Guido Imaguire & Stamatios Gerogiorgakis (eds.). 2017. *Handbook of mereology*. München: Analytica.

Chomsky, Noam. 1995. *The minimalist program*. Cambridge: The MIT Press.

Chomsky, Noam. 2011. Language and other cognitive systems: What is special about language? *Language Learning and Development* 7. 263–278.

Corbett, Greville G. 1983. Resolution rules: agreement in person, number, and gender. In Gerald Gazdar, Ewan Klein & Geoffrey K. Pullum (eds.), *Order, concord and constituency*, 175–206. Dordrecht: Foris.

Corbett, Greville G. 2006. *Agreement*. Cambridge University Press.

Crease, Robert P. & Alfred Scharff Goldhaber. 2014. *The quantum moment. How Planck, Bohr, Einstein, and Heisenberg taught us to love uncertainty*. New York: W.W. Norton and Company.

Darnell, Michael, Edith Moravcsik, Frederick J. Newmeyer, Michael Noonan & Kathleen Wheatley (eds.). 1998. *Functionalism and formalism in linguistics* (Studies in Language Companion Series 41). Amsterdam/Philadelphia: Benjamins.

Dixon, R. M. W. & Alexandra Aikhenvald. 2003. *Word: a cross-linguistic typology*. Cambridge University Press.

Droste, Flip G. & John E. Joseph (eds.). 1991. *Linguistic theory and grammatical description* (Current Issues in Grammatical Theory 75). Amsterdam: Benjamins.

Edmondson, Jerold A. & Donald A. Burquest. 1992. *A survey of linguistic theories*. Dallas, TX: The Summer Institute of Linguistics.

Francis, Elaine J. & Laura A. Michaelis (eds.). 2003. *Mismatch: form-function incongruity and the architecture of grammar*. Stanford, CA: Center for the Study of Language and Information.

Hagemann, Jörg & Sven Staffeldt (eds.). 2014. *Syntaxtheorien. Analysen im Vergleich*. Tübingen: Stauffenburg.

Halliday, Michael A. K. & Christian M. I. M. Matthiessen. 2004. *An introduction to Functional Grammar*, 3rd edn. London: Arnold.

Haspelmath, Martin. 2011. The indeterminacy of word segmentation and the nature of morphology and syntax. *Folia Linguistica* 45. 31–80.

Heine, Bernd & Heiko Narrog (eds.). 2010. *The Oxford Handbook of Linguistic Analysis*. Oxford University Press.

Horrocks, Geoffrey. 1987. *Generative grammar*. London: Longman.
Hudson, Richard. 1981. Some issues on which linguists can agree. *Linguistics* 17. 333–343.
Jackendoff, Ray. 2002. *Foundations of language. Brain, meaning, grammar, evolution*. Oxford University Press.
Kertész, András. 2004. *Cognitive semantics and scientific knowledge. Case studies in the cognitive science of science*. Amsterdam/Philadelphia: John Benjamins.
Kertész, András. 2017. *The historiography of generative linguistics*. Tübingen: Narr Francke Attempto.
Kertész, András & Csilla Rákosi. 2006. Inconsistency and plausible reasoning in the analysis of German affricates: a case study in the philosophy of linguistics. *Language Sciences* 26(4). 386–423.
Kertész, András & Csilla Rákosi. 2012. *Data and evidence in linguistics. A plausible argumentation model*. Cambridge University Press.
Kertész, András & Csilla Rákosi. 2013. Paraconsistency and plausible argumentation in generative grammar. A case study. *Journal of Logic, Language and Information* 22. 195–230.
Kiss, Tibor & Artemis Alexiadou. 2015. Introduction: syntax in linguistics. In Kiss & Alexiadou (eds.), *Syntax – Theory and Analysis, Volume I*, 1–15.
Kiss, Tibor & Artemis Alexiadou (eds.). 2015. *Syntax – theory and analysis. An international handbook* (Handbücher zur Sprach- und Kommunikationswissenschaft, 42.1) Berlin: De Gruyter Mouton.
Luraghi, Silvia & Claudia Parodi (eds.). 2013. *The Bloomsbury companion to syntax*. London, New Delhi: Bloomsbury.
Machamer, P. 2002. A brief historical introduction to the philosophy of science. In Peter Machamer & Michael Silberstein (eds.), *The Blackwell Guide to the Philosophy of Science*, 9–11. Malden & Oxford: Blackwell.
MacWhinney, Brian, Andrej Malchukov & Edith Moravcsik (eds.). 2014. *Competing motivations in grammar and usage*. Oxford University Press.
McGregor, William B. 1997. *Semiotic Grammar*. Oxford: Clarendon Press.
Moravcsik, Edith A. 1993. Why is syntax complicated. In Mushira Eid & Gregory Iverson (eds.), *Principles and predictions: the analysis of natural language*, 73–92. Amsterdam/Philadelphia: John Benjamins.
Moravcsik, Edith A. 2006. *An introduction to syntactic theory*. London, New York: Continuum.
Moravcsik, Edith A. 2009. Partonomic structures in syntax. In Vyvyan Evans & Stéphanie Pourcel (eds.), *New directions in cognitive linguistics*, 269–285. Amsterdam/Philadelphia: John Benjamins.
Moravcsik, Edith A. 2010. Conflict resolution in syntactic theory. *Studies in Language* 34(3). 636–669.
Moravcsik, Edith A. & Jessica R. Wirth (eds.). 1980. *Current approaches to syntax* (Syntax and Semantics 13). New York, London: Academic Press.
Müller, Stefan. 2010. *Grammatiktheorie*. Tübingen: Stauffenburg Verlag.
Müller, Stefan. 2016. *Grammatical theory. From transformational grammar to constraint-based approaches. Volumes 1 and 2*. Berlin: Language Science Press.
O'Grady, William. 2005. *Syntactic carpentry: an emergentist approach to syntax*. Mahwah, NJ: Erlbaum.
Rákosi, Csilla. 2005. *Nyelvészet és argumentációelmélet* (Linguistics and argumentation theory). Debrecen: DE Kossuth Egyetemi Kiadó.
Rákosi, Csilla. 2014. Data and the resolution of inconsistency in Optimality Theory. In András Kertész & Csilla Rákosi (eds.), *The evidential basis of linguistic argumentation*, 271–307. Amsterdam/Philadelphia: John Benjamins.

Rouse, Joseph. 1998. New philosophies of science in North America – twenty years later. *Journal for General Philosophy of Science* 29. 71–122.

Sadock, Jerrold M. 1991. Autolexical Syntax. In *A theory of parallel grammatical representations* (Studies in Contemporary Linguistics). Chicago: The University of Chicago Press.

Sampson, Geoffrey & Anna Babarczy. 2014. *Grammar without grammaticality. Growth and Limits of Grammatical Precision*. Berlin: Mouton de Gruyter.

Sanders, Gerald A. 1972. *Equational grammar*. Mouton: The Hague.

Sanders, Gerald A. 1975. *Invariant ordering*. Mouton: The Hague.

Shlonsky, Ur. 2010. The cartographic enterprise in syntax. *Language and Linguistics Compass* 4(6). 417–429.

Sells, Peter. 1985. *Lectures on contemporary syntactic theories: an introduction to Government-binding Theory, Generalized Phrase Structure Grammar, and Lexical-functional Grammar*. Stanford, CA: Center for the Study of Language and Information.

Simone, Raffaele. 2004. The object, the method, and the ghosts. Remarks on a terra incognita. *The Linguistic Review* 21. 235–256.

Stewart, Thomas W. 2016. *Contemporary morphological theories. A user's guide*. Edinburgh University Press.

van Riemsdijk, Hank. 1984. Introductory remarks. In Wim de Geest & Yuan Putseys (eds.), *Sentential complementation*, 1–9. Dordrecht: Foris.

Van Valin, Robert D. Jr. 2010. *Role and Reference Grammar as a frame for linguistic analysis*. In Heine & Narrog (eds.), *The Oxford Handbook of Linguistic Analysis*, 703–738.

Part I: **Approaches to syntax**

Cristiano Broccias

2 Cognitive Grammar

Abstract: This chapter discusses how syntax is handled in Langacker's Cognitive Grammar (CG). It first of all points out that CG rejects the notion of autonomy of language in general and syntax in particular in that various general cognitive abilities and cognitive models are appealed to in order to make sense of linguistic organization and no clear-cut boundaries are assumed to exist between syntax and lexicon. More generally, CG argues that much in language is a matter of degree. The chapter then discusses syntactic functions, which are not regarded as primitives, and constituency. In particular, it highlights that "classical constituency", traditionally represented by means of syntactic trees in formal approaches, is just one possible level of grammatical organization and description: attention to dynamicity or processing and discourse reveals that grammatical structure may be flat and serial, rather than hierarchical.[1]

1 Data

Cognitive Grammar (CG) is a theory of language whose claims can in principle be validated through the use of both elicited and unelicited data, that is data obtained through both experimental methods and corpora (see e. g. Langacker, 2016c). It must be pointed out, however, that the creator and main developer of CG, Ronald Langacker, uses mainly data based on introspection as far as English is concerned. Still, Langacker himself resorts to non-introspective data when dealing for example with applications of CG to non-Indo-European languages. It is also worth remarking that, as CG does not assume linguistic primitives such as syntactic functions or parts of speech (see below for more details), language variation is not modelled on the basis of the setting of some parameters as in traditional generative grammar. In fact, as in Croft's Radical Construction Grammar (RCG), see Croft (2001), linguistic categories may be viewed as language-specific in CG. Although typological studies cast in CG are lacking, this observation concerning the non-universal nature of linguistic categories would not prevent language comparison. As Croft shows within the largely compatible

1 CG makes extensive use of pictorial representations as a (hopefully) useful heuristics for the elucidation of linguistic phenomena. Thus, this chapter also includes numerous figures that may, admittedly, sometimes strike the reader as being quite complex. However, it is important to stress that a fine-grained understanding of such diagrams lies beyond the scope of this chapter. Rather, they are included in the hope that they may convey the gist of the phenomena under discussion and may prove to be a useful guide for the reader who is keen on exploring CG in more detail.

https://doi.org/10.1515/9783110540253-002

framework of RCG (see Broccias, 2002), language comparison is still possible once a universal conceptual space, modelled by basic cognitive abilities (see below for more details), is appealed to.

2 Goals

As its very name suggests, CG aims at offering a cognitively plausible theory of grammar. Langacker (2016a) distinguishes between two stages in CG research. "**Classical**" CG stretches from its inception in the 1980s to roughly the turn of the 21st century (see Langacker, 1987, 1991, 1999, 2002) and focuses on offering a description of language structure that is alternative to the dominant (generative) view of grammar as an **autonomous module** of an autonomous **language faculty** (see e. g. Taylor, 2007). CG is agnostic about the existence of a language faculty and does not view language as being independent of **general cognition**. Lexicon, morphology and syntax are not treated as autonomous **components** of an independent language faculty but are held to be identical in nature in that they are all symbolic, consisting in pairings of form and meaning (see below). Importantly, this implies that grammar is inherently "meaningful".

As any cognitive description of language structure must obviously be compatible with **language processing**, the second stage in the development of CG (see, in particular, Langacker, 2008, 2009, 2014, 2016a, 2016b), from the turn of the millennium onwards, focuses on integrating structure with **processing** as well as **discourse**.

2.1 Cognitive abilities

Before exploring the two stages in the development of CG in more detail, it must be emphasized that CG relies on a variety of independently existing cognitive processes in order to **describe** language. Here I will just concentrate on four, namely association, automatization, construal and categorization. As is repeatedly pointed out in CG, much in language (and cognition) is a matter of degree. Thus, these abilities are not necessarily distinct from one another. For example, as will be seen below, categorization requires association.[2] Further, these abilities may be evoked simultaneously, as the discussion of "constituency" in Section 3.3 will also show.

2 Nor are the four abilities mentioned above by any means exhaustive of our cognitive potential. Other abilities include, for example, our capacity for conceptual grouping, manipulating mental spaces, engaging in conceptual blending, fictive thought, and mental scanning through a domain (see Langacker, 2008, 34–36).

2.1.1 Association

Our cognitive ability to establish **associations** is manifest, for instance, in the linking between a semantic structure and a phonological structure in lexical items.[3] The former is also called the semantic pole or meaning of an expression while the latter is called the phonological pole or form of the expression. The lexical item *tree*, for example, can be characterized as a pairing of meaning, abbreviated as [TREE], and form, abbreviated as [tree], which gives rise to a higher-level entity called a **symbolic assembly**, [[TREE]/[tree]]. A symbolic assembly is thus made up of three “entities”: a semantic pole, a phonological pole *and* the symbolic link connecting the two poles. Also, the semantic pole is understood to include information that is traditionally regarded as “encyclopaedic” (for example, the knowledge that trees are used to produce paper). Similarly, the label “phonological pole” is construed broadly so as to also encompass other bodily manifestations such as gesture.

2.1.2 Automatization

CG stresses that language comprises many expressions that are units. These are structures that are used automatically, without much constructive effort. They do not only include lexical items such as *tree* but also multiword expressions, whether transparent or not, like *I’ve had enough*, *I love you*, *How are you holding up?*, *kick the bucket*, etc. The availability of units in language is to be related to the cognitive process of automatization or **entrenchment**, which accounts for our ability to perform more or less effortlessly tasks such as tying a shoelace or driving. In language, entrenchment leads to structures achieving unit status.

2.1.3 Construal

Also central to language description is our ability to **construe** the same situation in alternate ways. We are all familiar with the (entrenched) expression “Is the glass half empty or half full?”, which construes the same objective situation in two different ways. Construal involves various dimensions such as **schematicity** (or granularity), **prominence** and **perspective**.

Schematicity refers to our ability to construe a situation in terms of varying degrees of specificity. *Bowl* is a less schematic description than *container*, which in turn

3 Other examples of association include the establishment of a link between a source and a target in metaphor (e. g. *head of Department*) and metonymy (e. g. *Washington signed the agreement*) or the links in polysemy networks (e. g. *mouse* as an animal and as a computer device).

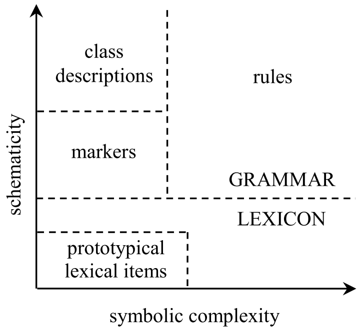


Figure 1: The grammar-lexicon continuum.

is less schematic than *thing*. Schematicity is of central importance in grammar because it constitutes one of the dimensions of variation that allow for the distinction between "lexical" items and "grammatical" structures. CG rejects the traditional dichotomy between lexicon and grammar because it views them as two opposite poles along a continuum, see Figure 1. "Prototypical" lexical items such as *tree* can be described as expressions that are low in schematicity and symbolic complexity; they tend to be fairly specific in terms of their meaning and form and are of limited size. What are traditionally described as "morphemes" such as agentive *–er* (*driver, hunter, murderer*, etc.) are structures that are higher in schematicity (the verbal event is not specified) but are still of limited complexity. Instead, grammatical patterns or, traditionally speaking, "rules" are viewed as schematic **abstractions** over specific usage events (actual instances of language use) and are more complex symbolically than lexemes and morphemes. The pattern NP V NP, for example, is a schematic representation of the commonality of specific expressions such as *Sally likes chocolate*, *The players in the red shirts won the match*, *She watched the telly*, etc.[4] Intermediate between "prototypical" lexical items and grammatical "rules" are partly-filled multi-word expressions such as V_sX in the N_b, where V_s is a verb of impact like *strike*, *kick*, *hit* and N_b is a body-part noun like *shin*, *back*, *face*. Such expressions are difficult to accommodate in theories that distinguish neatly between lexicon and grammar but are accounted for straightforwardly in CG.

Linguistic expressions also construe conceptual content in terms of **prominence** or focus of attention. The verb *write* and the nouns *writer* and *book* are said in CG to evoke the same content in that they presuppose the same conceptual base: someone

4 In other words, CG dispenses with "rules" in the traditional (e. g. generative) sense and replaces them with schemas. For example, the plural morpheme *–s*, as in *dogs*, may be regarded in CG as a partially schematic noun, whose semantic pole denotes a multiplex mass and whose phonological pole specifies that the final segment is e. g. [z] (see e. g. Langacker, 1987, 82–85 and, for a more recent account, Langacker, 2016b). As schemas can exist at different levels of specificity, it is not possible to enumerate a finite set of "rules" in CG.

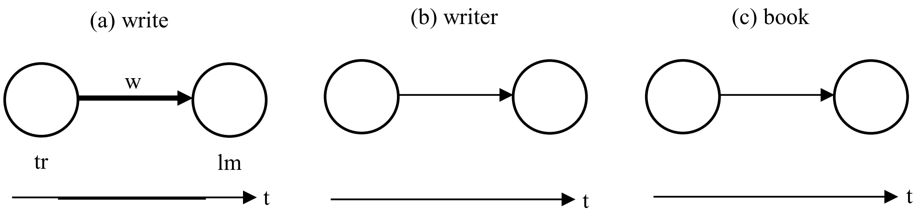


Figure 2: Profiling as prominence.

is involved in the production of a text. However, while the verb *write* **profiles** the process of composing a text, see Figure 2a, *writer* and *book* profile the two participants involved in it, see Figure 2b–c. Profiling, which is represented by heavy lines/contours in pictorial representations, thus determines the grammatical category of an expression. A noun profiles a "thing", represented as a circle. "Thing" is a technical term in CG, where it is used to refer to "a set of interconnected entities which function as a single entity at a higher level of organization" (Langacker, 2008, 107). A verb profiles a process, represented as an arrow in Figure 2. A process is a relationship that develops or is tracked through time, as is shown by the heavy time (t) line in Figure 2a. Adjectives and prepositions profile atemporal or non-processual relationships. Within a relationship, whether temporal or atemporal, one participant turns out to be more prominent than the other(s). This is called the **trajector** (tr) or primary focal participant while the secondary focal participant, if present, is called a **landmark** (lm). In Figure 2, the noun *writer* stands for the trajector in the relationship profiled by *write*, while *book* codes the landmark.

One more example will suffice. Consider the prepositions *in front of* and *behind*. They evoke the same content but differ in the entity that is given trajector status. In *Francis is sitting in front of Ben*, Francis is the trajector and Ben is the landmark, while the opposite is the case with *Ben is sitting behind Francis*, see Figure 3.

The choice of trajector and landmark in *in front of* and *behind* is obviously related to **perspective** or perspectivization. Perspectivization is pervasive in language. Consider, for instance, the proximal determiner *this* in *Sally likes this cat*. *This* can only by understood with reference to the ground (G), which is made up by the interlocutors/conceptualizers (the speaker and the hearer) and their immediate circumstances, see Figure 4. Still, *this* profiles a "thing", which is further specified by *cat* in our ex-

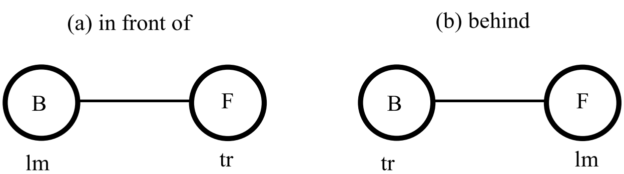


Figure 3: *In front of* vs. *behind*.

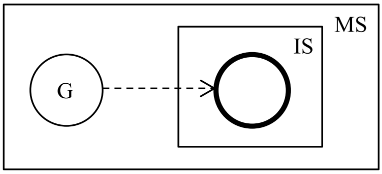


Figure 4: The determiner *this* as a grounding element.

ample, rather than the grounding relationship (the relationship between the ground and the "thing") as is illustrated in Figure 4. Thus, the ground is left "offstage" within a window of maximal scope of awareness (MS), while the "thing" is the "onstage" focus of attention, being the entity profiled within the window of immediate scope of awareness (IS).

The relation between what is onstage as the object of conceptualization and an offstage conceptualizer who acts as the subject of conception is also relevant diachronically. As is well-known, the English verb *go* has grammaticized into a marker of future time, as in *Sally is going to do the weekly shop*, which of course could also be used to describe Sally's actual motion. In CG, the grammaticization of *go* is viewed as an instance of **subjectification**. In the original, motion meaning of *go*, the conceptualizer, who is offstage and acts as the subject of conception, scans through time by tracking Sally's motion through space, which is onstage as the focus of attention. When *go* is used as a future marker, the subjective (i. e. the offstage conceptualiser's) scanning through time, which was immanent in the motion meaning of *go*, is no longer linked to spatial motion but is used to locate events in time. To put it differently, the schematic meaning inherent in spatial motion *go*, namely the scanning through time on the part of the conceptualizer, has been abstracted away from the spatial domain and put to use in a different domain, the temporal domain, to locate events.

2.1.4 Categorization

Schematization is linked to another fundamental conceptual ability, namely **categorization**. The motion meaning of *go* and the temporal meaning of *go* referred to in the previous section can be regarded as **instantiations** or elaborations in the spatial domain and the temporal domain, respectively, of the schematic meaning of *go*. The relation of categorization between two structures, however, is not necessarily one of instantiation but may involve **extension**. Under normal circumstances, the verb *sneeze*, for example, is unlikely to be categorized by English speakers as a force-dynamic verb in the same way as *kick* is, as in *Sally kicked the football over the fence*. Nevertheless, Goldberg's (1995) well-known example *Pat sneezed the napkin off the table*, where *sneeze* is used in the so-called caused-motion construction, can be made sense of only if *sneeze* is categorized as a force-dynamic verb by a process of extension because of some perceived similarity with force-dynamic verbs. The caused-motion construction prototypically makes use of force-dynamic verbs like *kick*. Sneezing may involve a sudden burst of air from one's mouth and this can be construed as a force capable of displacing an object from its current location so that *sneeze* becomes a plausible candidate for use in the caused-motion construction. Obviously, the force-dynamic meaning of *sneeze* has not yet gained unit status, at least among non-linguists, because this meaning is not accessed automatically. Some constructive effort is required to make sense of its occurrence in the caused-motion construction.

2.2 Cognitive models

Alongside cognitive abilities such as association, automatization, construal and categorization, CG postulates various cognitive models or conceptual archetypes which are held to account for linguistic organization. One example is the stage model or **baseline viewing arrangement**, which was represented in a compact way in Figure 4 and is shown in more elaborate fashion in Figure 5. As was observed above, we need to distinguish between an "offstage" region, which involves the ground (the speaker (S), the hearer (H) and their immediate circumstances, comprising their interaction), and an "onstage" region, which includes what is "viewed" as the focus of attention, represented as a generic entity by means of the emboldened square in Figure 5. As should be apparent from the discussion above, the stage model is relevant to grounding and grammaticization among other things.

Another important model is the **billiard-force model** (Langacker, 1991, 13), which basically describes an agent-patient interaction resulting in the patient's change of place or, more abstractly, state. Its relevance to grammar is evident in so-called (transitive) resultative constructions (Goldberg, 1995) such as *Pat ran her sneakers thread-*

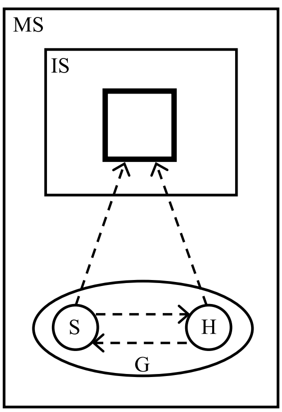


Figure 5: Baseline viewing arrangement.

bare. Broccias (2003) claims that the semantic pole of the resultative construction involves the activation of the billiard-ball model: the verbal event or process, *run* in the example at hand, must be construable (metaphorically, if necessary) as denoting a forceful interaction between an energy source, coded as the constructional subject, and an energy sink, coded as the constructional direct object. This metaphorical construal relies on the interpretation of running as being so frequent and/or "excessive" that it can have an impact on the "state" of the sneakers.

Also worth mentioning here is the conception of reality as a "growing" cylinder, whose face represents current reality, see Figure 6.[5] A conceptualizer's knowledge or conception of what is real represents, however, a portion of reality, which is called "conceived reality", whose growing face is "immediate reality". Lack of a modal (as is *She washed her hair*, *She is washing her hair*) indicates that the profiled process belongs in conceived reality. The use of "present tense" places the process in immediate reality while the use of "past tense" specifies non-immediacy by placing the profiled process in conceived reality outside immediate reality. A modal (e. g. *She may/might wash her hair*) signals that the profiled process is placed in "irreality" (Langacker, 2008, 302), the complement of conceived reality.

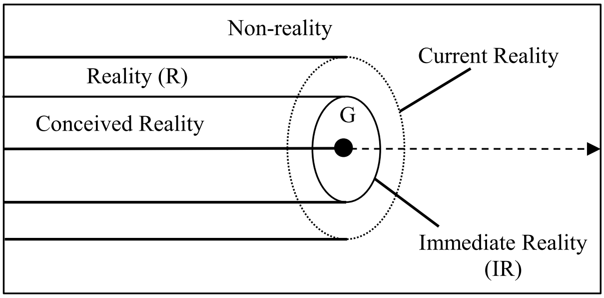


Figure 6: Our conception of reality.

3 Tools

CG rejects the notions of modularity, autonomy of syntax and traditional hierarchical constituency as represented in syntactic trees. It does not posit transformations or null elements. Rather, CG adheres to the content requirement, the idea that only semantic structures, phonological structures and symbolic structures linking the two

5 This model relies on another major model called the control cycle (see e. g. Langacker, 2009), which describes an agent, initially at rest within a dominion under its control, capturing a target so as to include it in the agent's dominion.

are necessary to describe language structure. As was pointed out in the previous section, grammar is meaningful in CG. This implies that CG strives to offer a conceptual characterization of word classes, syntactic functions and syntactic constructions. Importantly, syntactic functions are not seen as basic units of descriptions but just as convenient descriptive labels for a variety of conceptual operations.

3.1 Syntactic functions

Let us consider syntactic functions, starting with the functions "subject" and "object". Subject and object are defined both prototypically and schematically in CG. The schematic characterization is assumed to be valid for and to be immanent in all instances of the category in that it expresses the commonality inherent in the category's various realizations. From a semantic point of view, a subject is identified prototypically with an agent (an energy source) while, schematically, it is defined as "a nominal that codes the trajector of a profiled relationship" (Langacker, 2008, 364). Thus, the subject codes a primary focal relational element. An object, from a semantic point of view, is identified prototypically, at least in English, with a patient (an energy sink) while, schematically, it is defined as a nominal that codes the landmark of a profiled relationship (Langacker, 2008, 364). The referent of an object is therefore a secondary focal relational element.

If (the referent of) an object is construable as a patient, Langacker uses the more restrictive term "direct object" to describe it. "Direct object" and, hence, transitivity are thus employed to identify those nominals that allow passivization, since passivization is taken to be symptomatic of patient-like construal.

An important caveat is in order. Subject and object are best defined schematically as the "primary focal relational element" and the "secondary focal relational element" rather than the "primary focal participant" and the "secondary focal participant", respectively. This is a looser characterization, which I used in Section 2.1.3. A nominal trajector (a subject) can be a setting or a location rather than a participant, as in *The garden is swarming with bees* and *This book contains a lot of information on syntax*, where *the garden* is a setting and *this book* is a (metaphorical) location (see, e. g., Langacker, 2008, 361, 374 n. 19, and 387). Examples of objects that are not participants include paths (*We hiked a new trail*), locations (*The train approached the station*) and measurements (*It weighs ten kilos*).

Another type of object is the "indirect object", which, from a semantic point of view, corresponds prototypically to an experiencer and which is marked by a preposition (typically meaning 'to' or 'at') or dative case. In cases like Italian *A Paola piace il cioccolato* (lit. 'at/to Paola likes the chocolate'; i. e., 'Paola likes chocolate'), the indirect object, Paola, is a landmark, here a secondary focal participant.

When an indirect object occurs with a dative verb like Italian *dare* ('give'), as in *Luca ha dato il libro a Paola* (lit. 'Luca has given the book to Paola'), Langacker does not commit himself to the analysis of the indirect object as a landmark. He says that "[p]erhaps the indirect object should be considered a secondary landmark. If not, it is at least quite salient as a profiled participant" (Langacker, 2008, 393). As for English, the verb *give* can occur both in the double object construction (*Luke gave Paula the book*) and in the prepositional dative construction or caused motion construction (*Luke gave the book to Paula*). In the double object construction, Langacker analyzes the recipient (*Paula*) as the landmark. The landmark, although it is a recipient, is treated as a direct object because it can become the subject in the corresponding passive sentence (*Paula was given the book*). The analysis of the entity being transferred (*the book*) is however uncertain. Langacker (2008, 360) claims that while the agent and the recipient are focal participants, the transferred entity is a participant which is not focused as trajector or landmark and calls it a central participant. (This would imply that *the book* is not an object vis-à-vis the definition above.) In the caused motion construction, the analysis is similar to the one for Italian in that the landmark is identified with the transferred entity (see Langacker, 2008, 242 and 393–94).

3.2 Constituency and assemblies of symbolic structures

CG replaces hierarchical constituency with assemblies of symbolic structures. In fact, CG contends that the composition of symbolic structures does not necessarily result in strictly hierarchical assemblies. Consider the sentence or assembly *Sally likes this black cat*, which will also be used to introduce the notions of head, complement and modifier. A possible compositional path for this sentence is illustrated in Figure 7.

At the lowest level of this compositional path, the adjective *black* combines with the noun *cat*. *Black* profiles a relationship between a thing, represented by the emboldened circle, and a property, given as **b** in the diagram. The ellipses that appear in the diagram represent bundles of properties that serve to specify the various concepts. The thing in the representation for *black* serves as a trajector and is put in correspondence, as is shown by means of the dashed line, with *cat* and is elaborated by it. In CG parlance, the trajector of *black* is an elaboration site (or e-site) with respect to *cat*. As *cat* is the profile determinant or **head** in *black cat* (this expression refers to a cat, not to blackness), *black* functions as a **modifier** because a salient substructure of *black* is elaborated by the head.

At the next level in the compositional path, *this*, a grounding element, combines with the nominal *black cat*. Note that while the horizontal dimension in the diagram shows the composition of symbolic assemblies thanks to their conceptual overlap or correspondences, the vertical dimension in the diagram shows the categorizing relationships that exist between the various symbolic assemblies. For example, *black cat* is an instance of the more general nominal *cat*.

At a further level in the compositional hierarchy, the verb *likes*, which profiles a relationship between a trajector and a landmark, combines with *this black cat*. *This*

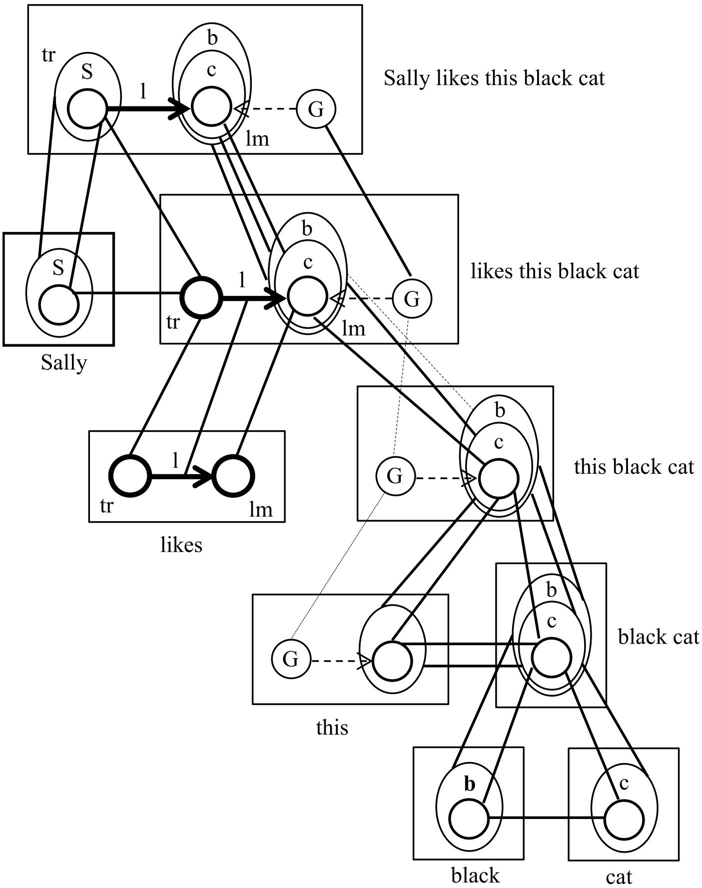


Figure 7: A compositional path for *Sally likes this black cat*.

black cat corresponds to or elaborates the landmark of *like*. As the assembly *likes this black cat* profiles a process (liking) rather than a thing (the cat), *likes* is the head. *This black cat* elaborates a salient substructure of the head and is, thus, described as a **complement** in CG.

Finally, *Sally* elaborates the trajector of *likes this black cat* and thus functions as the subject nominal of the overall expression.

Although the compositional path illustrated in Figure 7 resembles a traditional constituency tree, CG claims that alternate constituencies are possible. For example, Langacker (2016a, 29) points out that a topic construction such as *This black cat Sally likes* is not hierarchical but essentially serial, as is illustrated in Figure 8, where *this black cat* "serves as a *reference point* (R) for interpreting a *target* (T), the process [*Sally*] *likes*" (Langacker, 2016a, 29). (In Figure 8, only some of the correspondences are shown for the sake of simplicity.)

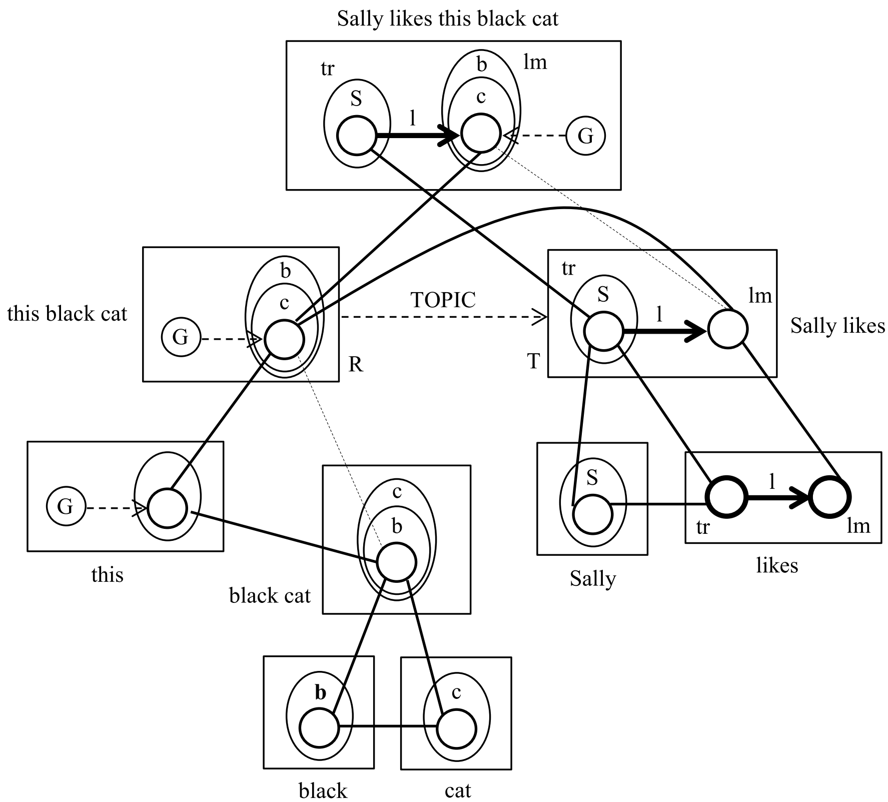


Figure 8: The topic construction *This black cat Sally likes*.

3.3 Four axes

The observation concerning sequential access in the topic construction leads us to the second phase of CG, where Langacker emphasizes the dynamic nature of grammar. Notations like [[TABLE]/[table]] may suggest that the "semantic and phonological poles are clearly delimited and exist independently" (Langacker, 2016a, 29). In reality, linguistic elements are embedded in a usage event (an actual instance of language use) that involves at least the four axes, shown in Figure 9, namely the individual axis, the interactive axis, the descriptive axis and the discursive axis. (In Figure 9, boxes are used only for the sake of clarity without implying the existence of clear boundaries at all.) Expressions such as *Ouch!* are primarily individual in that they express the speaker's feelings. Expressions like *hello* are primarily interactive in that they enact a social routine. Lexical and grammatical structures pertain primarily to the descriptive axis because they describe the objective scene. Finally, expressions like the topic construction or the connectors *moreover* and *so* are primarily discursive in that they serve to relate a usage event to others in the past or the future.

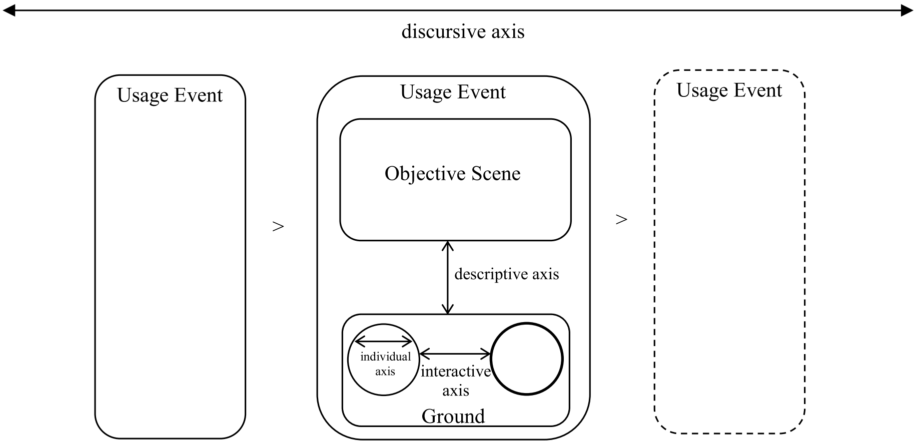


Figure 9: The axes involved in language use.

In Langacker's view, reference to the discursive axis or discursive substrate allows for a unified account of grammar and discourse. Consider the two sentences in (1) from Langacker (2016a, 36).[6]

(1) a. Alice kicked the dog, so it barked furiously.
    b. Alice kicked the dog. It barked furiously.

(1a) is a complex sentence while (1b) is a sequence of two independent clauses. Langacker claims that the pronominal anaphor *it* can be treated in the same way in either case, instead of viewing it as part of grammar in (1a) and as part of discourse in (1b). Both (1a) and (1b) involve two clause-sized "windows", which here correspond to intonation units, see Figure 10a and 10b.[7] The pronoun *it* profiles a neuter (n) entity within the current window of attention (the rightmost rectangle in Figure 10). This window is embedded within a larger scope of awareness (the dashed rectangle in Figure 10) that includes a previous window of attention (the leftmost rectangle in Figure 10). This previous window includes a nominal (*the dog*, in the example at hand) that corresponds to the profiled entity in the current window of attention. Both profiled entities therefore represent the same element in the descriptive target (DT), "that portion of our mental universe which is under discussion in a given discourse"

6 Here, the two sentences should be understood as depicting the same complex event, namely one where the barking is a consequence of Alice's kicking the dog. (2b) is also compatible with the interpretation that the barking is the cause for Alice's kicking the dog but this reading is not relevant to the present discussion.

7 Following Chafe (1994), Langacker observes that a window of attention or processing window is often clause-sized. These clause-sized windows are called intonation units by Chafe and describe what "can be fully active in the mind at one time" (Chafe, 1994, 69).

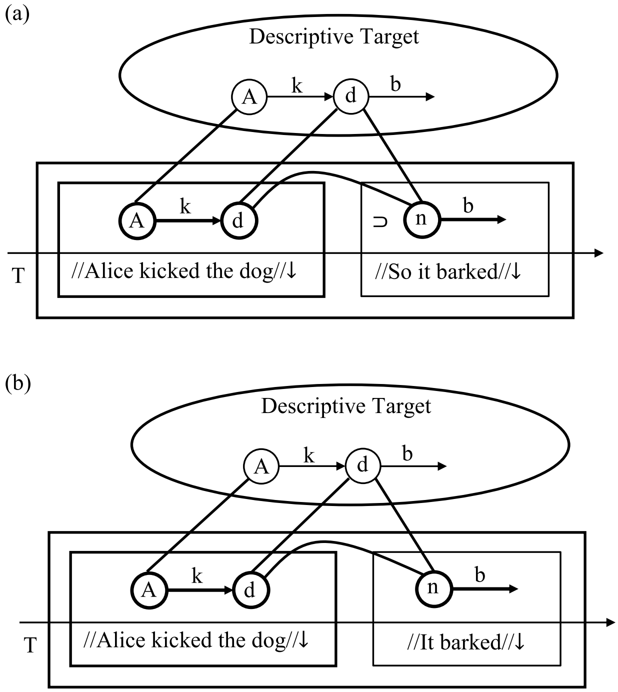


Figure 10: Processing representations of how two clauses are connected through anaphora.

(Langacker, 2016a, 35). Figure 10a and Figure 10b show that both sentences have a serial rather than hierarchical organization: each clause appears in a window of its own and the only substantive difference is the presence vs. lack of a connective element, *so*, represented as "⊃" in Figure 10a.

The discussion so far has implicitly shown that processing occurs simultaneously on different time scales (see also MacWhinney, 2014 for a similar position). Anaphora requires two windows of attention and a larger scope of awareness (see Figure 10), but the difference between the windows of attention and the larger scope of awareness is a matter of degree and intermediate processing windows may be active. This is illustrated, in a non-anaphoric case, in Figure 11a, where the nesting boxes represent progressively larger processing/prosodic windows, which, at each level, are assumed to have roughly the same duration, as is shown by their relative sizes. The largest corresponds to the intonation unit //*Sally* / *likes chocolate*//, while the subject (*Sally*) and the predicate (*likes chocolate*) occur in shorter prosodic windows of roughly the same duration. In this instance, the prosodic groupings coincide with "traditional" hierarchical constituency (subject + predicate). In other words, **discursive organization**, as evidenced by prosodic groupings, and **descriptive organization**, which pertains to grammar in the traditional sense, dovetail with each other. This situation may be the norm or baseline but need not be so all the time. Consider, for example, the vari-

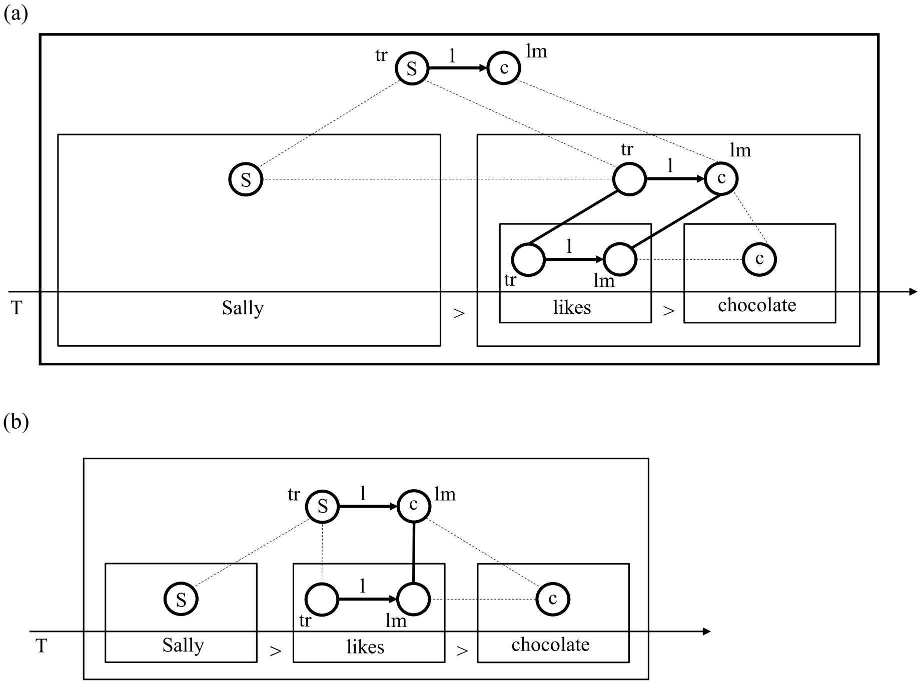


Figure 11: Processing and “constituency”.

ant //*Sally* / *likes* / *chocolate*//, where subject, verb and object occur in processing windows of roughly the same duration. In this case, see Figure 11b, there is no intermediate “constituent” *likes chocolate* so that the grammatical organization is “flat”.[8] The distinction between Figure 11a and Figure 11b may be a matter of degree in that the composite conception *likes chocolate* can emerge at some level of processing, as is shown in Figure 12 by means of the dashed box. Crucially, this composite conception may not be symbolized by any prosodic grouping so it does not form a grammatical constituent in the traditional sense.

Prosodic grouping does not only suggest that grammatical structure can be flat or **serial** rather than hierarchical. In some cases, discursive organization may override descriptive organization. While in //*The letter* / *Sally was expecting* // *just arrived* // the prosodic groupings correspond to traditional hierarchical constituents, in //*The letter* / *just arrived* // *that Sally was expecting* //, the descriptive grouping or compos-

8 Incidentally, this means that traditional criteria for determining constituency such as replaceability by pro-forms and deletability are just diagnostics for descriptive organization that does not necessarily coincide with discursive organization. In other words, they point to descriptions on the part of the linguist that do not necessarily correspond to processing because “constituency” is just one of the dimensions of conceptual organization and is, further, not necessarily fixed.

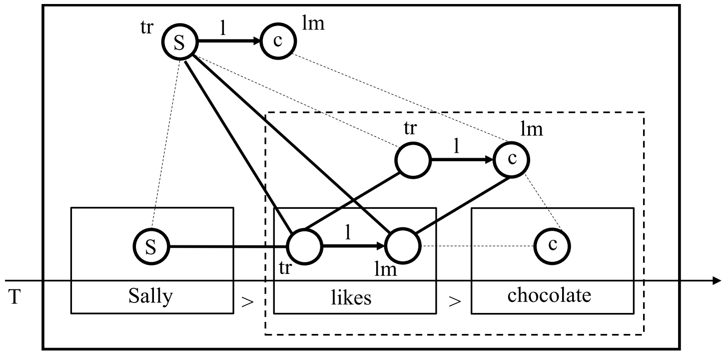


Figure 12: The emergence of composite conceptions that are not grammatical constituents.

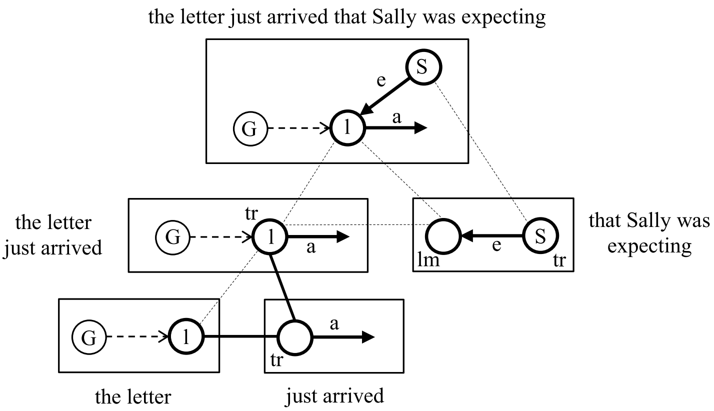


Figure 13: A compositional path for *The letter just arrived that Sally was expecting*.

ite conception *the letter that Sally was expecting* is not symbolized by a single prosodic grouping (an intonation unit) so that it does not form a constituent in the traditional sense of the term. Figure 13 follows Langacker (2014, Figure 14) and shows the compositional path for *//The letter / just arrived // that Sally was expecting //*. Note that both events are profiled at the composite structure level as is evidenced by the emboldened arrows for *arrived* and *expecting*. However, a more perspicuous representation for this example is along the lines of Langacker (2014, Figure 20), shown here as Figure 14, because the "two events are not profiled simultaneously [as Figure 13 seems to suggest, CB], but rather sequentially, in successive windows of attention at the basic level [the level corresponding to intonation units, CB]" (Langacker, 2014, 54–55). Sequential activation is shown in Figure 14 by emboldening first the arrow for *arrived* and then the one for *expecting*. (The index "0" stands for processing at the baseline level while "+1" stands for processing at a larger timescale.)

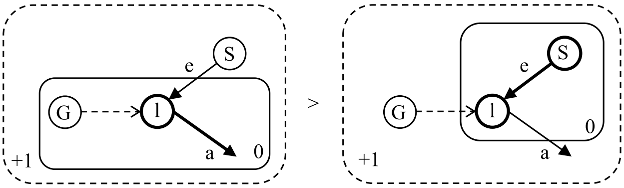


Figure 14: Sequential access in *The letter just arrived that Sally was expecting*.

The previous example shows that discursive organization is crucial to the analysis of subordination (see also Langacker, 2014). Let us explore this topic further. The complex sentence in (2) is usually assigned the constituency in (2a) while prosody suggests the groupings in (2b).

(2) a. [Amy says [Bob thinks [Chris believes [Doris left]]]].
b. // Amy says // Bob thinks // Chris believes // Doris left //.

Langacker points out that the bracketing in (2a) does not necessarily reflect grammatical constituency but, rather, conceptual layering, which has to do with the descriptive target. As is suggested by (2b), the clausal organization may just be serial rather than hierarchical, especially when multiple clauses are involved. As is shown in Figure 15a, the clauses are integrated by means of correspondences, thanks to the conceptual overlap between the landmark of a process and the trajector of the next. In this sense, each clause is thus "subordinate" to the previous one because it is accessed through it. The containment relation represented by the constituency in (2a) is still present but is now conceptual rather than grammatical, see Figure 15b. In other words, a distinction is drawn between grammatical subordination and conceptual subordination.

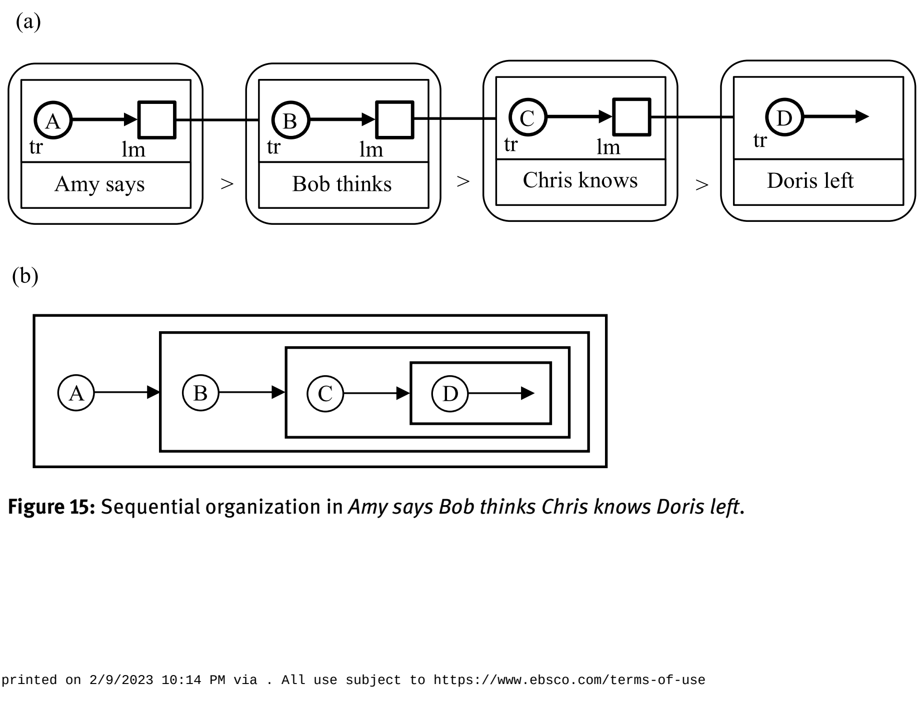


Figure 15: Sequential organization in *Amy says Bob thinks Chris knows Doris left*.

3.4 Clauses

After having observed that CG does not subscribe to classical grammatical constituency, it is time to take a closer look at what a clause is in CG (see Langacker, 2015 for more details).

In CG, a clause is said to profile a process, which is a relationship tracked through time, and to express a proposition, that is, a process assessable for validity. Although both a lexical verb and a clause profile a process, a lexical verb merely describes a **type** of occurrence (a basic process type). A clause, instead, is built by means of various dimensions of elaboration. The minimal or baseline elaboration, see Figure 16, involves 1) the specification of clausal participants, which results in an elaborated process type or profiled occurrence (**p**) and 2) grounding by tense, which results in a grounded process **instance** or proposition (P). An elaborated process type grounded by tense constitutes a **baseline** clause. The baseline clause presupposes as its conceptual substrate the **baseline viewing arrangement** shown in Figure 5 (see above), which describes the **baseline speech act** of **statement**. Elaborations of the baseline clause are effected by means of **perspectivalization** and **grounding**. Perspectivalization has to do with the use of the passive, progressive and perfect while grounding involves situating a process with respect to reality and immediacy.

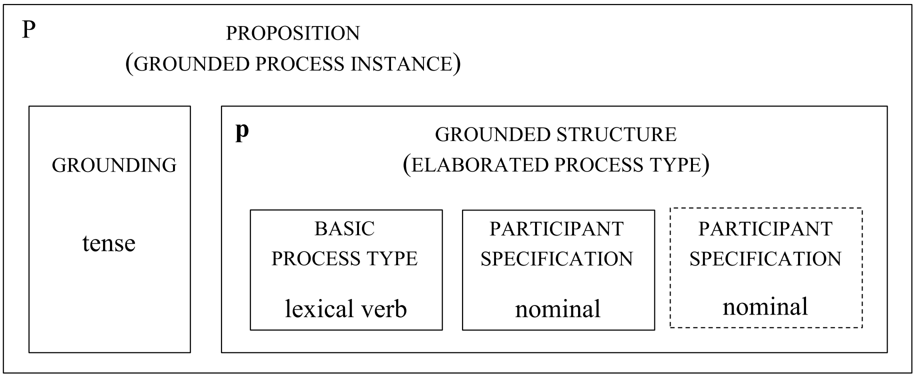


Figure 16: The baseline clause.

In a baseline clause, tense is either not marked, in the so-called present tense (with the exception of the third person singular), or marked, in the so-called past tense. The fact that English does not mark present tense is not problematic because grounding is implicit in the baseline viewing arrangement, where the profiled process **p** is placed, together with the ground, in reality (R), in fact in immediate reality (IR). This is linked to the metaphor of R as a cylinder growing with time, whose face is IR, see Figure 6 above. The use of a non-immediate or distancing form places **p** outside IR. Modals, instead, place **p** outside R.

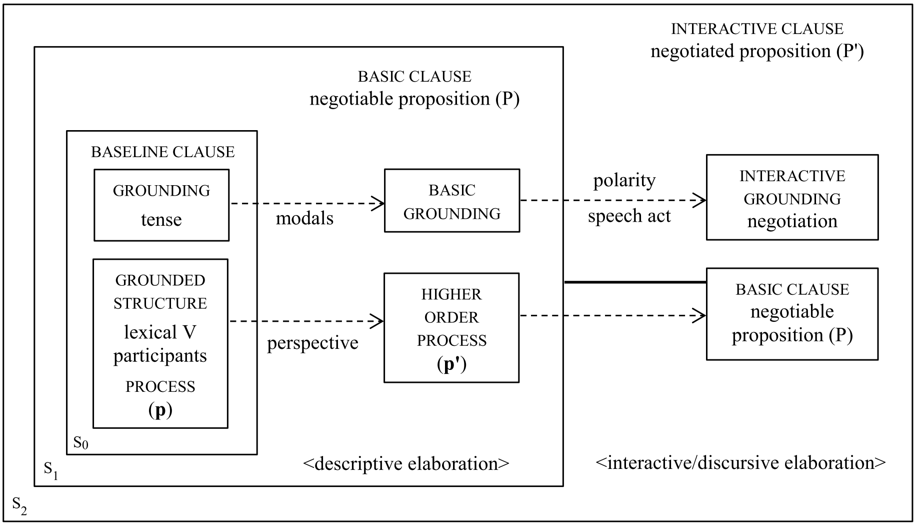


Figure 17: Basic and interactive clauses.

A basic clause represents a higher stratum compared to a baseline clause. An even higher stratum, see Figure 17, is an **interactive** clause, which expresses a proposition that is negotiated. In an interactive clause, interactive grounding augments basic grounding. Interactive grounding has to do with **polarity** and **speech act**. The baseline for polarity is POSITIVE (*She is smart*) while at a higher stratum we have NEGATIVE, marked by *not* (*She isn't smart*), and AFFIRMATIVE, marked by unreduced stress (*She IS smart*). The baseline for speech act is STATEMENT while at a higher stratum we have, for example, QUESTIONING (*Is she smart?*). The subject and the finite verb are thus very important when it comes to interactive grounding. Langacker suggests that they define a functional grouping called the existential core ($C_{\exists}$), within which the finite verb is called the existential verb ($V_{\exists}$). The existential core provides "a compact, clause-initial presentation" (Langacker, 2015, 24) of **existential negotiation**, which pertains to "establishing joint epistemic control or building up [. . .] a shared conception of reality" (Langacker, 2015, 38). In this sense, the so-called "auxiliary verbs" can be regarded as existential verbs: as they profile schematic processes, their fundamental contribution has to do with the existence of a relationship rather than the relationship itself.

Another important element in a clause is the anchor, which is defined as the initial element of a sequence. The anchor functions as a point of access. In a baseline clause, the subject, which corresponds to the trajector of the lexical verb, functions as anchor. Things get more complicated as we deal with higher strata. For example, in a passive, which is a basic rather than baseline clause, the anchor is still the subject but is no longer the trajector of the lexical verb because the subject corresponds to the landmark of the lexical verb. In interactive clauses, a variety of elements can function as discursive anchors (A'), such as non-subject nominals, prepositional phrases and adverbs, see (3), after Langacker (2015, example (2)).

(3) a. Trump she would never vote for.
b. In parts of Ireland it rains almost every day.
c. From Milan he will drive to Paris.
d. Therefore she decided to leave.
e. On the counter it goes!
f. Carefully she unwrapped the present.

Various options are detailed in Figure 18. (a) is an interactive clause but contains only a descriptive anchor (A) for both $C_\exists$ and the clause. (b) includes a discursive anchor

(a)

$C_\exists$		R'
A	$V_\exists$	
She	will	visit Genoa.

(b)

A'	$C_\exists$		R'
	A	$V_\exists$	
Genoa	she	will	visit.

(c)

$C_\exists$			R'
A'/A	$V_\exists$	R	
Never	will	she	visit Genoa.

(d)

$C_\exists$			R'
A'/A	$V_\exists$	R	
What	will	she not	visit?

(e)

$C_\exists$		R'
A'/A	$V_\exists$	
Nobody	will	visit Genoa.

(f)

$C_\exists$		R'
A'/A	$V_\exists$	
Who	will	visit Genoa?

(g)

$C_\exists$		R'
A'/A	$V_\exists$	
SHE	will	visit Genoa.

(h)

$C_\exists$		R'
$A'/A/V_\exists$	R	
Will	she	visit Genoa?

Figure 18: Anchors and the English clause.

(A'), the clause-internal topic *Genoa*, but, unlike in (c), there is no subject-auxiliary inversion. In (c), Langacker claims that *never* takes up a two-fold role as both a discursive anchor (A') and a descriptive anchor (A). It is a descriptive anchor because, trivially, it comes first in the clause. It is a discursive anchor because it instantiates (negative) polarity, which pertains to existential negotiation. As the core gives an indication of existential negotiation, it makes sense to assign it to the existential core. Other elements that lend themselves to the same characterization are question words and negative pronouns, as is shown in (d)–(h). Importantly, as the core has the structure $A > V_\exists > R$, it is not possible to have cases such as (4), where there are two distinct anchors within it.

(4) a. *Trump would she never vote for.
b. *In parts of Ireland does it rain almost every day.
c. *From Milan will he drive to Paris.
d. *Therefore did she decide to leave.
e. *On the counter does it go!
f. *Carefully did she unwrap the present.

In English, in such cases, the subject, a core element, is thus found after $V_\exists$ in the remainder (R). Nevertheless, the subject itself can be used as a discursive anchor, when it corresponds to a negative pronoun, an interrogative pronoun or a clause-internal topic, see (e)–(g). Finally, even an auxiliary or existential verb can be a discursive anchor, as in (h). Langacker thus shows that inversion is not a purely "formal" rule but follows from discursive factors.

4 Evaluation

As a theory of grammar, CG is closest to Construction Grammar (CxG) in that they share key assumptions such as the grammar-lexicon continuum and the importance of cognitive abilities such as profiling and categorization, which implies that there is no clear-cut separation between language and general cognition. There are however important differences (see also Evans and Green, 2006, Chapter 22 for some introductory discussion). For example, a construction in CxG is defined as a pairing of semantic structure and syntactic structure, while syntax in CG has no independent role, being a part of "semantics". A construction in CG is defined as the pairing of a semantic pole and a phonological pole (broadly construed), see Section 2.1.1 above. Also, CG offers a conceptual characterization of syntactic functions such as subject and object, which instead seem to be taken as primitives in CxG, although, admittedly, this is not so in Croft's (2001) Radical Construction Grammar (which, however, still defines constructions as pairings of semantics and syntax rather than semantics and phonology).

The point about the non-primitive nature of syntactic functions in CG also distinguishes it from Hudson's Word Grammar (see e. g. Hudson, 2007), which despite sharing many basic (cognitive) assumptions with CG, views grammatical relations are primitives.

Similarities also exist between CG and Halliday's Functional Grammar (see e. g. Halliday and Matthiessen, 2014). The latter theory, like CG, stresses the continuum nature of grammar and lexis and has always incorporated discourse in its formulation.

In terms of cognitive abilities and models, CG bears some similarity to Talmy's framework (see e. g. Talmy, 2000), although it must be observed that Talmy also distinguishes between lexical and grammatical subsystems, which is not in tune with the CG grammar-lexicon continuum hypothesis.

The main challenge that CG (still) faces is testing its claims with the help of data obtained from psycholinguistic studies. While the dynamic/discourse view espoused by Langacker in the second stage in the development of CG makes CG even more truly a cognitive theory of language, the view of processing illustrated in this chapter requires theoretical clarification and experimental verification. It is not clear, for instance, whether the processing described by Langacker is carried out by the speaker and/or the hearer. In a cognitive theory of language, it is vital to distinguish between the two. While assemblies of symbolic structures especially in the first stage of CG seem to pertain to processing on the part of the speaker (as is also the case with the assembly of syntactic trees in generative grammar), the emphasis on windows of attention in the second stage of CG appears to describe processing on the part of the hearer or, at least, to portray the speaker in similar fashion to the hearer in that both speaker and hearer may be described as attending to material which is "out there" or "onstage".

5 Sample analysis

The discussion of subordination highlighted that hierarchical constituency is not guaranteed, especially with large clauses, as they may be better analyzed in serial terms. In the sample sentence *After Mary introduced herself to the audience, she turned to a man she had met before*, we have what are traditionally analyzed as a temporal adverbial clause (introduced by *after*) and a temporal adverb (*before*). While it may make sense to view the adverb *before* in terms of classical constituency, see Langacker (2014, Figure 17), it is more likely that the adverbial clause is processed in its own basic level window and preserves its profile, see Langacker (2014, Figure 25). Thus, this complex sentence may be taken to profile two clauses (two events), each of which is probably accessed in its own basic level window. This means that the two clauses are accessed sequentially rather than simultaneously so that the overall sentence does not illustrate classical constituency.

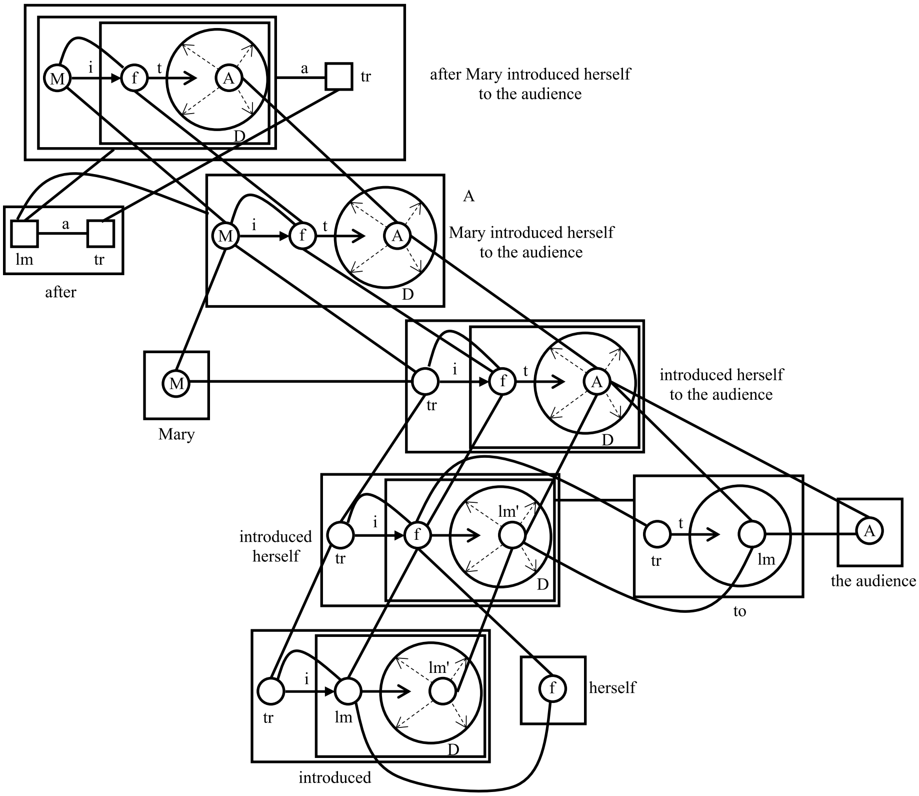


Figure 19: A compositional path for *after Mary introduced herself to the audience*.

In the following analysis, I have opted for a rather flat structure, although other options are plausible. For reasons of space and simplicity, I have omitted depicting either nominal or clausal grounding. Figure 19 shows a possible assembly for the adverbial clause a*fter Mary introduced herself to the audience*. The verb *introduce* (abbreviated as 'i') describes an interaction between a trajector and a (primary) landmark that results in the landmark's entering, metaphorically speaking, a dominion (D) of attention over which a secondary landmark (lm') has control. The primary landmark is elaborated by the reflexive pronoun *herself* (abbreviated as 'f'), whose coreferentiality with the trajector is depicted by means of the dashed line connecting the trajector and the landmark (a detailed treatment of reflexivity in CG is offered by van Hoek, 1997). The next step consists in the integration of *introduce herself* with *to the audience*. The preposition *to* (abbreviated as 't'), which profiles the motion of a trajector into a region close to a landmark, is put in correspondence with the path traced out by the referent of *herself* into the secondary landmark's dominion of *introduce*. The nominal *the audience* (abbreviated as 'A') elaborates the landmark of *to*, which in turn is put in correspondence with the secondary landmark of *introduced*. The present analysis thus assumes that *introduced herself to the audience* is a classical constituent, whose overall trajector is elaborated by the nominal *Mary* (abbreviated as ‘M’). Finally, *Mary introduced herself to the audience* elaborates the landmark of *after* (abbreviated as ‘a’), which profiles a temporal relation.

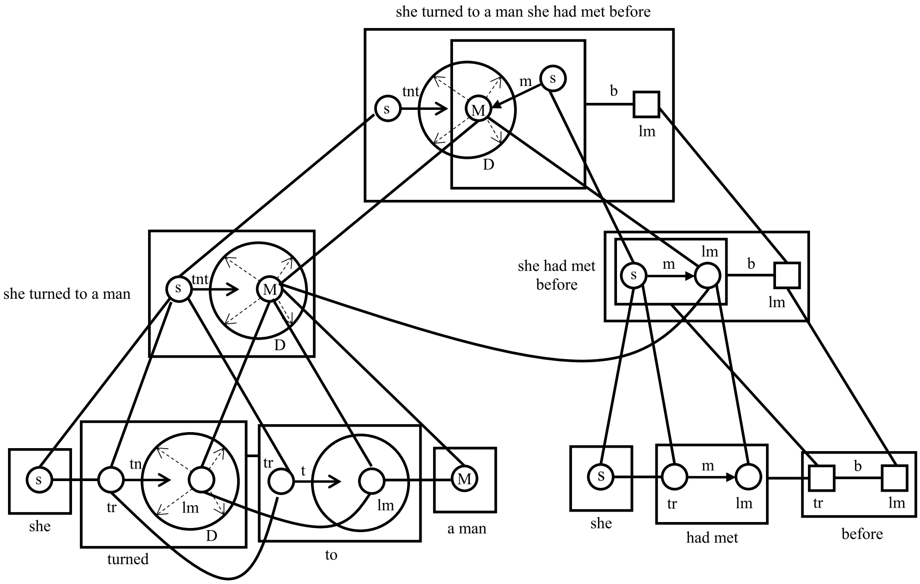


Figure 20: A compositional path for *she turned to a man she had met before*.

Figure 20 shows a plausible analysis of the “main clause” *She turned to a man she had met before*. As is apparent from this Figure, I have opted for a very flat structure. The verb *turn* (abbreviated as ‘tn’) depicts the rotation (not shown in the diagram) of a trajector, who, as in the case of the verb *introduce*, enters the dominion of attention of a landmark. The elaboration of the trajector and landmark is similar to that of the trajector and landmarks in Figure 19 (‘tnt’ stands for the merger of *turned* and *to*). The “main” clause also contains a relative clause, whose landmark corresponds to the landmark of *turned* (i. e., *a man* (abbreviated as ‘M’)). Note that the trajector of the adverb *before* (abbreviated as ‘b’) is elaborated by the clause *she had met* (the verb (*had*) *met* is abbreviated as ‘m’). At the highest level, the two clauses *she turned to a man* and *she had met before* are integrated with each other. The profile lines for both of them are heavy because, as was pointed out in connection with Figure 13, it is assumed that both clauses are accessed sequentially in basic windows of attention. (Remember, however, that a diagram like Figure 20 is not capable of showing sequential access to the two clauses, as was remarked in connection with Figure 13 above, hence the need for a more dynamic representation like Figure 14, which is not offered here for the sake of simplicity.)

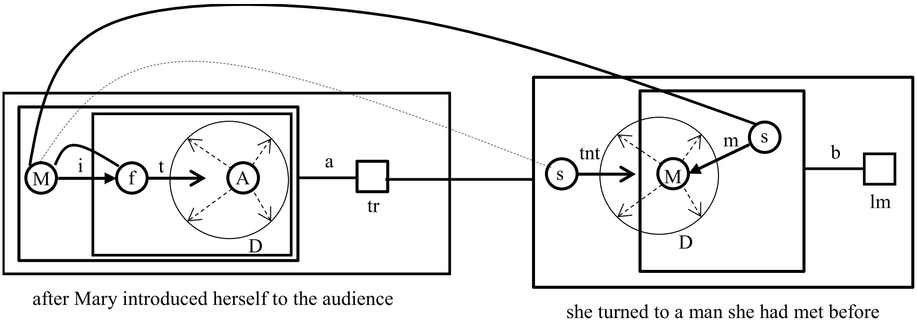


Figure 21: Sequential access in *After Mary introduced herself to the audience, she turned to a man she had met before*.

Finally, Figure 21 shows the referential identity of the two instances of *she* with *Mary* and that the temporal clause and the "main" clause are accessed sequentially and do not result in a classical constituent, i. e. there is, diagrammatically, no single box comprising both clauses in Figure 21.

References

Broccias, Cristiano. 2002. Review of Croft, William. 2001. In *Radical Construction Grammar*. Oxford: Oxford University Press. LINGUIST List: Vol–13–1516. Fri May 24 2002. http://https://linguistlist.org/issues/13/13-1516.html (accessed 24 June 2018).

Broccias, Cristiano. 2003. *The English Change Network: Forcing Changes into Schemas*. Berlin, New York: Mouton de Gruyter.

Chafe, Wallace. 1994. *Discourse, Consciousness and Time: The Flow and Displacement of Conscious Experience in Speaking and Writing*. Chicago: Chicago University Press.

Croft, William. 2001. *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press.

Evans, Vyvyan & Melanie Green. 2006. *Cognitive Linguistics: An Introduction*. Edinburgh: Edinburgh University Press.

Goldberg, Adele E. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.

Halliday, M. A. K. & Christian M. I. M. Matthiessen. 2014. *Halliday's Introduction to Functional Grammar*. London: Routledge.

Hudson, Richard. 2007. *Language Networks: The New Word Grammar*. Oxford: Oxford University Press.

Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar. Vol. 1, Theoretical Prerequisites*. Stanford: Stanford University Press.

Langacker, Ronald W. 1991. *Foundations of Cognitive Grammar. Vol. 2, Descriptive Application*. Stanford: Stanford University Press.

Langacker, Ronald W. 1999. *Grammar and Conceptualization*. Berlin, New York: Mouton de Gruyter.

Langacker, Ronald W. 2002 [1990]. *Concept, Image, and Symbol: The Cognitive Basis of Grammar*, 2nd edn. Berlin, New York: Mouton de Gruyter.

Langacker, Ronald W. 2008. *Cognitive Grammar: A Basic Introduction*. New York: Oxford University Press.

Langacker, Ronald W. 2009. *Investigations in Cognitive Grammar*. Berlin, New York: Mouton de Gruyter.

Langacker, Ronald W. 2014. Subordination in a Dynamic Account of Grammar. In Laura Visapää, Jyrki Kalliokoski & Helena Sorva (eds.), *Contexts of Subordination: Cognitive, Typological and Discourse Perspectives*, 17–72. Amsterdam & Philadelphia: John Benjamins.

Langacker, Ronald W. 2015. How to Build an English Clause. *Journal of Foreign Language Teaching and Applied Linguistics*, 5. n.p.

Langacker, Ronald W. 2016a. Toward an Integrated View of Structure, Processing, and Discourse. In Grzegorz Drodżdż (ed.), *Studies in Lexicogrammar: Theory and Applications*, 23–53. Amsterdam & Philadelphia: John Benjamins.

Langacker, Ronald W. 2016b. Baseline and Elaboration. *Cognitive Linguistics* 27. 405–439.

Langacker, Ronald W. 2016c. Working Toward a Synthesis. *Cognitive Linguistics* 27. 465–477.

MacWhinney, Brian. 2014. Conclusions: competition across time. In Brian MacWhinney, Anderj Malchukov & Edith Moravcsik (eds.), *Competing motivations in grammar and usage*, 364–386. Oxford University Press.

Talmy, Leonard. 2000. *Toward a Cognitive Semantics. 2 vols*. Cambridge, MA: The MIT Press.

Taylor, John. 2007. Cognitive linguistics and autonomous linguistics. In Dirk Geeraerts & Hubert Cuyckens (eds.), *The Oxford Handbook of Cognitive Linguistics*, 421–462. Oxford: Oxford University Press.

van Hoek, Karen. 1997. *Anaphora and Conceptual Structure*. Chicago: The University of Chicago Press.

Rui P. Chaves

3 Construction Grammar

Abstract: Broadly construed, Construction Grammar is a constraint-based, generative, non-derivational, mono-stratal grammatical approach to the modeling of linguistic knowledge that is committed to incorporating the cognitive and interactional aspects of language. The central tenet of Construction Grammar is the claim that language is a repertoire of more or less complex and conventionalized templactic patterns of phonologic, morphologic syntactic, semantics, and/or pragmatic information. Such conventionalized templates (constructions) form intricate networks of overlapping and complementary patterns that are used during comprehension and production to encode and decode linguistic expressions in context, while attending to extralinguistic information.

1 Introduction

The study of constructions and their typology has played a crucial role in linguistics since Structuralism, and some of its ideas go as far back as to the time of Aristotle. The advent of Phrase-Structure Grammar (PSG) placed constructions in a more precise footing (Harris, 1951; Chomsky, 1957), but as PSGs struggled to cope with discontinuity phenomena (Chomsky, 1975, 190), and as movement became the key mechanism for arriving at cross-constructional generalizations (Chomsky, 1981), constructions came to be seen as epiphenomena rather than explicit part of grammatical knowledge (Chomsky, 1989, 43).[1]

The idea of viewing constructions as a fundamental component of natural language emerged in the mid-eighties, with the work of Charles Fillmore and colleagues. In such a Construction Grammar (CxG) framework, the linguistic knowledge that speakers acquire includes a large system of templates or schemata consisting of conventional associations of grammatical information (including morphosyntactic, semantic, pragmatic, and/or phonological information), assumed to range from the totally regular to the totally idiosyncratic, and can be lexical, phrasal or in between. In CxG the term 'construction' is used to refer to the templates that comprise the grammar, and the term 'construct' refers to the utterances structures built from those tem-

1 The motivation for movement remains controversial, however. See Borsley (2012) for empirical criticism.

Acknowledgement: I am very grateful to Laura Michaelis, Adele Goldberg, Jean-Pierre Koenig and Erika Bellingham for comments and discussion about an earlier draft, and to the editors and reviewers for their feedback. I am responsible for the views reflected here and for any errors that remain.

https://doi.org/10.1515/9783110540253-003

plates. However, in the years that followed Fillmore's work the term CxG came to mean slightly different things to different linguists. See for example Berkeley Construction Grammar (Fillmore et al., 1988; Fillmore and Kay, 1996; Kay and Fillmore, 1999), Cognitive Construction Grammar (Goldberg, 1995, 2006), Radical Construction Grammar (Croft, 2001), Embodied Construction Grammar (Bergen and Chang, 2005; Feldman et al., 2009), Fluid Construction Grammar (Steels, 2011), and Sign-Based Construction Grammar (Michaelis, 2012; Sag, 2012), among others. This paper provides an overview of the tenets and evidence for CxG grammar, as well as a formal and computational fragment to illustrate how a constructional account can be articulated.

2 Goals

Like many other approaches to language, the goal of CxG is to arrive at an explicit, contradiction-free, and generalization-prone model of natural language which has the widest possible empirical coverage, from the highly idiomatic and rigid to the fully productive and compositional. Hence, in a construction-based conception of language, it is to be expected that some regular clausal types have both regular and idiosyncratic uses. Mismatches between form and function such as (1), for example, are therefore not surprising from a constructivist perspective, and cannot be dismissed as mere marginalia. In fact, their hybrid status can shed light on phenomena that would otherwise remain undetected.

(1) a. What does she care? (*assertion*)
b. Why don't you just be quiet? (*command*)
c. Don't tell me you lost the keys again! (*interrogative*)
d. I don't suppose you'd like to buy this from me. (*interrogative*)

Consequently, there is no methodological separation between 'core' and 'peripheral' phenomena in CxG; a complete theory of any given natural language must account for all linguistic facts, including the interaction of highly idiosyncratic constructions with other, more regular ones, as the former and the latter are inextricably interdependent.

Second, CxG aims to be is maximally consistent with the available psycholinguistic and cognitive evidence about human language acquisition and processing. The constructivist null hypothesis is that grammars are composed of constructions, nothing else, and that they are acquired without a language-specific genetic endowment. More specifically, constructivist theories make the following claims.

I. Constructions are form-function templates that are stored in the mind of speakers as part of their grammar, acquired from the input via general cognitive mechanisms, and restricted by the stages of brain development. Such linguistic knowledge is to some extent processing-neutral, and deployed during both comprehension and production.

II. Constructions can introduce lexical, syntactic, semantic, pragmatic, prosodic constraints over and above those contributed by the expressions they combine, and induce varying degrees of regularity. For example, the same word is compatible with a wide range of different subcategorization patterns and novel interpretations.
III. The wellformedness of a complex linguistic expression is a matter of simultaneous constraint satisfaction, sensitive not only to various kinds of grammatical knowledge, but also knowledge of the discourse context, world knowledge, gestures and other kinds of visual information, social knowledge, and knowledge of style and genre.
IV. Constructions are clustered into networks, much like those assumed to represent non-linguistic knowledge in the mind, enabling generalizations that permit speakers to understand, acquire, and produce novel structures, sometimes through analogy.
V. Constructions exhibit degrees of language-internal irregularity, and vary across language families and genera. Typological patterns are likely due to historical, functional, and cognitive factors rather than language-specific genetic endowment.

The idea that grammars contain large inventories of constructions may appear to some researchers as a step backwards, away from deeper generalizations. There are several flaws with such a view. First, grammars that lack constructions come at the cost of increased complexity in other theoretical components and of limited empirical coverage (Johnson and Lappin, 1999; Culicover and Jackendoff, 2005). In other words, once a sufficiently large range of syntactic phenomena is taken into consideration – including the more idiosyncratic – the conclusion that constructions are a component oh human language is difficult to avoid. Second, a grammar consisting of a rich network of constructions is arguably a more cognitively plausible model of the linguistic knowledge that speakers *de facto* acquire and use during language processing. Embodied Construction Grammar (Bergen and Chang, 2005), for example, goes as far as focusing not just on what constructions are but on how they are used, as one of its primary goals to understand what cognitive and neural mechanisms do speakers engage while using human language. See for example Bryant (2008) for a psychologically-plausible best-fit probabilistic construction-based model of parsing and interpretation that aligns well with behavioral (human) sentence processing data.

Hence, CxG is in principle experimentally testable: if the linguistic knowledge in the brains of speakers does not include a large repertoire of constructions, then the constructionist view of grammar would be deemed incorrect. In this sense, CxG is closer to being an implementation-level theory of language, borrowing the terminology of Marr (1982, 25). See also Jackendoff (2002, ch. 2) on the distinction between ‘hard’ and ‘soft’ idealizations.

To be sure, there is no conceptual, linguistic or psychologic obstacle with assuming that grammatical knowledge involves a large repertoire of constructions. The number of lemmas that the average adult native speaker of American English knows has been estimated to be around 40,000 (Brysbaert et al., 2016), for example, and therefore it is not unreasonable that they also learn hundreds of grammatical constructions. Even a small sample like (2) should suffice to illustrate the range of constructions that one and the same verb can appear in. As Goldberg (2006, 18) put it, ‘it is constructions all the way down’.

(2) a. Sam laughed. (*strict intransitive*)
 b. Sam laughed his maniacal laugh. (*cognate object*)
 c. Sam laughed the lyrics (rather than singing them). (*transitive*)
 d. Sam laughed her his promise. (*ditransitive*)
 e. Sam out-laughed Robin. (*comparative compound*)
 f. Sam laughed the kids off the stage. (*caused motion*)
 g. Sam laughed about the incident. (*cause*)
 h. Sam laughed at me. (*directional*)
 i. Sam laughed her way out of the room. (*way-manner*)
 j. Sam laughed all the way to the bank. (*way-path*)
 k. Sam laughed her throat hoarse. (*resultative*)
 l. Sam laughed herself to tears. (*fake reflexive resultative*)
 m. Sam laughed her head off. (*off-resultative*)
 n. Sam laughed the idea off. (*phrasal verb idiom*)
 o. Sam laughed it up. (*particle idiom*)
 p. Sam laughed and laughed... (*X-and-X intensification*)
 q. Sam laughs, and the world laughs with her. (*X and Y Xs with K*)
 r. Sam laughed: *ho! ho! ho!* (*sound emission*)

Many of the above constructions involve idiosyncratic meaning and structure, both of which must be stipulated somewhere in the grammar, regardless of which theory one adopts. The advantage of CxG is that it allows the linguist to capture the regularities and irregularities more directly, at the level of the construction that captures the relevant patterns.[2] The phenomena in (2) underscore an important problem that CxG aims to tackle head-on: once a sufficiently broad range of linguistic phenomena are considered, simple overarching generalizations tend to vanish and a wide range

2 Such grammars have a better chance of being acquired via statistical learning than those relying on highly abstract information for which the learners have no direct observable evidence (Fodor and Sakas, 2001; Newmeyer, 2004; Clark and Lappin, 2011).

of variation and idiosyncrasy often emerges, with different degrees of structural and semantic sub-regularity.

Crucially, constructions can interact with others in very complex ways. For example, in (3) we see various uses of *laugh* interacting with extraction, raising, control, and passivization. Thus, any account of (2) must also take into account the myriad of ways in which verbal arguments can be alternatively realized.

(3) a. It was the kids who Sam supposedly tried to laugh off the stage.
b. What Sam seemed to be laughing was his maniacal laugh.
c. Pictures were taken, laughs were laughed, and food was eaten.

Importantly, such interactions sometimes reveal constraints and phenomena that would otherwise remain undetected. For instance, it is not obvious why the complement of (2r) can be clefted as in *It was* [*ho! ho! ho!*] *that Sam laughed* _ , but the complement of the *way*-Manner construction in (2i) cannot, viz. **It was* [*her way*] *that Sam laughed* _ *out of the room*. Similarly, the complement in (2n) can be extraposed *Sam laughed* _ *off* [*the idea*], but not that of (2m), viz. **Sam laughed* _ *off* [*her head*], and so on.

The key to a constructivist account of phenomena like (2) and (3) is the recognition that different kinds of construction impose constraints on different kinds of linguistic dimensions. Thus, some constructions govern how semantic arguments are linked to morphosyntactic categories (Linking Constructions), others govern the range of possible grammatical roles that such categories can have (Valence and Voice Constructions), and so on. For example, a standard assumption in CxG is that there is one lexical entry for verbs like *laugh*, neutral with regard to the possible realizations in (2); see for example Goldberg (1995, 50,99), Croft (2001, 54,168), Fillmore and Baker (2009, 120), and Sag (2012, 133–139) for different implementations of this insight, consistent with the fact that thematic roles and subcategorization frames associated with a verb are available shortly *after* that verb is accessed, as experimentally shown by Boland (1993), Trueswell and Kim (1998) and others. Such an underspecified lexical entry is a construction in itself, but one that concerns a single lexeme. If combined with the intransitive construction we obtain uses like (2a), if combined with the ditransitive construction we obtain (2d) and so on, as informally depicted in Figure 1. As an analogy, suppose that each of the boxes below is a transparency that can be overlaid on top of another. As long as the result is legible, their combination is well-formed. Of course, each construction can impose particular morphosyntactic, semantic, pragmatic and/or phonological constraints on the word it combines with.

Although the lexical entries of most words are underspecified as depicted above, and therefore can be used in a wide range of ways, the lexical entries of other words specifies additional constraints that restrict the range of constructions that they can combine with. Thus, the only lexical entry for the verb *rumor* is intrinsically passive (compare **We rumored Kim to be rich* with *Kim was rumored to be rich*). In other cases

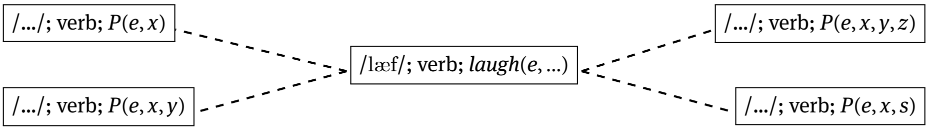


Figure 1: An underspecified word and constructions it can combine with.

still, it is up to particular constructions to introduce idiosyncratic constraints. Thus, certain uses of *assure* require an argument to be *ex situ* (compare **I can assure you him to be the most competent* with *Who can you assure me _ to be the most competent?*), and obligatory transitive verbs like *devour* can drop their object only when used with constructions like the *way* construction (e. g. compare **He devoured* with *He devoured his way to victory by eating dozens of roaches*[3]).

For Kay (2002), Bergen and Chang (2005), Fillmore and Baker (2009), Steels (2011) and others, the operation responsible for combining constructions is unification (Shieber, 1986; Carpenter, 1992). Hagoort (2003, 2005) interprets various electrophysiological and neuroimaging findings in terms of a unification-based process that acts on syntactic, semantic, and phonological representations simultaneously. As in CxG, words are stored in the lexicon as part of a template, and that parsing involves a single combinatorial operation (unification) that joins such templates. Others like Goldberg (1995) and Croft (2001) are less committal about the nature of the operation that instantiates information across constructions.

Lexical entries such as the one at the center of Figure 1 are taken to be the result of grammatical generalizations made during acquisition. Learners eventually abstract the lexical entry of 'laugh' away from its multiple uses, and arrive at a number of templates that can be used for other verbs as well. In some cases, particular constructional realizations are so frequent that they become integral part of the grammar, rather than computed on-the-fly. In most versions of CxG, the constellation of constructions that can constrain a lexical construction forms a network, based on the information that such constructions have in common. For example, for Goldberg (1995, 135) and Goldberg (2006) such networks act as attractors and play an important role in giving rise to generalizations across verb classes during language acquisition, as well as in the coining of novel uses.

3 Data

Empirical adequacy, generality, simplicity, psychological reality and alignment with data about language acquisition, usage, historical change, and the evolution of language are all relevant sources of data to consider in rejecting or accepting a given

3 http://www.browardpalmbeach.com/news/edward-archbold-guy-who-dropped-dead-after-roach-eating-contest-died-of-asphyxia-6466687

constructivist account. In particular, CxG is in principle experimentally testable: psycholinguistic and neurolinguistic evidence should be brought to bear to determine if an analysis is consistent with the behavioral facts; for recent in-depth discussion see Goldberg (2019).

For example, there is much evidence that even compositional expressions can attain independent representation in the mental grammar, as a way of making their processing more efficient (Corrigan et al., 2009). For example, Alegre and Gordon (1999) and various others found wholeword frequency effects for regularly inflected words, suggesting that such wordforms can be memorized, and Bannard and Matthews (2008) showed that two and three-year-olds were faster and better at repeating higher frequency phrases compared to lower frequency ones, even though the two strings were equally plausible and matched on all other frequency measures. In addition, there is a growing body of historical evidence suggesting that complex forms can be memorized (Traugott and Trousdale, 2014; Bybee, 2006).[4] The process of storing the output of a commonly used function so that the solution can be simply looked up rather than computed from scratch is called *memoization* in computer science and *chunking* in psycholinguistics, and its redundancy provides a simple, robust, and efficient solution to a hard computational problem. This is one of the aspects of CxG which places it somewhere in-between a computational level theory of language (i. e. one that abstracts away from processing details) and an algorithmic level theory (which does not). The explicit goal of some CxG approaches, like that of Bergen and Chang (2005), for example, is precisely to bridge the gap between these levels. Indeed, construction-based frameworks have been informed by models of collocation analysis (Stefanowitsch and Gries, 2005), acquisition and syntactic processing (Abbot-Smith and Tomasello, 2006), computational modeling of concept learning (Steels and Beule, 2006), and models of activation of neural motor programs during perception (Feldman and Narayanan, 2004), among others.

If constructions are in fact part of the knowledge of language and in some cases can contribute with meaning over and above the meanings of the expressions they combine then there should be many linguistic examples of non-idiomatic structures in which the meaning of the whole is greater than that of its parts and most straightforwardly analyzable as the result of a constructional rule. I now turn to such evidence below.

3.1 Lexical constructions

Productive reduplication morphology is perhaps the strongest kind of evidence for constructional approaches, as it is cross-linguistically widespread and typically in-

4 See Bybee (2013) and Diessel (2015) for more on the compatibility of usage-based approaches with CxG.

volves idiosyncratic meaning (Ghomeshi et al., 2004; Inkelas and Zoll, 2005; Masini and Thornton, 2008; Kay and Zimmer, 1990). For example, in Afrikaans complete reduplication is productive, as all lexical categories can be reduplicated to mean 'increase':

(4) Bakke-bakke veld-blomme versier die tafels
Bowls-bowls wild-flowers decorate the table
'The tables are decorated with wild flowers by the bowlful'
(Botha, 1988, 92)

The construction for reduplication of plural nouns can be assumed to be [N_{pl} N_{pl}] and to mean 'many Ns'. The alternative, of course, is to more indirectly stipulate the existence of a zero-affix that introduces the 'increase' denotation and selects two plural nouns. However, the very existence of reduplicative patterns such as these is predicted by the constructional approach to grammar. Similarly problematic for the zero-affix approach is the case of productive exocentric VN compounds in Romance languages, such as European Portuguese:

(5) a. lava-pratos
wash-dishes
'dish washer'

b. lava-carros
wash-cars
'car washer'

c. lava-janelas
wash-windows
'window washer'

d. lava-sanitas
wash-toilets
'toilet washer'

These compounds are nominals that denote an agent, not the action or the patient. As Booij (2010, 37) notes, there is no independent motivation for postulating a nominalizing zero-suffix, other than theory-internal assumptions.

Some productive compounding constructions similarly exhibit exotic structure as well as idiosyncratic meaning. This is the case of *Paired-Argument* compounds, illustrated by (6), from Jackendoff (2010). Such compounds involve two nominals that combine exocentrically to form a collective of sorts, which is then interpreted reciprocally by the following noun.

(6) a. a [love-hate] relationship

b. a [Port-cornstarch] mixture

The simplest account of such data is one where a dedicated construction imposes the appropriate form-meaning constraints, given that nothing else in the grammar derives these from independently motivated mechanisms.

But even more canonical compounding processes often exhibit peculiar structural and semantic constraints. For example in (7) the first noun is interpreted as a generic kind, and the second noun is interpreted as having been created with a benefactive goal. Hence, expressions like (7a) are interpreted as 'food created for generic dogs'.

(7) a. dog food
  b. baby diapers
  c. car seat

The compounds in (8), however, establish a meronymy relation between the first nominal and the head nominal, and existentially quantifies the former. Jackendoff (2010) argues that there are at least fourteen classes of semantic relationships in English compounds, and for evidence concerning adjective-noun combinations see Langacker (1987).

(8) a. cheese omelet
  b. brick building
  c. pebble beach

Interestingly, such compounds can interact with other compounds, such as the Numeral-N compounding construction illustrated in (9). Here we see a rather unusual combination of a plural numeral expression with a nominal root, which in turn combine with a nominal head.

(9) a. a [[two cheese] omelet]
  b. this [[six valve] engine]
  c. that [[ten story] building]
  d. one [[five page] letter]

Nothing requires the numeral expression to be simple, as illustrated by *a* [[[*two-hundred thousand*] *mile*] *race*], or for it to be plural, e. g. *a* [[*one party*] *state*], *a* [[*one man*] *show*], *a* [[*no cholesterol*] *omelet*], or *a* [[*zero latency*] *engine*].

3.2 Phrasal constructions

Constructions in which the meaning of the whole is richer than that of its parts are in no way restricted to compounding as the sample in (10) illustrates. A straightforward account of such phenomena is one where each construction is brought about by a different grammatical template, with its own selectional constraints, semantic contribution, and/or prosodic phrasing. Either way, the grammar must be made more complex, as there is no way to derive the above patterns from more general rules.

For construction-based accounts of (10a), for example, see Culicover and Jackendoff (1999), and Borsley (2004).

(10) a. The more you drink, the drunker you'll get.
(*Comparative correlative*)
(= 'If you drink more, you will get proportionally drunker')

b. It's a joke the way they run that place.
(*Extraposed exclamative*)
(= 'The way in which they run that place is a joke')

c. Miserable week after miserable week, we memorized the entire play, paragraph by paragraph, word for word.
(*N-P-N construction*)
(= 'During several weeks, we memorized every paragraph and word in the play')

d. They are planning to get engaged, war or no war.
(*X or no X*)
(= 'They are planning to get engaged, regardless of there being a war or not')

e. I like him, but I don't LIKE-HIM–like-him.
(*Focus reduplication*)
(= 'I like him to a moderate degree only')

Other phenomena that are consistent with a constructional account are illustrated in (11), which suggest a coertion-based approach via unary-branching rules like 'S → NP' that add the appropriate semantics and introduces the appropriate morphosyntactic information. Various other phenomena require such unary branching rules, such as bare NPs, grinding/packaging alternations, name-to-common-noun shifts, etc. See Fillmore and Kay (1996), Ginzburg and Sag (2000), Michaelis (2003), and Fillmore and Baker (2009) for specific proposals.

(11) a. A: Who owns a dog?
B: [Kim], and it's a dachshund. (= 'Kim owns a dog, and it's a dachshund')

b. A: Does Tom know ROBIN?
B: No, [Frank]. (= 'No, he knows Frank')

c. A: What do you think Robin wants?
B: Probably [Drugs]. (= 'Robin probably wants drugs')

One major advantage of construction-based approaches concerns the ability to model systematic constructional relations across constructions. Consider Subject-Auxiliary Inversion (SAI), for example, seen in (12).

(12) a. You have read the paper? (*non-inverted*)

b. Have you read the paper? (*inverted*)

Inverted and non-inverted uses of the verb can be modeled without any appeal to movement operations (Fillmore, 1999). For example, suppose that verbs come with an attribute that indicates whether they occur in inverted or non-inverted verbal structures. Thus, non-auxiliary verbs are lexically specified as INV-, and and (most) auxiliary verbs are underspecified. Hence, both kinds of verb can appear in the VP construction shown in (13), because no constraint is imposed on the value of INV. The PSG rule format adopted in (13) is highly simplified, but will do for the present purposes.[5]

(13) VP CONSTRUCTION
$\mathrm{VP} \rightarrow \mathrm{V}_{\mathrm{INV-}}\ X_1 \ldots X_n$

Thus, (13) licenses both [*lifted* [*Kim*]] and [*will* [*lift* [*Kim*]]. The type of complement is restricted by the verb's semantics. In turn, the Subject construction in (14) combines such verb phrases with a specifier in order to saturate the semantic arguments and obtain a clause.

(14) SUBJECT CONSTRUCTION
$\mathrm{S} \rightarrow X\ \mathrm{VP}$

Again, regardless of the verb being auxiliary or not, the construction in (14) derives both [[*We*][*lifted* [*Kim*]]] and [[*We*][*will* [*lift* [*Kim*]]]], by combining subjects and verb phrases. Finally, inverted structures are obtained via the construction in (15), which requires INV+ verbs. Hence, *Can we go* is licit because the verb is lexically underspecified for INV, but **Try we to go* is not licit because the verb is lexically specified as INV-.

(15) SAI CONSTRUCTION
$\mathrm{S} \rightarrow \mathrm{V}_{\mathrm{INV+}}\ X_1\ \ldots\ X_m$

Most auxiliary verbs are therefore compatible with (13) and (15) because they are lexically underspecified for INV. Hence, the same verb is free to appear in inversion and non-inversion constructions, without necessitating any movement operation or anything equivalent to it. Moreover, such an account allows for a range of lexical exceptions. For example, *aren't* must be specified as INV+ so that it is only compatible with (15) (e. g. *Aren't I invited?* / **I aren't invited?*), the auxiliary *better* is specified as INV- so it is only compatible with (13) (e. g. *You better not cry* / **Better you not cry?*), and whereas future *shall* is INV- (e. g. *I shall go downtown* = 'I will go downtown'), deontic *shall* is INV+ (e. g. *Shall I go downtown?* = 'I should go downtown?). Clark and Eyraud (2007) and others show that SAI phenomena can be learned automatically us-

5 The reader can assume that the semantics of the mother is the combination of the semantics of the daughters plus semantics contributed by the construction (if any). In CxG, constructions are nothing but static information about the internal structure of grammatical units, very much like phrase-structure grammar rules if viewed as declarative statements about mother-daughter configurations, rather than as string-rewriting functions (McCawley, 1968; Gazdar, 1982). Hence, constructions like (13) can be used by both the production and comprehension modules.

ing phrase structure grammar fragments of this kind, without explicit instruction, on the basis of a small set of positive data, and Bod (2009) obtains a similar result using child-directed data from CHILDES.

CxG takes the account just sketched above to the next level, by recognizing that there are in fact many SAI constructions, not just one like (15). As illustrated in (16) these have different distributions, function, and varying degrees of idiomaticity (Fillmore, 1999).

(16) a. [Shall we leave]?
(*canonical Y/N interrogative*)

b. Where [did they go]?
(*canonical wh-interrogative*)

c. What [does it matter if it's 2pm or 3pm]? We're late!
(*idiomatic wh-interrogative*)

d. [Wasn't that brave of him]?!
(*idiomatic exclamative*)

e. (Wow/Boy,) [can she sing]!
(*modal exclamative*)

f. [Don't you be late young man]!
(*negative imperative*)

g. (Oh) [don't I know it]...
(*expletive negation idiom*)

h. [May you have a fantastic birthday].
(*blessings/curses*)

i. The course was more confusing than [was the workshop].
(*comparative*)

j. [Had you warned me], we would be ready by now.
(*counterfactual conditional*)

k. [Should there be a storm], we will stay indoors.
(*concessive modifier*)

l. Rarely [had we seen them laugh so hard].
(*adverbial fronting*)

m. You're curious, and so [are we].
(*fronted so/as/neither conjuncts*)

A different SAI construction is responsible for each of the clause types in (16), all having in common the same SAI form seen in (15). These constructions thus form a form-resemblance family, and are typically depicted in CxG as a hierarchy with '$\mathrm{S} \rightarrow \mathrm{V}_{\mathrm{INV+}}\ X_1\ \ldots\ X_m$' at the top, or alternatively, at the center of the cluster. Although

Fillmore (1999) argues that there is no general semantics shared by all aux-initial constructions, this is a controversial point, however. For Goldberg (2006, 179), for example, the construction in (15) is associated with certain functional properties; see also Goldberg (2009), Borsley and Newmeyer (2009), and the references cited there.

For Kay (2002) and others the constructional hierarchies that are common in CxG research are best seen as a (non-redundant) taxonomic characterization of the constructional knowledge shared by a cluster of constructions. In practice, it is the cluster of constructions that matters for speakers, not abstract taxonomies. The latter capture all the generalizations potentially available to the speaker of a language, though it is not assumed that the internal representation of the language in the mind of each speaker contains every generalization inherent in the data. Variability among speakers is an appropriate research question for psycholinguistics and variation studies. It is the grammarian's job to lay out the initial possibilities by identifying the full range of candidates. Thus CxG predicts that speakers of English have at least 13 constructions that share the general form in (15), subsets of which have semantic and pragmatic properties in common with each other and/or with non-SAI constructions in addition to their own morphosyntactic, semantic, pragmatic and/or prosodic idiosyncrasies (Langacker, 1987). Again, this is one of the aspects of CxG that places it somewhere in-between a theory of language that abstracts away from processing details and an algorithmic level theory, since it aims to be consistent with the knowledge that is *de facto* used by speakers during sentence processing.

4 Tools

CxG aims at a comprehensive description of the grammar of all languages, focusing both on regular and irregular constructions, and on their interaction. There is generally no limit to the size of local syntactic structures, as they can be unary branching, binary or longer. However, different constructional variants adopt different formalisms and different representational tools. As already mentioned, some assume the grammar is unification-based, others do not. Some adopt grammatical relations (such as subject and object), others do not.

For some constructivist researchers, CxG is 'generative' in the original sense of Chomsky (1965, 4), whereby a grammar is nothing but an explicit formal statement of the rules of the language, but for others the distinction between competence and performance is rejected. Like other monostratal frameworks (e. g. LFG, HPSG, CCG, etc; see present volume), all dimensions of linguistic information – phonology, morphosyntax, semantics and pragmatics – co-exist in lexical entries, grammar rules, and in the expressions licensed by the grammar. Moreover, all linguistic information, be it lexical or otherwise, is represented with the same basic format. Some constructivist approaches adopt featural representations, others reject them and remain noncommittal about how best to represent linguistic information.

5 Evaluation

Although the idea that a same small set of mechanisms derive all constructions in all languages of the world is conceptually appealing, constructivist researchers find it difficult to justify, given the extremely wide range of typological variation across the languages of the world (Dryer, 1997; Croft, 2001; Hawkins, 2004; Haspelmath, 2007). Rather than assuming that all of this intra- and cross-linguistic diversity and idiosyncrasies should be the product of one and the same hyper-abstract language module that has somehow come to be part of the human genetic endowment, CxG is more conservative in that it assumes that human grammars differ to the extent that their respective languages differ.[6] And since CxG does not assume that all languages must be described in terms of the same core components, linguistic tests and even the repertoire of parts of speech can be language-specific (Croft, 2001). Recurrent cross-linguistic patterns are argued to be best described in terms the result of historical, functional, and cognitive pressures (Dryer, 1997; Croft, 2001; Hawkins, 2004; Newmeyer, 2005; Culicover and Jackendoff, 2005; Goldberg, 2006; Haspelmath, 2007). Thus, the constructional grammarian usually first focuses on each language independently, in their own terms, and only later identifies any emergent cross-linguistic similarities. For construction-based crosslinguistic research on word order, for example, see Kathol (2000, Ch. 7) and Wetta (2014).

6 A Grammar Fragment

It is impossible to do justice to all extant variants of CxG, as they differ in the formalism they adopt, their degree of explicitness, and in the analysis of certain phenomena (Hoffman and Trousdale, 2013). In what follows is a formally explicit grammar

6 It is now known that 98.8 % of the human and chimp gene sequences are identical, not merely similar, and that the only human-specific genes that concern the brain simply govern the number of rounds of cell division during fetal brain development (Liu et al., 2012): Whereas cortical synaptogenesis in humans extends to five years it is only a few months in chimpanzees and macaques. Genes like FOX2P are in fact not specific to humans (Enard et al., 2002), and their mutation causes a diffuse range of effects, including problems with movement of the face and mouth unrelated to language, and significantly reduced IQ in the non-verbal domain (Vargha-Khadem et al., 1995). For more discussion see Elman (1999), Marcus (2001), Marcus and Fisher (2003), and Atkinson et al. (2018).

fragment that incorporates insights from a variety of sources, such as Fillmore and Kay (1996), Kay and Fillmore (1999), Goldberg (1995), Croft (2001), and Bergen and Chang (2005). Although the Sign Based Construction Grammar formalism (Sag, 2012) is adopted here in broad terms, various revisions are made in order to come closer to the spirit of the aforementioned strands of CxG, and to better highlight the key differences between CxG and other frameworks in this volume. To be clear, what follows is a grammar of English, rather than a grammar that is supposed to also model languages unrelated to English. Thus, the attributes and rules shown below should not be assumed to be necessarily appropriate for all other languages.[7]

Let us begin with a snapshot of the information that characterizes an English verb and how it is organized. The Attribute-Value Matrix (AVM) in Figure 2 states that signs of type *word* are composed of several major dimensions of linguistic information: FORM (phonology), SYN (morphosyntax), FRAMES (semantics), and ARG-ST (argument structure).[8]

$$\begin{bmatrix} \textit{word} & \\ \text{FORM} & \begin{bmatrix} \textit{w-phon} & \\ \text{PHON} & \textit{list(phoneme)} \\ \text{PREFIX} & \textit{list(phoneme)} \\ \text{STEM} & \textit{list(phoneme)} \\ \text{SUFFIX} & \textit{list(phoneme)} \end{bmatrix} \\ \text{SYNTAX} & \begin{bmatrix} \textit{syn} & \\ \text{CAT} & \begin{bmatrix} \textit{verb} & \\ \text{VFORM} & \textit{vform} \\ \text{INV} & \textit{boolean} \end{bmatrix} \\ \text{VAL} & \textit{list(sign)} \\ \text{GAP} & \textit{list(sign)} \\ \text{REL} & \textit{list(index)} \end{bmatrix} \\ \text{FRAMES} & \textit{list(frame)} \\ \text{ARG-ST} & \textit{list(sign)} \\ \text{DTRS} & \textit{list(word)} \end{bmatrix}$$

Figure 2: General attribute-value geometry of English verbs.

7 A small computational grammar fragment created to illustrate the implementation of the present theory can be downloaded from https://github.com/RuiPChaves/SBCG

8 Information structure is omitted due to space limitations; see Engdahl and Vallduví (1996) for example.

The notation [PHON *list(phoneme)*] means that the value of PHON is required to be a list of phonemes. The value of PHON is oversimplified here since phonological representations have complex structure (syllables, feet, prosodic words, intonational phrases, etc.). The type *phr-phon* differs from *w-phon* in that it lacks *pref*, SUFF, and STEM attributes, and instead has phonological attributes that are appropriate for phrases. Analogously, the notation [ARG-ST *list(sign)*] requires that the value of ARG-ST must be a list of signs, which can be wither AVMs of type *word* or *phrase*. The value of any given attribute is typed (shown in italics) and some types introduce further attributes. Whenever an attribute has no additional information beyond that shown in Figure 2, it is referred to as being 'underspecified' and the attribute is not made explicit in the AVM, for exposition purposes. Thus, applying this convention to the AVM in Figure 2 would result in an AVM that shows only the non-underspecified information, as depicted in Figure 3.

$$\begin{bmatrix} \textit{word} & \\ \text{SYNTAX} & \begin{bmatrix} \textit{syn} & \\ \text{CAT} & \begin{bmatrix} \textit{verb} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Figure 3: Attribute-value geometry of English verbs (underspecified attributes omitted).

The attribute D(AUGH)T(E)RS lists the sign's daughters. For example, verbal compounds like *body-shame*, *slut-shame*, *dog-shame*, etc. are verbs that have two daughters: a root and a verb, as licensed by a binary branching verbal construction. Conversely, words like *dog* have no daughters, and therefore the value of their DTRS attribute is the empty list ⟨⟩. For exposition purposes, the DTRS attribute is depicted more conventionally from now on, as in Figure 4. The tree notation pertains to combinatorical signs licensed by the grammar (constructs) whereas the PSG notation pertains to the rules (constructions) that license them. Following Sag (2012), constructs are displayed inside a box.

As a more concrete example, consider the verb *laugh* given in (17). This construction characterizes the idiosyncratic association of form, morphosyntax and meaning

$$\begin{array}{c} \text{X} \\ \text{Y}_1 \;\; \ldots \;\; \text{Y}_n \end{array} \quad \text{is a shorthand for branching constructs} \quad \boxed{\begin{bmatrix} \text{X} & \\ \text{DTRS} & \left\langle \text{Y}_1,\ldots,\text{Y}_n \right\rangle \end{bmatrix}}$$

$$X \rightarrow Y_1 \ldots Y_n \quad \text{is a shorthand for branching constructions} \quad \begin{bmatrix} \text{X} & \\ \text{DTRS} & \left\langle \text{Y}_1,\ldots,\text{Y}_n \right\rangle \end{bmatrix}$$

Figure 4: The representation of immediate dominance information in AVM format.

pertaining to a particular lexeme. This AVM states that there is a uninflected, uninvertable, verb stem /læf/ describing an action frame that involves a laughing state-of-affairs. This word is a verb because the value of CAT(EGORY) is of type *verb*, uninflected because no information about PREFIX and SUFFIX is given, and uninvertable because the value of INV is '–'. As stated above, underspecified information is not shown in the AVM, for perspicuity.[9]

(17) THE 'LAUGH' LEXICAL CONSTRUCTION

$$\begin{bmatrix} \textit{word} & \\ \text{FORM} & \begin{bmatrix} \textit{w-phon} & \\ \text{STEM} & \langle \text{l,æ,f} \rangle \end{bmatrix} \\ \text{SYN} & \begin{bmatrix} \textit{syn} & \\ \text{CAT} & \begin{bmatrix} \textit{verb} & \\ \text{INV} & - \end{bmatrix} \end{bmatrix} \\ \text{FRAMES} & \left\langle \begin{bmatrix} \textit{action-pred-fr} & \\ \text{PRED} & \textit{laugh} \\ \text{INDEX} & e \end{bmatrix} \ldots \right\rangle \\ \text{DTRS} & \langle \, \rangle \end{bmatrix}$$

Semantic representations are cast in Frame Semantics (Fillmore, 1982, 1985; Fillmore and Baker, 2010; Fillmore et al., 2012), a framework which crucially assumes that meanings are relativized to rich representations that go beyond typical lexical semantics and include broader situational information. Thus, the *action-pred-fr* frame type has many sub-types, and a wide range of frame elements as illustrated in Figure 5.[10] The type *monadic-pred-fr* allows no additional frame elements other than those introduced by *action-fr* and hence it corresponds to an intransitive use of the verb. The type *action-process-fr* introduces a theme participant and therefore licenses transitive uses. Similarly, the *action-result-fr* frame type corresponds to uses where the predicate causes the theme to undergo a change as in *Sam laughed the beer out of his nose*.

Meaning postulates constrain how such arguments can be interpreted, given the semantics of the predicate. For example, as their name indicates, sound emission verbs like *laugh* involve the production of a sound signal, and as such the theme is required to describe a sound (e. g. *Sam laughed 'ha ha ha'*, *Sam laughed the lyrics*, *Sam laughed a nervous laugh*). Hence, *Sam laughed the shoe* is not felicitous because the

9 The ellipses '...' in FRAMES indicate that the list may or not contain other frames. Departing from Sag (2012), the first frame in the FRAMES list corresponds to the semantics of the head, in order to simplify the syntax-semantics interface. See below for discussion.

10 The type *sit-index* is for events and states *e* and *s*, *i-index* is for individuals *x*...*y*, and *t-index* for temporal indices.

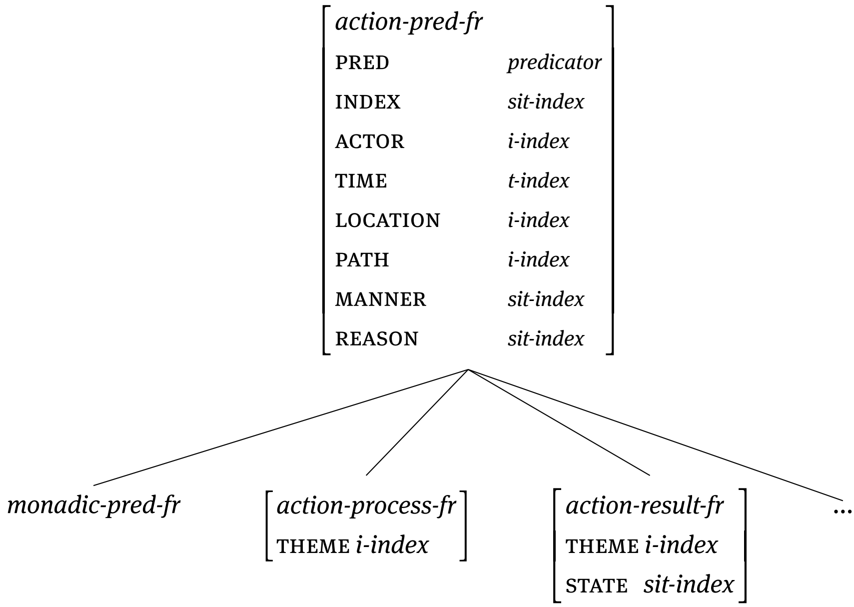


Figure 5: Type hierarchy of action-predicate semantics (not shown in full).

theme is incompatible with the constraints imposed on the theme of a sound emission verb by predicate-specific meaning postulates. Thus, not all semantic frame elements are equally compatible with all predicators. For example, the sentence *Sam laughed my Nikes threadbare* is not felicitous because of world knowledge: sound emission actions such as laughing cannot ordinarily have the described effect on sneakers, except in contrived contexts where, for example, the soles of the sneakers in question happen to be molecularly so extremely unstable that human speech suffices to cause them to shed material.[11] In other words, the compatibility between the predicate and its frame is a matter of degree, constrained by meaning postulates, and contingent on contextual and extralinguistic information.

How semantic frame elements map into morphosyntactic categories is the purview of linking constructions such as the one in (18), which bind each variable in FRAMES to a an argument structure element, ordered in terms of obliqueness. Note that the morphosyntax of core frame elements like 'actor' and 'theme' is more specific than that of non-core frame elements like state, time, and location. This is because the latter can be realized by phrases of varying categories, whereas the former are required to be NPs by this kind of verb.

11 See Müller (2005) for a similar explanation for contrasts like *Robin ran her Nikes threadbare/*purple*, and other partial productivity phenomena.

(18) THE CAUSED-MOTION LINKING CONSTRUCTION

$$\begin{bmatrix} \textit{word} & \\ \text{SYN} & \begin{bmatrix} \textit{syn} & \\ \text{CAT} & \begin{bmatrix} \textit{verb} \\ \text{INV} \quad - \end{bmatrix} \end{bmatrix} \\ \text{FRAMES} & \left\langle \begin{bmatrix} \textit{action-result-fr} & \\ \text{INDEX} & e \\ \text{ACTOR} & x \\ \text{THEME} & y \\ \text{STATE} & s \\ \text{TIME} & t \\ \text{LOCATION} & k \\ \vdots & \\ \text{REASON} & z \end{bmatrix} \ldots \right\rangle \\ \text{ARG-ST} & \langle NP_x, NP_y, XP_s, XP_t, XP_k, \ldots, XP_z \rangle \end{bmatrix}$$

Any representation that is nominal and has an empty valence list is abbreviated as 'NP' as shown in Figure 6. In other words, noun phrases are signs of with 'noun' part of speech, no valents, and determiner frames (i. e. a semantics where the index is bound to a overt or covert determiner). See below for more discussion about nominal semantics. The abbreviation 'XP' is similar except that no constraints are imposed on the part-of-speech or on the index type of that linguistic entity. The symbols 'PP', 'VP', 'S', etc. are similarly nothing but abbreviations for AVMs with certain part-of-speech and VAL(ENCE) specifications.

By unifying (17) with (18) we obtain the word in Figure 7. Such a combination is possible because there is no conflicting attribute-value information between the two constructions (i. e. the AVMs are *unifiable*).

Another linking construction analogous to (18) specifies instead that the frame is of sub-type *monadic-pred-fr* and that ARG-ST is $\langle NP_x, XP_t, XP_k, \ldots, XP_z \rangle$, giving us the canonical use of the verb, e. g. *Sam laughed*, and so on for many other argument structure constructions compatible with the constraints specified by (17). Some constructions will impose very specific constraints on some of their arguments, such as the 'way' constructions in (2i,j), or the 'off' constructions in (2m,n), for example. Other

$$NP_x \quad \text{is shorthand for} \quad \begin{bmatrix} \text{SYN} & \begin{bmatrix} \textit{syn} & \\ \text{CAT} & \begin{bmatrix} \textit{noun} \end{bmatrix} \\ \text{VAL} & \langle\,\rangle \end{bmatrix} \\ \text{FRAMES} & \left\langle \begin{bmatrix} \textit{determined-fr} \\ \text{INDEX} \quad x \end{bmatrix} \ldots \right\rangle \end{bmatrix}$$

Figure 6: The representation of phrasal categories.

$$
\begin{bmatrix}
\textit{word} & \\
\text{FORM} & \begin{bmatrix} \textit{w-phon} & \\ \text{STEM} & \langle \text{l,æ,f} \rangle \end{bmatrix} \\
\text{SYN} & \begin{bmatrix} \textit{syn} & \\ \text{CAT} & \begin{bmatrix} \textit{verb} & \\ \text{INV} & - \end{bmatrix} \end{bmatrix} \\
\text{FRAMES} & \left\langle \begin{bmatrix} \textit{action-result-fr} & \\ \text{INDEX} & e \\ \text{PRED} & \textit{laugh} \\ \text{INDEX} & e \\ \text{ACTOR} & x \\ \text{THEME} & y \\ \text{STATE} & s \\ \text{TIME} & t \\ \text{LOCATION} & k \\ \vdots & \\ \text{REASON} & z \end{bmatrix} \ldots \right\rangle \\
\text{ARG-ST} & \left\langle \text{NP}_x, \text{NP}_y, \text{XP}_s, \text{XP}_t, \text{XP}_k, \ldots, \text{XP}_z \right\rangle \\
\text{DTRS} & \langle\,\rangle
\end{bmatrix}
$$

Figure 7: Unification of the constructions in (17) and (18).

verbs are compatible with a different (though often overlapping) range of frames than that of sound emission verbs, and therefore will also partially overlap in the argument structure realizations they can have. Verbs with similar meaning will tend to be compatible with similar argument structure patterns, though there is always the possibility for particular verbs to lexically introduce additional constraints on argument structure so that only a more limited range of uses is possible.

Every sign (lexical or otherwise) is licensed if it satisfies all the constraints imposed by its constructional class, and if it satisfies all the constraints imposed by whatever combinatoric constructions it is a part of. In the former case we have several orthogonal dimensions that simultaneously constrain the sign and further instantiate it (constructional class constructions), in the latter case the sign functions as a daughter of a construction (combinatoric constructions). Both kinds of constraint are illustrated below.

6.1 Constructional Class Constructions

Constructional class constructions are organized into a network or cluster of constructions, each of which characterizes a different dimension of linguistic information.

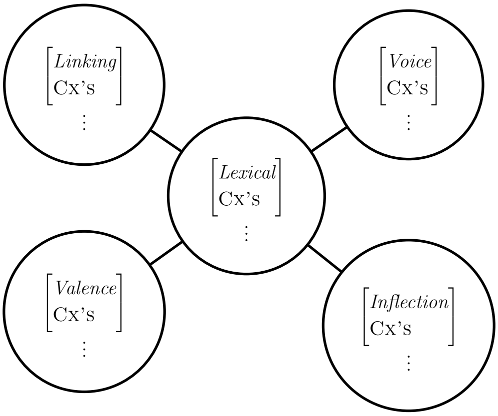


Figure 8: Constructional verb clusters.

Such constructions can pertain to lexical signs or phrasal signs. Consider for example the cluster of verbal lexical constructions illustrated in Figure 8. At the center we have a set of verb forms, i. e. a large number of constructions like (17), and connected to these verb forms we have verbal templates that instantiate different information in different ways, imposing constraints on voice, inflection, argument structure, and valence. Thus, in the linking construction cluster we have a large number of constructions like (18), and so on.[12]

As already discussed in § 2, such clusters arise once learners realize that different uses of the verbal meaning are 'allolexemes', i. e. different combinations of the same core verbal form with various constructions that instantiate different dimensions of linguistic information. Learners begin by memorizing particular uses of the same verb, but given enough experience, arrive at underspecified versions of such verbs by factoring out regular patterns concerning linking, voice, valence, and inflection information. The end result are clusters of different verbs that behave similarly, connected to constructions that constraint their use in various ways. In production, enough constructions must be factored in so that the constraints on the form of the sign is sufficiently instantiated given the constraints on the meaning and function, and in comprehension, enough constructions must be factored in so that semantic information is sufficiently instantiated given the form information. In some cases, the unification of some of these constructions is so frequent that the fully resolved form becomes part of the grammar as well, and can be accessed directly. Through anal-

12 In HPSG and SBCG, the range of possible constructional combinations is encoded as multi-inheritance hierarchies, and as Koenig (1999) shows, such inheritance hierarchies can be computed on-the-fly rather than be listed explicitly. In the approach presented here, closer in spirit to CxG, cluster networks such as the one in Figure 8 are simply rules over word classes, where any given verb use is a conjunction of constructions from each of the verbal construction classes. In order to use a verb in a particular way, speakers must select one construction from each of these clusters and unify them into a single word.

ogy, constructional clusters and networks are expanded, and in the absence of certain uses, speakers counterfactually assume that certain verbs are incompatible with certain constructions.

An alternative way to conceptualize the cluster network in Figure 8 is as a rule that defines any given verb use as the unification of a core (underspecified) verbal lexical entry with one linking construction, one valence construction, one inflection construction, and so on, for all of the different kinds of construction that restrict the space of possible verb uses. Beyond this, the grammar has nothing to say about how speakers effectively choose to combine two given constructions. It is up to the production module to select the appropriate constructions based on the semantic frame *F* that the speaker wishes to covey, also taking into consideration contextual, visual and gestural information, as well as social knowledge, style, and genre. In other words, the particular linking, valence, voice, and inflection constructions that the verb combines with at any point are selected given their compatibility with the frame *F* in question, and their likelihood given the utterance context. Similarly, the comprehension module selects the constructions that are most likely and consistent with the input.

Let us take a closer look at the construction classes in Figure 8. Following Koenig (1999, Ch. 3), Goldberg (2006, 5), Jackendoff and Audring (2014), and others, inflectional phenomena are modeled by constructions such as (19), which instantiate the affixal phonology, set the value of VFORM accordingly, and add the necessary tense information to FRAMES. Of course, further constraints should be added to (19) so that only certain kinds of stem are appropriate for this construction. Irregular inflection is handled by other constructions which in turn select different kinds of stem.

(19) THE REGULAR PAST TENSE INFLECTION CONSTRUCTION

$$\begin{bmatrix} \textit{word} & \\ \text{FORM} & \begin{bmatrix} \textit{w-phon} & \\ \textsc{prefix} & \langle\,\rangle \\ \textsc{suffix} & \langle \text{d} \rangle \end{bmatrix} \\ \text{SYN} & \begin{bmatrix} \textit{syn} & \\ \textsc{cat} & \begin{bmatrix} \textit{verb} & \\ \textsc{vform} & \textit{finite} \end{bmatrix} \end{bmatrix} \\ \text{FRAMES} & \left\langle \begin{bmatrix} \textit{action-pred-fr} & \\ \textsc{arg} & e \end{bmatrix}, \begin{bmatrix} \textit{tense-fr} & \\ \textsc{pred} & \textit{past} \\ \textsc{arg} & e \end{bmatrix} \right\rangle \end{bmatrix}$$

Thus, some constructions introduce information about the stem, others about the affixes. It is up to the construction in (20) to determine how the PHON value of words in general is computed. Capitalized letters in italics are variables over attribute values. Thus, the prefix phonological information in P_1 is concatenated (via the list append relation '$\oplus$') with that in STEM, P_2 and the result is concatenated with that of SUFFIX P_3.

(20) THE LEXICAL PHONOLOGY CONSTRUCTION

$$\begin{bmatrix} \textit{word} & \\ \text{FORM} & \begin{bmatrix} \textsc{phon} & P_1 \oplus P_2 \oplus P_3 \\ \textsc{prefix} & P_1 \\ \textsc{stem} & P_2 \\ \textsc{suffix} & P_3 \end{bmatrix} \end{bmatrix}$$

Unifying the AVM in Figure 7 with (19) and (20) results in the past tense verb in Figure 9. As in Sag (2012), the VFORM attribute has as its value the type *vform* which has two sub-types: *finite*, *infinitive*, *base*, *present-participle*, *past-participle*, and *pass-participle*.

The word in Figure 9 is not quite sufficiently instantiated, however. For example, there is no information about the grammatical function of the arguments, or whether they are locally realized or not. Two kinds of construction from the network in Figure 8 are responsible for constraining valent realization, namely, valence constructions and

$$\begin{bmatrix} \textit{word} & \\ \text{FORM} & \begin{bmatrix} \textit{w-phon} & \\ \textsc{phon} & \langle \text{l,æ,f,d} \rangle \\ \textsc{pref} & \langle\,\rangle \\ \textsc{stem} & \langle \text{l,æ,f} \rangle \\ \textsc{suffix} & \langle \text{d} \rangle \end{bmatrix} \\ \text{SYN} & \begin{bmatrix} \textit{syn} & \\ \textsc{cat} & \begin{bmatrix} \textit{verb} & \\ \textsc{form} & \textit{finite} \\ \textsc{inv} & - \end{bmatrix} \end{bmatrix} \\ \text{FRAMES} & \left\langle \begin{bmatrix} \textit{action-result-fr} & \\ \textsc{index} & e \\ \textsc{pred} & \textit{laugh} \\ \textsc{actor} & x \\ \textsc{theme} & y \\ \textsc{state} & s \\ \textsc{time} & t \\ \textsc{location} & k \\ \vdots & \\ \textsc{reason} & z \end{bmatrix}, \begin{bmatrix} \textit{tense-fr} & \\ \textsc{pred} & \textit{past} \\ \textsc{arg} & e \end{bmatrix} \right\rangle \\ \text{ARG-ST} & \left\langle \text{NP}_x, \text{NP}_y, \text{XP}_s, \text{XP}_t, \text{XP}_k, \ldots, \text{XP}_z \right\rangle \end{bmatrix}$$

Figure 9: Past-tense inflected verb *laughed* with resultative use.

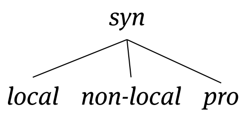


Figure 10: Type hierarchy of *syn* types.

voice constructions. For example, the valence construction in (21) requires that the list of arguments of a word be (non-deterministically) mapped into three (potentially empty) sub-lists: a sublist X corresponds to locally realized valents, a sublist Y corresponds to non-locally realized valents (e. g. clefted, topicalized, extraposed, etc.) and a third list corresponds to unrealized valents (e. g. null complements, passive subjects, etc.).[13]

(21) THE VALENCE CONSTRUCTION

$$\begin{bmatrix} \textit{word} & \\ \text{SYN} & \begin{bmatrix} \textit{syn} & \\ \text{VAL} & X_{list(local)} \\ \text{GAP} & Y_{list(non-local)} \end{bmatrix} \\ \text{ARG-ST} & X \bigcirc Y \bigcirc Z_{list(pro)} \end{bmatrix}$$

Based on Sag (2012, 98), the type *syn* is assumed to have three sub-types: *local* (which indicates that the sign is locally realized), *non-local* (which indicates that it is non-locally realized), and *pro* (which indicates that the sign is not realized). The type hierarchy of *syn* types is given in Figure 10. The construction in (21) requires the members of the VAL list to be typed as *local*, GAP members are *non-local*, and the members of Z must be typed *pro*.

Depending on how the '$\bigcirc$' constraints in (21) are resolved, a wide range of lexemes will be licensed, each with a different constellation of local, non-local, and unrealized valents. Semantic and pragmatic constraints should be added to (21) so that the realization of dependents is constrained, like those in Fillmore and Kay (1996), Goldberg (1995), Goldberg (2006, Ch. 9), Lee-Goldman (2011) and Ruppenhofer and Michaelis (2014). For example, suppose that the verb *laughed* in Figure 9 is unified with (21) so that: (i) X is resolved as a sublist containing NP_x and PP_s, (ii) Y is resolved as a singleton list containing NP_y, and (iii) Z is resolved as containing the remainder. The result is in Figure 11.

This use of *laugh* corresponds to one where the object is not *in situ*, as in *It was Sam* [*who*]$_i$ *the crowd laughed* $\__i$ *out of the room* or [*Who*]$_i$ *Kim laughed* $\__i$ *off the stage were*

13 The 'shuffle' relation '$\bigcirc$' (Reape, 1994) is a non-deterministic version of list concatenation, defined in terms of '$\oplus$'. For example, the unification $\langle \text{NP}, \text{PP} \rangle = A \bigcirc B$ has a total of four possible solutions: $A = \langle \text{NP}, \text{PP} \rangle$ & $B = \langle \rangle$, $A = \langle \text{NP} \rangle$ & $B = \langle \text{PP} \rangle$, $A = \langle \text{PP} \rangle$ & $B = \langle \text{NP} \rangle$, and $A = \langle \, \rangle$ & $B = \langle \text{NP}, \text{PP} \rangle$.

$$
\begin{bmatrix}
\textit{word} & \\
\text{FORM} & \begin{bmatrix} \textit{w-phon} & \\ \textsc{phon} & \langle \text{l,æ,f,d} \rangle \\ \textsc{stem} & \langle \text{l,æ,f} \rangle \\ \textsc{suffix} & \langle \text{d} \rangle \end{bmatrix} \\
\text{SYN} & \begin{bmatrix} \textit{syn} & \\ \textsc{cat} & \begin{bmatrix} \textit{verb} & \\ \textsc{vform} & \textit{finite} \\ \textsc{inv} & - \end{bmatrix} \\ \textsc{val} & \left\langle \text{NP}_x, \text{PP}_s \right\rangle \\ \textsc{gap} & \left\langle \text{NP}_y \right\rangle \end{bmatrix} \\
\text{FRAMES} & \left\langle \begin{bmatrix} \textit{action-result-fr} & \\ \textsc{index} & e \\ \textsc{pred} & \textit{laugh} \\ \textsc{actor} & x \\ \textsc{theme} & y \\ \textsc{state} & s \\ \textsc{time} & t \\ \textsc{location} & k \\ \vdots & \\ \textsc{reason} & z \end{bmatrix}, \begin{bmatrix} \textit{tense-fr} & \\ \textsc{pred} & \textit{past} \\ \textsc{arg} & e \end{bmatrix} \right\rangle \\
\text{ARG-ST} & \left\langle \text{NP}_x, \text{NP}_y, \text{XP}_s, \text{XP}_t, \text{XP}_k, \ldots, \text{XP}_z \right\rangle \\
\text{DTRS} & \langle\, \rangle
\end{bmatrix}
$$

Figure 11: A possible unification between (21) and the AVM in Figure 9.

the kids.[14] As detailed in Ginzburg and Sag (2000), Sag (2010), and Sag (2012), any elements in GAP are constructionally shared across mother-daughter configurations, so that they 'percolate' in syntactic structure until they can be unified with a fronted element as illustrated in Figure 12. As we shall see below, phrasal constructions require the GAP values of the mother and daughters to combine via unification, which in effect allows the information about a missing argument to be propagated on the tree, much the same way that information about the phonology, semantics, valence, or part of speech of a phrase propagates in the tree via unification.

14 If instead the subject phrase is in GAP and complements in VAL then one obtains subject extraction patterns like [*Who*]$_i$ *I think* __$_i$ *laughed the kids off the stage was Kim*, and similarly, if both the object and the subject are in GAP then we license extraction patterns like [*A comedian* THAT *experienced*]$_i$, *I doubt* [*even the worst hecklers*]$_j$ *would easily* __$_j$ *laugh* __$_i$ *off the stage*, and so on. See below for more details.

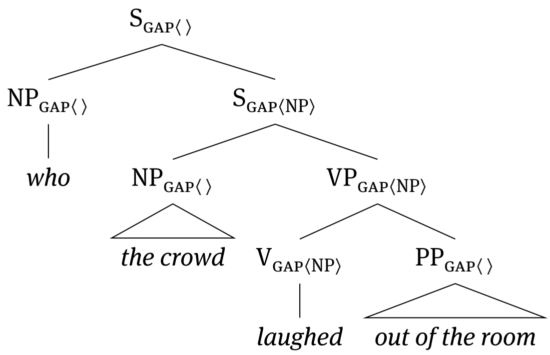


Figure 12: Propagation of GAP information (AVMs abbreviated).

For most words, the type of the ARG-ST members is underspecified as *syn*, so that they can be locally realized, *ex situ* or elided. For other words, however, certain arguments in ARG-ST are more restrictively typed, by stipulation of the lexical entry itself, or by stipulation of some of the constructions that they can be unified with. For example, the lexical entry for the verb *rumor* is necessary passive, whereas only certain uses of *assure* require a dependent to be *ex situ*.

The valence construction in (21) also entails that only arguments can be extracted, not modifiers. And since we have taken a broad view of what counts as a valent, following Fillmore (1982) and Croft (2001), we account for extractions like those in (22) and the impossibility of extracting modifiers like (23).[15]

(22) a. It was [yesterday/mistakenly] that I think Sam skipped school _ .
b. [How often] did you say that Robin was late this week _ ?
c. Was it [by accident] that the driver didn't run over the squirrel _ ?
d. [For what reason] do you believe that Sam resigned _ ?

(23) a. *It was [definitely/never] that I think Sam _ skipped school.
b. *[How happy] did you say that Robin has a _ dog?
c. *It was [by them] that Sam was rumored to be rich _ .
d. *Was it [this] that the driver almost ran over _ squirrel?

Let us now consider voice constructions. Drawing from Fillmore and Kay (1996), Koenig (1999, Ch. 3), Goldberg (1995, 57), Croft (2001, 216), and Davis (2001, Ch. 6)

15 Just like some nominal, verbal, adjectival and prepositional phrases lead a double life as modifiers and as arguments, it is however possible that some temporal, locative, path, manner, and reason phrases also lead a double life, functioning either as modifiers or as arguments. There is a construction that allows sequences of adverbials to characterize the trajectory along a path, or to narrow down a spatial or temporal location, shown in (i) and (ii). Such complex adverbial sequences form a complex constituent, as seen below.

i. Was it [in 1945, on the 16th of July, at 5:39 a.m.] that the first nuclear bomb was detonated _ ?
ii. Q: When was the first nuclear bomb detonated _ ? A: [In 1945, on the 16th of July, in the morning].

among others, constructions like the passive in (24) state that the verb must be inflected appropriately and that the type of *syn* of the first element in ARG-ST must be resolved as *pro*. The latter effectively prevents the argument from being realized, since (21) does not allow *pro*'s to reside in VAL or GAP.[16]

(24) THE PASSIVE VOICE CONSTRUCTION (regular case)

$$\begin{bmatrix} \textit{word} & \\ \text{FORM} & \begin{bmatrix}\text{SUFFIX} & \langle \textit{d} \rangle\end{bmatrix} \\ \text{SYN} & \begin{bmatrix} \textit{syn} & \\ \text{CAT} & \begin{bmatrix} \textit{verb} & \\ \text{VFORM} & \textit{pass-participle}\end{bmatrix}\end{bmatrix} \\ \text{ARG-ST} & \left\langle \begin{bmatrix}\text{SYN} & [\textit{pro}]\end{bmatrix} \ldots \right\rangle \end{bmatrix}$$

Following Koenig and Davis (2003), the passive by-PP phrase is a VP modifier that binds its index to the actor role of the verb heading the modified VP. The passive by-PP phrase can access the correct role by inspecting the first frame listed in the VP's FRAMES. The fact that the passive by-PP phrase is a VP modifier predicts that it can in principle be coordinated with other kinds of modifiers, as shown in (25).

(25) Every chapter of his book was written (both) [[by students] and [for students]].

In contrast, the active construction in (26) remains neutral about inflection, requires that the first element in ARG-ST of the verb is not *pro*, and that its index is the same as the highest ranking thematic role in the verb's semantic frame, as depicted in (26). Additional constraints can be imposed on the verbal frame, and more specifically on the value of PRED, so that the range of passivizable verbs can be semantically restricted as appropriate.

(26) THE ACTIVE VOICE CONSTRUCTION

$$\begin{bmatrix} \textit{word} & \\ \text{SYN} & \begin{bmatrix} \textit{syn} & \\ \text{CAT} & [\textit{verb}]\end{bmatrix} \\ \text{FRAMES} & \left\langle \begin{bmatrix} \textit{action-pred-fr} & \\ \text{ACTOR} & x \end{bmatrix} \ldots \right\rangle \\ \text{ARG-ST} & \left\langle \begin{bmatrix} \text{SYN} & [\neg \textit{pro}] \\ \text{SEM} & \begin{bmatrix} \textit{sem} & \\ \text{INDEX} & x \end{bmatrix} \end{bmatrix} \ldots \right\rangle \end{bmatrix}$$

16 As in the case of other inflectional constructions, this passive construction is simplified here given that constraints on the stem must be added, so that the *-d* suffix is added to only with certain stems. Other passive rules impose different constraints on the stem, and introduce a different suffix, e. g. *-en* as in *He was seen yesterday*.

Another type of lexical construction that further restricts morphosyntax concerns case assignment. Such a construction requires that the first non-*pro* NP in the ARG-ST of a finite verb must have its value of CASE resolved as *nom*, and that CASE value of any other NP must be instantiated as *acc*. For heads that are not finite verbs, case can be assigned lexically. For example, prepositions require accusative objects, and gerunds non-nominative subjects.

All of the constructional class constructions discussed above are lexical in nature, but constructional class constructions can also be phrasal in the sense that they further constrain signs that have daughters. For example, such constructions may specify how phonological phrases combine to form intonational phrases, or how phrasal semantic composition is to proceed. In what follows I focus on constructions (lexical or otherwise) that have daughters.

6.2 Combinatoric constructions and sample analysis

Whereas class constructions serve to make underspecified signs more instantiated, combinatoric constructions take one of more instantiated signs and license a different sign (though when there is only one daughter, the name 'combinatoric' is somewhat of a misnomer). For example, compounding constructions are binary combinatoric constructions, as discussed in § 3.1, but derivational morphology constructions are combinatorical unary constructions, following Koenig (1999), Booij (2010), and Sag (2012). The latter is illustrated by the *-er* construction in (27), which is essentially a PSG rule of the form 'N → V'. The notation $Y{:}\alpha$ means that the value of the variable Y has at least the information in the AVM α. Thus the *action-pred-fr* semantics Y of the verbal daughter is also part of the nominal mother.

(27) THE '-ER' NOMINALIZATION CONSTRUCTION

$$
\begin{bmatrix}
\textit{word} & \\
\text{FORM} & \begin{bmatrix} \textit{w-phon} & \\ \text{STEM} & X \end{bmatrix} \\
\text{SYN} & \begin{bmatrix} \textit{syn} & \\ \text{CAT} & \begin{bmatrix} \textit{noun} \end{bmatrix} \end{bmatrix} \\
\text{FRAMES} & \left\langle \begin{bmatrix} \textit{common-n-fr} \\ \text{INDEX } x \end{bmatrix}, Y \ldots \right\rangle
\end{bmatrix}
\rightarrow
\begin{bmatrix}
\textit{word} & \\
\text{FORM} & \begin{bmatrix} \textit{w-phon} & \\ \text{PHON} & X \\ \text{PREFIX} & \langle\,\rangle \\ \text{SUFF} & \langle \text{ɹ} \rangle \end{bmatrix} \\
\text{SYN} & \begin{bmatrix} \textit{syn} & \\ \text{CAT} & \begin{bmatrix} \textit{verb} & \\ \text{VFORM} & \textit{base} \\ \text{INV} & - \end{bmatrix} \end{bmatrix} \\
\text{FRAMES} & \left\langle Y{:}\begin{bmatrix} \textit{action-pred-fr} \\ \text{ACTOR } x \end{bmatrix} \right\rangle
\end{bmatrix}
$$

Note that the VFORM value of the verbal daughter is *base* (preventing it from having been inflected), and the mother node is required to be a case-underspecified common noun co-indexed with the actor of the verb's frame Y. Other examples of unary-branching constructions include those of the form 'NP → N', which license determinerless NPs by directly adding the correct determiner to the top of the frames list of the nominal.[17] See Fillmore and Baker (2009) for more detailed account of a broad range of such bare NP constructions. According to Chaves (2014), another kind of unary branching construction is the one responsible for so-called Right Node Raising (RNR) structures. As illustrated below, RNR can apply to a wide range of constructions other than coordination.

(28) a. Explain how signals move from a PRE- to a POST-[synaptic neuron].
 b. Are you talking about A NEW or about AN EX-[boyfriend]?
 c. Robin does NOT PLAY – or PRETENDS not to play – [with a full deck].
 d. This is the difference between AN INTERESTING and A BORING [book].
 e. I said that John – and you said that Mary – [were wonderful students].

In order to illustrate how branching phrases are obtained, we turn to the example sentence in (29) and show how it is decomposed, piecemeal.

(29) After Mary introduced herself to the audience, she turned to a man that she had met before.

I will start by focusing on the sentence *Mary introduced herself to the audience*, and in particular, on its sub-constituents. I assume that pronouns and most English proper names already have a determiner frame in their semantics, which in effect means that such expressions are ready to function as NPs. For illustration, consider the determiner *the* in (30). From now on I omit the affixal and stem attributes, for convenience. The type *the-fr* (specific to the word 'the') is a subtype of *definite-fr* (the type for all definite nominals), which in turn is a sub-type of *determined-fr* (the type for all nominal phrases).

(30) a. THE 'THE' LEXICAL CONSTRUCTION

$$\begin{bmatrix} \textit{word} & \\ \text{FORM} & \begin{bmatrix} \textit{w-phon} & \\ \text{PHON} & \langle \eth,\text{ʌ} \rangle \end{bmatrix} \\ \text{SYN} & \begin{bmatrix} \textit{syn} & \\ \text{CAT} & \textit{det} \\ \text{VAL} & \langle\rangle \\ \text{GAP} & \langle\rangle \end{bmatrix} \\ \text{FRAMES} & \left\langle \begin{bmatrix} \textit{definite-fr} & \\ \text{INDEX} & x \end{bmatrix} \right\rangle \end{bmatrix}$$

17 English bare NP uses are not restricted to plurals and mass nominals, see for example *Mother told me to go home* and *This can help baby sit up independently*.

b. The ‘audience’ Lexical Construction

$$\begin{bmatrix} \textit{word} & \\ \text{FORM} & \begin{bmatrix} \textit{w-phon} & \\ \textsc{phon} & \langle \text{ɔ,d,i,j,ə,n,t,s} \rangle \end{bmatrix} \\ \text{SYN} & \begin{bmatrix} \textit{syn} & \\ \textsc{cat} & \begin{bmatrix} \textit{noun} \end{bmatrix} \\ \textsc{val} & \langle\rangle \\ \textsc{gap} & \langle\rangle \end{bmatrix} \\ \text{FRAMES} & \left\langle \begin{bmatrix} \textit{person-fr} & \\ \textsc{index} & x \begin{bmatrix} \textit{agr} & \\ \textsc{num} & \textit{plur} \\ \textsc{per} & \textit{3rd} \end{bmatrix} \\ \textsc{pred} & \textit{audience} \end{bmatrix} \right\rangle \end{bmatrix}$$

The ‘NP → DP N’ construction in (31) allows determiners to combine with common noun nominal heads, based on Fillmore and Kay (1996) and Bergen and Chang (2005). Either daughter can be of type *word* or *phrase*, and therefore can be lexical or phrasal. The x indices of the two daughters are unified, thus binding the quantified variable to the variable introduced by the nominal. Since agreement information is recorded at the index level, the unification of the two indices causes them to agree in number, gender and person. The index of the determiner in (30a) is underspecified for agreement and therefore it can combine with any nominal.

(31) The Determination Construction

$$\begin{bmatrix} \textit{phrase} & \\ \text{FORM} & \begin{bmatrix} \textit{phr-phon} & \\ \textsc{phon} & P_1 \oplus P_2 \end{bmatrix} \\ \text{SYN} & X \\ \text{FRAMES} & \left\langle Y{:}\begin{bmatrix} \textsc{index} & x \\ \textsc{arg} & Z \end{bmatrix} \right\rangle \end{bmatrix} \rightarrow \begin{bmatrix} \text{FORM} & \begin{bmatrix} \textsc{phon}\ P_1 \end{bmatrix} \\ \text{SYN} & \begin{bmatrix} \textit{syn} \\ \textsc{cat}\ \textit{det} \end{bmatrix} \\ \text{FRAMES} & \langle Y \rangle \end{bmatrix} \begin{bmatrix} \text{FORM} & \begin{bmatrix} \textsc{phon}\ P_2 \end{bmatrix} \\ \text{SYN} & X{:}\begin{bmatrix} \textit{syn} & \\ \textsc{cat} & \textit{noun} \\ \textsc{val} & \langle\,\rangle \end{bmatrix} \\ \text{FRAMES} & Z{:}\left\langle \begin{bmatrix} \textit{common-n-fr} & \\ \textsc{index} & x \end{bmatrix}, \ldots \right\rangle \end{bmatrix}$$

Note also that (31) requires all SYN information of the nominal daughter to be unified via the SYN information of the mother node, via the variable X.[18] The result of combining (30a) with (30b) via (31) is seen in Figure 13.

The obtained semantic representation is interpreted as ι_x *audience*(x). Because the case is underspecified, this NP can function as a subject or as a complement. Determiners cannot combine with NPs because the latter have *determiner-fr* at the top of

18 Thus if the noun is specified as [GAP ⟨PP⟩], for example, then so is the NP mother node, and vice-versa, which is necessary to license extraction patterns like *It was* [*from Sue*] *that I got* [*a letter*__].

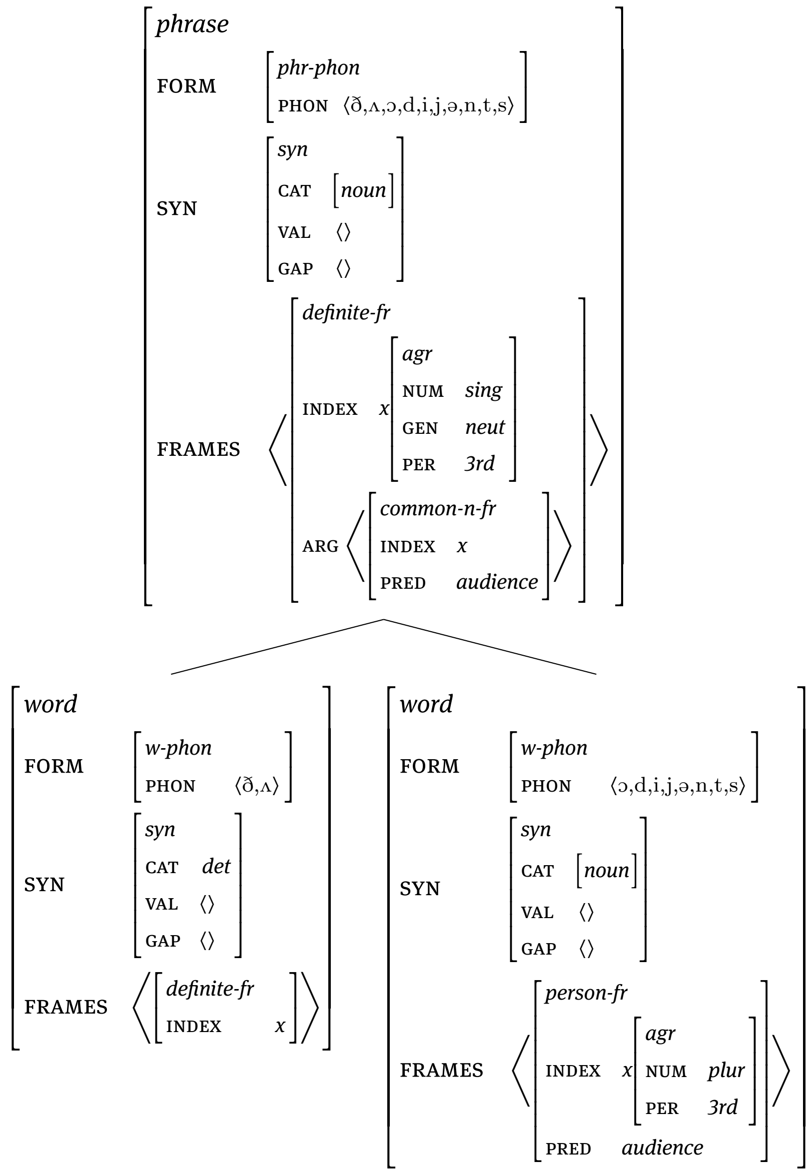


Figure 13: The representation of *the audience*.

their FRAMES list. Conversely, pre-determiner expressions are only allowed to combine with expressions with a *determiner-fr* already at the top of FRAMES.

Some words can combine with NPs without adding much meaning to them at all. This is the case of argument-marking prepositions like (32), which merely impose a thematic role on their complements (Wechsler, 1995; Davis, 2001), and have a singleton ARG-ST list $\langle \mathrm{NP}_z \rangle$. The use of *to* shown in (32) arises when the valence construction in (21) determines that the NP must reside in VAL, rather than resolving it as a null (*pro*) complement (a null complement use), or as a member of GAP (a preposition stranding use).

(32)
$$\begin{bmatrix} \textit{word} & \\ \text{FORM} & \begin{bmatrix} \textit{w-phon} \\ \text{PHON}\ \langle \text{t,u} \rangle \end{bmatrix} \\ \text{SYN} & \begin{bmatrix} \textit{syn} & \\ \text{CAT} & \textit{prep} \\ \text{VAL} & \left\langle X{:}\mathrm{NP}_z\begin{bmatrix}\text{CASE}\ \textit{acc}\end{bmatrix} \right\rangle \\ \text{GAP} & \langle\rangle \end{bmatrix} \\ \text{FRAMES} & \left\langle \begin{bmatrix} \textit{recipient-fr} & \\ \text{INDEX} & z \end{bmatrix} \right\rangle \\ \text{ARG-ST} & \langle X \rangle \end{bmatrix}$$

In general, prepositions combine with their complements via the ‘PP → P XP’ construction seen in (33).

(33) THE PREPOSITIONAL-COMPLEMENT CONSTRUCTION

$$\begin{bmatrix} \textit{phrase} & \\ \text{FORM} & \begin{bmatrix} \textit{phr-phon} \\ \text{PHON}\ P_1 \oplus P_2 \end{bmatrix} \\ \text{SYN} & \begin{bmatrix} \text{CAT} & \textit{prep} \\ \text{VAL} & \langle\ \rangle \\ \text{GAP} & G \end{bmatrix} \\ \text{FRAMES} & F_1 \oplus F_2 \end{bmatrix} \rightarrow \begin{bmatrix} \textit{word} & \\ \text{FORM} & \begin{bmatrix}\text{PHON} & P_1\end{bmatrix} \\ \text{SYN} & \begin{bmatrix} \text{CAT} & \textit{prep} \\ \text{VAL} & \langle X \rangle \end{bmatrix} \\ \text{FRAMES} & F_1 \end{bmatrix} X{:}\begin{bmatrix} \text{FORM} & \begin{bmatrix}\text{PHON} & P_2\end{bmatrix} \\ \text{SYN} & \begin{bmatrix}\text{GAP}\ G\end{bmatrix} \\ \text{FRAMES} & F_2 \end{bmatrix}$$

Thus, combining (32) and the AVM in Figure 13 via (33) yields the AVM in Figure 14. In the constructions discussed so far in the present grammar fragment, the mother’s semantic representation is simply the concatenation of the semantic representations of the daughters, but that need not be always the case, of course. As already discussed in § 3, some constructions make a semantic contribution over and above the contributions of the daughters. Thus, semantic composition is construction-specific, and can range from the completely transparent to the highly irregular.

Let us move on to the verb *introduce*, shown in (34). As in the case of *laugh*, this verb is compatible with many different uses, and therefore its frame and argument structure will be further instantiated by linking constructions, as long as they are mutually compatible.

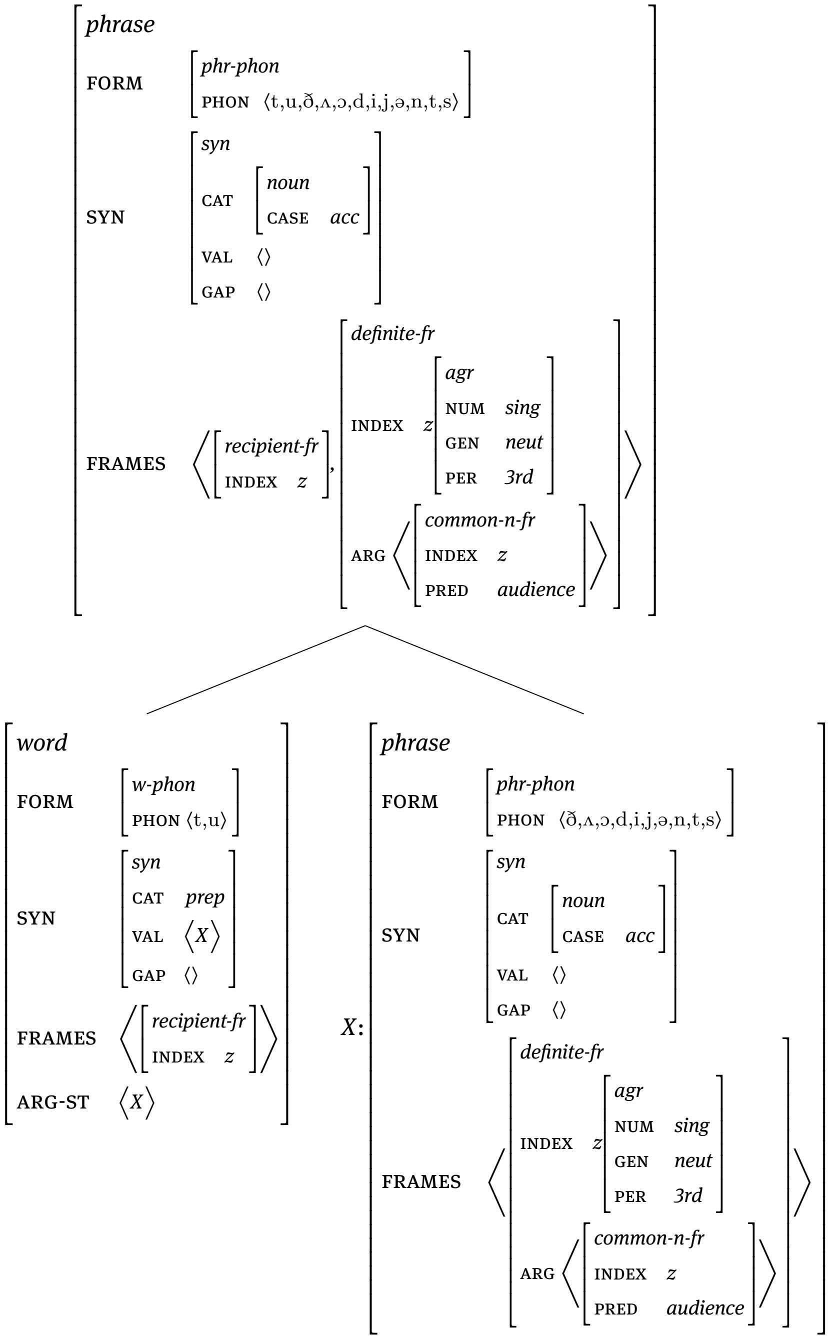


Figure 14: The representation of *to the audience*.

(34) The ‘introduce’ Lexical Construction

$$\begin{bmatrix} \textit{word} & \\ \text{FORM} & \begin{bmatrix} \textit{w-phon} & \\ \text{STEM} & \langle \text{ɪ,n,t,ɹ,o,d,u,s} \rangle \end{bmatrix} \\ \text{SYN} & \begin{bmatrix} \textit{syn} & \\ \text{CAT} & \begin{bmatrix} \textit{verb} & \\ \text{INV} & - \end{bmatrix} \end{bmatrix} \\ \text{FRAMES} & \left\langle \begin{bmatrix} \textit{cause-poss-fr} & \\ \text{PRED} & \textit{introduce} \\ \text{INDEX} & e \\ \text{ACTOR} & x \\ \text{THEME} & y \\ \text{RECIPIENT} & z \\ \text{TIME} & t \\ \text{LOCATION} & k \\ \vdots & \\ \text{REASON} & z \end{bmatrix} \right\rangle \\ \text{DTRS} & \langle\rangle \end{bmatrix}$$

Among the linking constructions that are compatible with the frame type introduced by (34) are those responsible for the dative alternation. Thus, if (34) is combined with (35a) we obtain the use of the verb in which the recipient is an NP, and if combined with (35b) we obtain the use of the verb in which the recipient is an oblique. Which frame is chosen depends on the speaker’s intentions, and which constellation of frames are syntactically, semantically, and pragmatically compatible with the verb.

(35) a. The Ditransitive Argument-Structure Construction

$$\begin{bmatrix} \textit{word} & \\ \text{SYN} & \begin{bmatrix} \textit{syn} & \\ \text{CAT} & \begin{bmatrix} \textit{verb} \end{bmatrix} \end{bmatrix} \\ \text{FRAMES} & \left\langle \begin{bmatrix} \textit{caused-poss-fr} & \\ \text{INDEX} & e \\ \text{ACTOR} & x \\ \text{THEME} & y \\ \text{RECIPIENT} & z \\ \text{TIME} & t \\ \text{LOCATION} & k \\ \vdots & \\ \text{REASON} & z \end{bmatrix} \right\rangle \\ \text{ARG-ST} & \left\langle \text{NP}_x, \text{NP}_z, \text{NP}_y, \text{XP}_t, \text{XP}_k, \ldots, \text{XP}_z \right\rangle \end{bmatrix}$$

b. The Transitive Argument-Structure Construction

$$\begin{bmatrix} \textit{word} & \\ \text{SYN} & \begin{bmatrix} \textit{syn} & \\ \text{CAT} & \begin{bmatrix} \textit{verb} \end{bmatrix} \end{bmatrix} \\ \text{FRAMES} & \left\langle \begin{bmatrix} \textit{caused-poss-fr} & \\ \text{INDEX} & e \\ \text{AGENT} & x \\ \text{THEME} & y \\ \text{RECIPIENT} & z \\ \text{TIME} & t \\ \text{LOCATION} & k \\ \vdots & \\ \text{REASON} & z \end{bmatrix} \right\rangle \\ \text{ARG-ST} & \left\langle \text{NP}_x, \text{NP}_y, \text{PP}_z, \text{XP}_t, \text{XP}_k, \ldots, \text{XP}_z \right\rangle \end{bmatrix}$$

By unifying (34) with the argument-structure construction in (35b), the valence construction, the lexical phonology construction, the past tense inflection construction, the active voice construction, and the case construction, we obtain uses like the one in Figure 15.

Let us assume that the direct object is the reflexive pronoun *herself* in (36). The type *reflx* is a sub-type of *definite-fr*, so this nominal can directly function as an NP.[19] This semantic representation in FRAMES is equivalent to $\iota_x(x = \ldots)$.

(36)
$$\begin{bmatrix} \textit{word} & \\ \text{FORM} & \begin{bmatrix} \textit{w-phon} & \\ \text{PHON} & \langle \text{h,ɹ,s,ɛ,l,f} \rangle \end{bmatrix} \\ \text{SYN} & \begin{bmatrix} \textit{syn} & \\ \text{CAT} & \begin{bmatrix} \textit{noun} & \\ \text{CASE} & \textit{acc} \end{bmatrix} \\ \text{VAL} & \langle\rangle \\ \text{GAP} & \langle\rangle \end{bmatrix} \\ \text{FRAMES} & \left\langle \begin{bmatrix} \textit{reflx-fr} & \\ \text{INDEX} & x \begin{bmatrix} \textit{agr} & \\ \text{NUM} & \textit{sing} \\ \text{GEN} & \textit{fem} \\ \text{PER} & \textit{3rd} \end{bmatrix} \\ \text{ARG} & \langle\rangle \end{bmatrix} \right\rangle \end{bmatrix}$$

19 As in HPSG and SBCG, Binding Theory is stated at the *word* level, as a constraint on ARG-ST. For example, Principle A states that if an ARG-ST list *L* has a non-*pro* member *K* with a reflexive nominal index *x*, then *K* must be preceded in *L* by some other non-*pro* element that is co-indexed with *x*.Recall that *pro* phrases are not allowed to reside in VAL or GAP, and as per the Valence Construction in (21) are suppressed. Hence, Binding Theory ignores such members of ARG-ST, as intended.

Uninverted verbs combine with whatever complements they lexically select in VAL via the 'VP → XP_1 ... XP_n' construction formalized in (37). This construction requires that all subcategorized valents of the first daughter (except the subject) be unified with its sisters X_1 ... X_n. All their phonologies are concatenated, as are their frames and gaps.[20]

(37) THE PREDICATE-COMPLEMENT CONSTRUCTION

$$\begin{bmatrix} \textit{phrase} & \\ \text{FORM} & \begin{bmatrix} \textit{phr-phon} \\ \text{PHON } P_0 \oplus P_1 \oplus \ldots \oplus P_n \end{bmatrix} \\ \text{SYN} & \begin{bmatrix} \textit{syn} & \\ \text{CAT} & K \\ \text{VAL} & \langle X_0 \rangle \\ \text{GAP} & G_0 \cup \ldots \cup G_n \end{bmatrix} \\ \text{FRAMES} & F_0 \oplus \ldots \oplus F_n \end{bmatrix} \rightarrow \begin{bmatrix} \textit{word} & \\ \text{FORM} & \begin{bmatrix} \text{FORM } P_0 \end{bmatrix} \\ \text{SYN} & \begin{bmatrix} \textit{syn} & \\ \text{CAT} & K{:}\begin{bmatrix} \textit{verb} & \\ \text{INV} & - \end{bmatrix} \\ \text{VAL} & \left\langle X_0, X_1, \ldots, X_n \right\rangle \\ \text{GAP} & G_0 \end{bmatrix} \\ \text{FRAMES} & F_0 \end{bmatrix}$$

$$X_1\begin{bmatrix} \text{FORM} & \begin{bmatrix} \text{PHON } P_1 \end{bmatrix} \\ \text{SYN} & \begin{bmatrix} \text{GAP } G_1 \end{bmatrix} \\ \text{FRAMES} & F_1 \end{bmatrix} \ldots\ X_n\begin{bmatrix} \text{FORM} & \begin{bmatrix} \text{PHON } P_n \end{bmatrix} \\ \text{SYN} & \begin{bmatrix} \text{GAP } G_n \end{bmatrix} \\ \text{FRAMES} & F_n \end{bmatrix}$$

If the verb in Figure 15 is unified with the first daughter of (37) then its VAL list will consist of an NP and a recipient PP, which must appear in that order. If the NP is (36) and the PP is the one in Figure 14, we obtain the VP below.

The construction that allows VPs to combine with their subjects is of the form 'S → X VP', as shown in (38). This construction requires that the verbal daughter's unsaturared valent X is unified with the first daughter.[21]

---

20 I assume that the symbol '∪' is a non-deterministic operator that treats lists as if they were sets. Thus, $\langle NP_x \rangle \cup \langle NP_y \rangle$ can be resolved as $\langle NP_x, NP_y \rangle$ or as $\langle NP_z \rangle$ (where $x = y = z$). The former is necessary when there are multiple gaps linked to different fillers as in *Robin is someone who I never know what to say _ to _*, and the latter is necessary when there are multiple gaps linked to the same filler, as in *Robin was the client who we forgot to send pictures of _ to _*.

21 As in many phrasal constructions, the mother's GAP values are a combination of the GAP values of the daughters, allowing gaps to propagate from/to the subject phrase (e. g. [*Which president*] *would* [*the impeachment of _*] *have caused the most outrage?*), the object phrase (e. g. [*Which president*] *would you have welcomed* [*the impeachment of _*]*?*), both subject and objects (e. g. [*Which president*] *would* [*the impeachment of _*] *have surprised _ the most?*), or neither, depending on the GAP values of each daughter.

$$
\begin{bmatrix}
\textit{word} & \\
\text{FORM} & \begin{bmatrix} \textit{w-phon} & \\ \text{PHON} & \langle \text{ɪ,n,t,ɹ,o,d,u,s,d} \rangle \end{bmatrix} \\
\text{SYN} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \textit{syn} & \\ \text{CAT} & \textit{verb} \\ \text{VFORM} & \textit{fin} \\ \text{INV} & - \end{bmatrix} \\ \text{VAL} & \left\langle \text{NP}_x, \text{NP}_y, \text{PP}_z \right\rangle \\ \text{GAP} & \langle \rangle \end{bmatrix} \\
\text{FRAMES} & \left\langle \begin{bmatrix} \textit{caused-poss-fr} & \\ \text{INDEX} & e \\ \text{PRED} & \textit{introduce} \\ \text{AGENT} & x \\ \text{THEME} & y \\ \text{RECIPIENT} & z \\ \text{TIME} & t \\ \text{LOCATION} & k \\ \vdots & \\ \text{REASON} & z \end{bmatrix}, \begin{bmatrix} \textit{tense-fr} & \\ \text{PRED} & \textit{past} \\ \text{ARG} & e \end{bmatrix} \right\rangle \\
\text{DTRS} & \langle \rangle
\end{bmatrix}
$$

Figure 15: Possible unification of (34) with linking, valence, voice, and inflection cxs.

(38) THE SUBJECT-PREDICATE CONSTRUCTION

$$
\begin{bmatrix}
\textit{phrase} & \\
\text{FORM} & \begin{bmatrix} \textit{phr-phon} & \\ \text{PHON} & P_1 \oplus P_2 \end{bmatrix} \\
\text{SYN} & \begin{bmatrix} \text{CAT} & K \\ \text{VAL} & \langle \, \rangle \\ \text{GAP} & G_1 \cup G_2 \end{bmatrix} \\
\text{FRAMES} & F_1 \oplus F_2
\end{bmatrix}
\rightarrow X\!: \begin{bmatrix} \text{FORM} & \begin{bmatrix} \text{PHON } P_1 \end{bmatrix} \\ \text{SYN} & \begin{bmatrix} \text{GAP} & G_1 \end{bmatrix} \\ \text{FRAMES} & F_1 \end{bmatrix}
\begin{bmatrix} \text{FORM} & \begin{bmatrix} \text{PHON } P_2 \end{bmatrix} \\ \text{SYN} & \begin{bmatrix} \text{CAT } K\!: \begin{bmatrix} \text{INV } - \end{bmatrix} \\ \text{VAL } \langle X \rangle \\ \text{GAP } G_2 \end{bmatrix} \\ \text{FRAMES} & F_2 \end{bmatrix}
$$

The fact that the second daughter has the attribute INV means that it must be verbal, as no other part-of-speech bears that attribute. The phrase licensed by combining the lexical entry for *Mary* in (39) with the VP in Figure 16 via (38) is shown in Figure 17.

(39)
$$
\begin{bmatrix}
\textit{word} & \\
\text{FORM} & \begin{bmatrix} \textit{w-phon} \\ \text{PHON}\ \langle \text{m,ɛ,ɹ,ɪ} \rangle \end{bmatrix} \\
\text{SYN} & \begin{bmatrix} \textit{syn} & \\ \text{CAT} & \begin{bmatrix} \textit{noun} \end{bmatrix} \\ \text{VAL} & \langle\rangle \\ \text{GAP} & \langle\rangle \end{bmatrix} \\
\text{FRAMES} & \left\langle \begin{bmatrix} \textit{definite-fr} & \\ \text{INDEX} & x \begin{bmatrix} \textit{agr} & \\ \text{NUM} & \textit{sing} \\ \text{GEN} & \textit{fem} \\ \text{PER} & \textit{3rd} \end{bmatrix} \\ \text{ARG} & \left\langle \begin{bmatrix} \textit{name-fr} & \\ \text{INDEX} & x \\ \text{PRED} & \textit{mary} \end{bmatrix} \right\rangle \end{bmatrix} \right\rangle
\end{bmatrix}
$$

The CxG analysis of raising and control is similar to that of Categorial Grammar, Lexical-Functional Grammar, and Head-driven Phrase Structure Grammar (see this volume). Basically, the auxiliary selects a VP complement and requires that the subject *X* subcategorized by that VP be unified with the subject subcategorized by the auxiliary. As shown in (40), the lexical entry of an auxiliary verb like *have* requires that the *X* subject valent of the VP complement is unified with the valent of the auxiliary.

(40)
$$
\begin{bmatrix}
\textit{word} & \\
\text{FORM} & \begin{bmatrix} \textit{w-phon} \\ \text{PHON}\ \langle \text{h,æ,d} \rangle \end{bmatrix} \\
\text{SYN} & \begin{bmatrix} \textit{syn} & \\ \text{CAT} & \begin{bmatrix} \textit{verb} & \\ \text{VFORM} & \textit{fin} \end{bmatrix} \\ \text{VAL} & \left\langle X, \begin{bmatrix} \text{SYN} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \textit{verb} & \\ \text{VFORM} & \textit{fin} \end{bmatrix} \\ \text{VAL} & \langle X \rangle \end{bmatrix} \\ \text{FRAMES} & \left\langle \begin{bmatrix} \text{INDEX}\ e \end{bmatrix}, \ldots \right\rangle \end{bmatrix} \right\rangle \\ \text{GAP} & \langle\rangle \end{bmatrix} \\
\text{FRAMES} & \left\langle \begin{bmatrix} \textit{tense-fr} & \\ \text{PRED} & \textit{past} \\ \text{ARG} & e \end{bmatrix} \right\rangle
\end{bmatrix}
$$

$$
\begin{bmatrix}
\textit{phrase} & \\
\text{FORM} & \begin{bmatrix} \textit{phr-phon} & \\ \text{PHON} & \langle \text{ɪ,n,t,ɹ,o,d,u,s,d,h,ɹ,s,ɛ,l,f,t,u,ð,ʌ,ɔ,d,i,j,ə,n,t,s} \rangle \end{bmatrix} \\
\text{SYN} & \begin{bmatrix} \textit{syn} & \\ \text{CAT} & \begin{bmatrix} \textit{verb} & \\ \text{INV} & - \\ \text{VFORM} & \textit{fin} \end{bmatrix} \\ \text{VAL} & \langle \text{NP}_x \rangle \\ \text{GAP} & \langle\rangle \end{bmatrix} \\
\text{FRAMES} & \left\langle \begin{bmatrix} \textit{caused-poss-fr} & \\ \text{INDEX} & e \\ \text{PRED} & \textit{introduce} \\ \text{ACTOR} & x \\ \text{THEME} & y \\ \text{RECIPIENT} & z \\ \text{TIME} & t \\ \text{LOCATION} & k \\ \vdots & \\ \text{REASON} & z \end{bmatrix}, \begin{bmatrix} \textit{tense-fr} & \\ \text{PRED} & \textit{past} \\ \text{ARG} & e \end{bmatrix}, \begin{bmatrix} \textit{reflx-fr} & \\ \text{INDEX} & x \begin{bmatrix} \textit{agr} & \\ \text{NUM} & \textit{sing} \\ \text{GEN} & \textit{fem} \\ \text{PER} & \textit{3rd} \end{bmatrix} \\ \text{ARG} & \langle\rangle \end{bmatrix}, \begin{bmatrix} \textit{definite-fr} & \\ \text{INDEX} & y \\ \text{ARG} & \left\langle \begin{bmatrix} \textit{person-fr} & \\ \text{INDEX} & y \begin{bmatrix} \textit{agr} & \\ \text{NUM} & \textit{plur} \\ \text{PER} & \textit{3rd}\end{bmatrix} \\ \text{PRED} & \textit{audience} \end{bmatrix} \right\rangle \end{bmatrix} \right\rangle
\end{bmatrix}
$$

Figure 16: The VP *introduced herself to the audience* (daughter nodes omitted).

Thus, in a sentence like *She had met him* the matrix subject is co-indexed with the subject required by the embedded VP *met him* as depicted in Figure 18 using familiar abbreviations for the respective AVMs, where 'S' and 'VP' stands for any AVM of part-of-speech *verb* with an empty VAL list and with a singleton VAL list, respectively. As before, the subject combines with the matrix verb phrase via the subject-predicate construction in (38), and the auxiliary combines with its VP complement via the predicate-

$$
\begin{bmatrix}
\textit{phrase} & \\
\text{FORM} & \begin{bmatrix} \textit{phr-phon} & \\ \text{PHON} & \langle \text{m,ɛ,ɹ,ɪ,ɪ,n,t,ɹ,o,d,u,s,d,h,ɹ,s,ɛ,l,f,t,u,ð,ʌ,ɔ,d,i,j,ə,n,t,s} \rangle \end{bmatrix} \\
\text{SYN} & \begin{bmatrix} \textit{syn} & \\ \text{CAT} & \begin{bmatrix} \textit{verb} & \\ \text{INV} & - \\ \text{VFORM} & \textit{fin} \end{bmatrix} \\ \text{VAL} & \langle\rangle \\ \text{GAP} & \langle\rangle \end{bmatrix} \\
\text{FRAMES} & \left\langle \begin{array}{l}
\begin{bmatrix} \textit{caused-poss-fr} & \\ \text{INDEX} & e \\ \text{PRED} & \textit{introduce} \\ \text{ACTOR} & x \\ \text{THEME} & y \\ \text{RECIPIENT} & z \\ \text{TIME} & t \\ \text{LOCATION} & k \\ \vdots & \\ \text{REASON} & z \end{bmatrix}, \begin{bmatrix} \textit{tense-fr} & \\ \text{PRED} & \textit{past} \\ \text{ARG} & e \end{bmatrix}, \\
\begin{bmatrix} \textit{reflx-fr} & \\ \text{INDEX} & x\begin{bmatrix} \textit{agr} & \\ \text{NUM} & \textit{sing} \\ \text{GEN} & \textit{fem} \\ \text{PER} & \textit{3rd} \end{bmatrix} \\ \text{ARG} & \langle\rangle \end{bmatrix}, \begin{bmatrix} \textit{definite-fr} & \\ \text{INDEX} & y \\ \text{ARG} & \left\langle \begin{bmatrix} \textit{person-fr} & \\ \text{INDEX} & y\begin{bmatrix} \textit{agr} & \\ \text{NUM} & \textit{sing} \\ \text{GEN} & \textit{mas} \\ \text{PER} & \textit{3rd} \end{bmatrix} \\ \text{PRED} & \textit{audience} \end{bmatrix} \right\rangle \end{bmatrix}
\end{array} \right\rangle
\end{bmatrix}
$$

Figure 17: Representation of *Mary introduced herself to the audience* (daughters omitted).

complement construction in (37). The latter construction is also responsible for combining *met* with *him*.

In a more complex structure like *a man that she had met before* various constructions are at play. The Valence Construction causes the object of *met* to appear in GAP

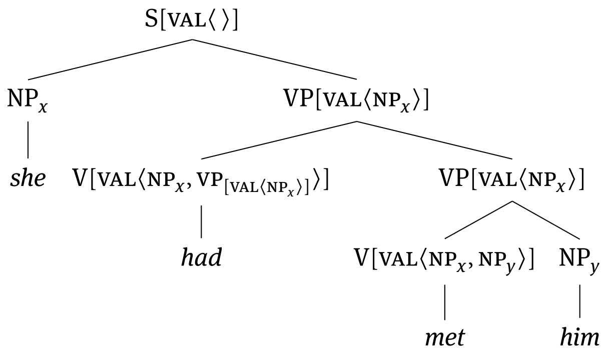


Figure 18: Structure of the S *she had met him* (AVMs abbreviated).

instead of VAL, which in effect prevents it from appearing *in situ*, and the phrasal constructions discussed above force it to be percolated in the tree structure.

(41)
$$\begin{bmatrix} \textit{word} & \\ \text{FORM} & \begin{bmatrix} \textit{w-phon} \\ \text{FORM } \langle \text{m,}\varepsilon\text{,t} \rangle \end{bmatrix} \\ \text{SYN} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \textit{syn} & \\ \text{CAT} & \textit{verb} \\ \text{VFORM} & \textit{fin} \\ \text{INV} & - \end{bmatrix} \\ \text{VAL} & \left\langle \text{NP}_x, \text{PP}_z \right\rangle \\ \text{GAP} & \left\langle \text{NP}_y \right\rangle \end{bmatrix} \\ \text{FRAMES} & \left\langle \begin{bmatrix} \textit{action-process-fr} & \\ \text{INDEX} & e \\ \text{PRED} & \textit{meet} \\ \text{AGENT} & x \\ \text{THEME} & y \\ \text{TIME} & t \\ \text{LOCATION} & k \\ \vdots & \\ \text{REASON} & z \end{bmatrix}, \begin{bmatrix} \textit{tense-fr} & \\ \text{PRED} & \textit{past} \\ \text{ARG} & e \end{bmatrix} \right\rangle \end{bmatrix}$$

An NP$_i$ containing a relative pronoun such as *which*, *who*, and *that* bears an attribute-value specification [REL $\{x\}$] where x is the variable of said pronoun. Like PHON, GAP, and FRAMES, the value of REL is assumed to be percolated in the tree structure by phrasal constructions. In most cases, the value of the mother's REL is the union of the daughters' REL values (Sag, 1997, 2010; Kay and Michaelis, 2016), and in generall a REL-bearing filler phrase combines with a gapped clause via the construction in (42).

(42) The Wh-Relative Construction

$$\begin{bmatrix} \textit{phrase} & \\ \text{FORM} & \begin{bmatrix} \textit{phr-phon} \\ \text{PHON}\ P_1 \oplus P_2 \end{bmatrix} \\ \text{SYN} & \begin{bmatrix} \text{CAT} & K \\ \text{VAL} & \langle\,\rangle \\ \text{GAP} & G \end{bmatrix} \\ \text{FRAMES}\ F_1 \oplus F_2 & \end{bmatrix} \rightarrow X{:}\begin{bmatrix} \text{FORM} & \begin{bmatrix}\text{PHON}\ P_1\end{bmatrix} \\ \text{SYN} & \begin{bmatrix}\text{REL} & \{x\}\end{bmatrix} \\ \text{FRAMES}\ F_1 & \end{bmatrix} \begin{bmatrix} \text{FORM} & \begin{bmatrix}\text{PHON}\ P_2\end{bmatrix} \\ \text{SYN} & \begin{bmatrix} \text{CAT}\ K{:}\begin{bmatrix}\text{INV}\ -\end{bmatrix} \\ \text{VAL}\ \langle\rangle \\ \text{GAP}\ \langle \text{X}\rangle \bigcirc G \end{bmatrix} \\ \text{FRAMES}\ F_2 & \end{bmatrix}$$

The resulting clause $S_{[\text{REL}\{x\}]}$ can then combine with an NP via a construction that unifies the REL index of the clause with the index of the modified NP, i. e. '$NP_{[\text{REL}\{\,\}]} \rightarrow NP_y\ S_{[\text{REL}\{x\}]}$, as in Figure 19.[22]

In a sentence like *After Mary introduced herself to the audience, she turned to a man that she had met before*, the fronted clause is extracted from the main verb *turned* and consists of a preposition that selects a clause like the one discussed in Figure 18 as its complement. The fronted phrase combines with the matrix via the construction '$S_{[\text{GAP}\langle\rangle]} \rightarrow XP_x\ S_{[\text{GAP}\langle XP_x\rangle]}$' based on Sag (2010). Finally, the linking construction that the verb *turned* combines with require that it selects a directional PP complement. The structure of the entire sentence is depicted below in Figure 20. Only the attribute GAP is shown, for perspicuity.

7 Conclusions

Construction Grammar is a surface-driven, non-modular, generative, non-derivational, and monostratal approach to linguistic theory, which aims at cognitive plausibility and full coverage of the facts of any language under study without loss of generalization, within and across languages. The empirical commitment of construction grammar is that grammatical theory must in principle account for the totality of facts of any language, not recognizing a priori any theoretically privileged set of core grammatical phenomena, as the data appear to demand a cline of constructions, from

22 I assume relatives combine with NPs because of examples like those below.

i. [[The man and the woman] [(that) the priest married]] were Tim and Sue.

ii. [[Every man and every woman] [who appeared in the same picture]] exchanged numbers.

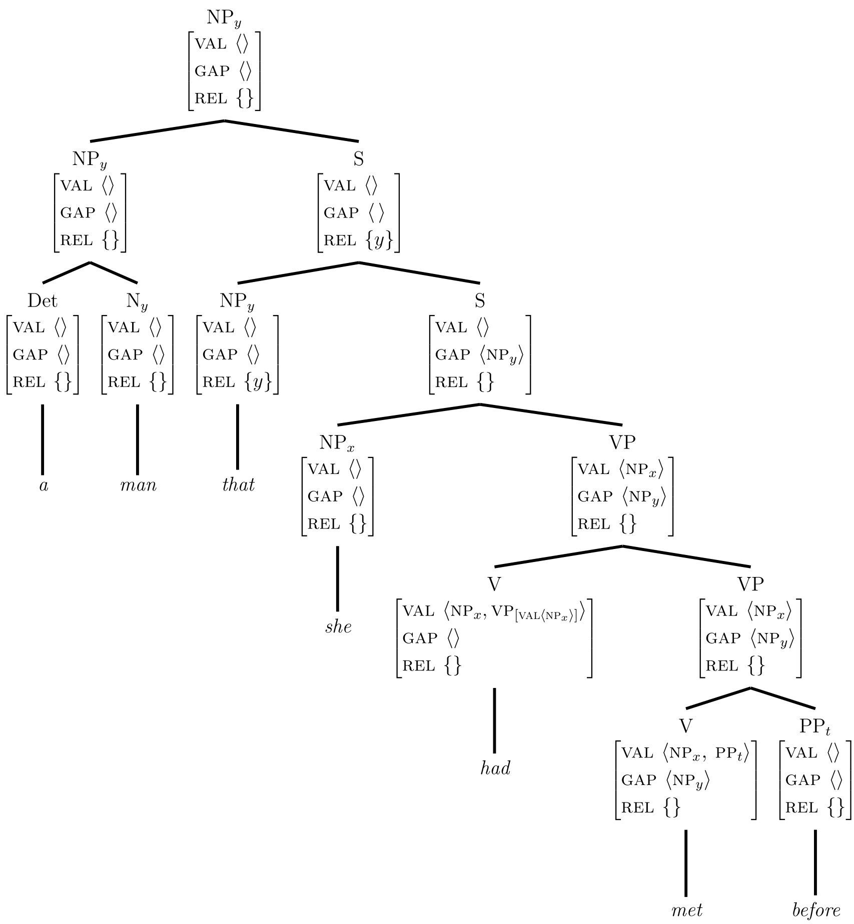


Figure 19: Structure of the NP *a man that she had met before* (AVMs abbreviated).

the relatively productive to the relatively frozen. The non-modular character of constructivist approaches assumes that form and meaning are part of each grammatical element, rather than located in separate components of the grammar. Construction Grammar aims to identify all the generalizations potentially available to the speaker of a language, though it is not assumed that the internal representation of the language in the mind of each speaker contains every generalization inherent in the data, as different speakers plausibly arrive at different generalizations and different grammars, and may regard different compositional structures as chunks, depending on the frequency to which they are exposed to such expressions during their daily life.

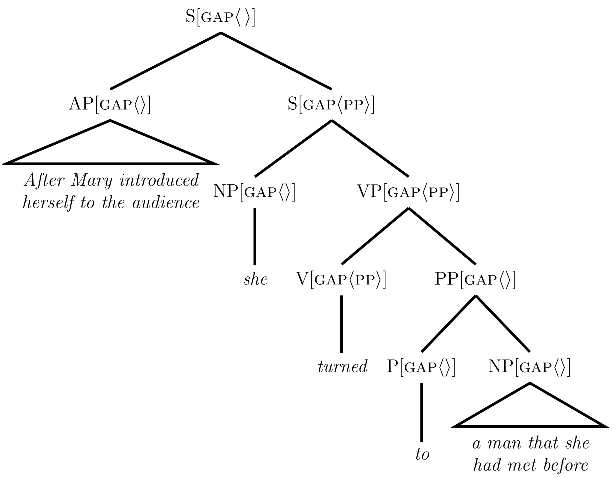


Figure 20: Structure of the S *After Mary introduced herself to the audience, she turned to a man that she had met before* (abbreviated AVMs).

References

Abbot-Smith, Kirsten & Michael Tomasello. 2006. Exemplar-learning and schematization in a usage-based account of syntactic acquisition. *The Linguistic Review* 23. 275–290.

Alegre, Maria & Peter Gordon. 1999. Frequency effects and the representational status of regular inflections. *Journal of Memory and Language* 40. 41–61.

Atkinson, Elizabeth Grace, Amanda Jane Audesse, Julia Adela Palacios, Dean Michael Bobo, Ashley Elizabeth Webb, Sohini Ramachandran & Brenna Mariah Henn. 2018. No evidence for recent selection at FOXP2 among diverse human populations. *Cell* 174. 1–12.

Bannard, Colin & Danielle Matthews. 2008. Stored word sequences in language learning: the effect of familiarity on children's repetition of four-word combinations. *Psychological Science* 19. 241–248.

Bergen, Benjamin K. & Nancy Chang. 2005. Embodied Construction Grammar in simulation-based language understanding. In J.-O. Östman & Mirjam Fried (eds.), *Construction Grammar(s): Cognitive and Cross-Language Dimensions*, 147–190. John Benjamins.

Bod, Rens. 2009. From Exemplar to Grammar: A Probabilistic Analogy-Based Model of Language Learning. *Cognitive Science* 33(5). 752–793.

Boland, Julie. 1993. The role of verb argument structure in sentence processing: Distinguishing between syntactic and semantic effects. *Journal of Psycholinguistic Research* 22. 133–152.

Booij, Geert. 2010. Compound construction: schemas or analogy? A construction Morphology perspective. In Sergio Scalise & Irene Vogel (eds.), *Compounding*, 93–108. Amsterdam/Philadelphia: Benjamins.

Borsley, Robert D. 2004. An approach to English comparative correlatives. In Stefan Müller (ed.), *Proceedings of the HPSG-2004 Conference, Center for Computational Linguistics, Katholieke Universiteit Leuven*, 70–92. Stanford: CSLI Publications. http://cslipublications.stanford.edu/HPSG/5/.

Borsley, Robert D. 2012. Don't move!. *Iberia* 4(1). 110–139.

Borsley, Robert D. & Frederick J. Newmeyer. 2009. On Subject-Auxiliary Inversion and the Notion 'Purely Formal Generalization'. *Cognitive Linguistics* 20(1). 135–143.

Botha, Rudolf P. 1988. *Forma and meaning in word formation: a study of Afrikaans reduplication*.

Bryant, John Edward. 2008. *Best-Fit Constructional Analysis*. Ph.d. dissertation, University of California at Berkeley.

Brysbaert, Marc, Michaël Stevens, Pawel Mandera & Emmanuel Keuleers. 2016. How many words do we know? Practical estimates of vocabulary size dependent on word definition, the degree of language input and the participant's age. *Frontiers in Psychology* 7. 1–11.

Bybee, Joan L. 2006. From usage to grammar: The minds response to repetition. *Language* 82. 711–733.

Bybee, Joan L. 2013. Usage-based theory and exemplar representation. In T. Hoffmann & G. Trousdale (eds.), *The Handbook of Construction Grammar*, 49–69. Oxford University Press.

Carpenter, Bob. 1992. *The Logic of Typed Feature Structures*. Cambridge University Press.

Chaves, Rui P. 2014. On the disunity of Right-Node Raising phenomena: extraposition, ellipsis, and deletion. *Language* 90(4). 834–888.

Chomsky, Noam. 1957. *Syntactic Structures*. Mouton: The Hague.

Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge, Massachusetts: The MIT Press.

Chomsky, Noam. 1975. *The Logical Structure of Linguistic Theory*. Chicago: University of Chicago Press.

Chomsky, Noam. 1981. *Lectures on Government and Binding*. Dordrecht: Foris.

Chomsky, Noam. 1989. Some notes on economy of derivation and representation. *MIT Working Papers in Linguistics* 10. 43–74.

Clark, Alexander & Rémi Eyraud. 2007. Polynomial Identification in the Limit of Substitutable Context-free Languages. *Journal of Machine Learning Research* 8. 1725–1745.

Clark, Alexander & Shalom Lappin. 2011. *Linguistic Nativism and the Poverty of the Stimulus*. Oxford: Wiley-Blackwell.

Corrigan, Roberta, Edith A. Moravcsik & Kathleen M. Wheatley. 2009. *Formulaic language. Volumes I–II*. Amsterdam/Philadelphia: John Benjamins.

Croft, William. 2001. *Radical Construction Grammar – Syntactic Theory in Typological Perspective*. Oxford University Press.

Culicover, Peter W. & Ray Jackendoff. 1999. The view from the periphery: The English comparative correlative. *Linguistic Inquiry* 30. 543–571.

Culicover, Peter W. & Ray S. Jackendoff. 2005. *Simpler Syntax*. Oxford University Press.

Davis, Anthony. 2001. *Linking by Types in the Hierarchical Lexicon*. Stanford, CA: CSLI Publications.

Diessel, Holger. 2015. Usage-based construction grammar. In Ewa Dabrowska & Dagmar Divjak (eds.), *Handbook of Cognitive Linguistics*, 295–321. Berlin: Mouton de Gruyter.

Dryer, Matthew S. 1997. Are grammatical relations universal. In Joan L. Bybee, John Haiman & Sandra Thompson (eds.), *Essays on Language Function and Language Type: Dedicated to T. Givon*, 115–143. Amsterdam: John Benjamins.

Elman, Jeff L. 1999. Origins of language: A conspiracy theory. In B. MacWhinney (ed.), *The emergence of language*, 1–27. Hillsdale, NJ: Lawrence Erlbaum Associates.

Enard, Wolfgang, Molly Przeworski, Simon E. Fisher, Cecilia S. L. Lai, Victor Wiebe, Takashi Kitano, Anthony P. Monaco & Svante Pääbo. 2002. Molecular evolution of FOXP2, a gene involved in speech and language. *Nature* 418. 869–872.

Engdahl, Elisabet & Enric Vallduví. 1996. Information Packaging in HPSG. In Claire Grover & Enric Vallduví (eds.), *Edinburgh Working Papers in Cognitive Science* (Studies in HPSG 12), Chapter 1, 1–32. Scotland: Centre for Cognitive Science, University of Edinburgh.

Feldman, Jerome & Srinivas Narayanan. 2004. Embodied Meaning in a Neural Theory of Language. volume 89, pages 385–392.

Feldman, Jerome, Ellen Dodge & John Bryant. 2009. Embodied Construction Grammar. In Bernd Heine & Heiko Narrog (eds.), *The Oxford Handbook of Linguistic Analysis*, 111–138. Oxford: Oxford University Press.

Fillmore, Charles J. 1982. Frame semantics. In The Linguistic Society of Korea (ed.), *Linguistics in the Morning Calm*, 111–137. Soeul: Hanshin.

Fillmore, Charles J. 1999. Inversion and constructional inheritance. In G. Webelhuth, J.-P. Koenig & A. Kathol (eds.), *Lexical and Constructional Aspects of Linguistic Explanation*, 113–128. Stanford: CSLI Publications.

Fillmore, Charles J. & Collin Baker. 2009. A Frames Approach to Semantic Analysis. In Bernd Heine & Heiko Narrog (eds.), *The Oxford Handbook of Linguistic Analysis*, 791–816. Oxford: Oxford University Press.

Fillmore, Charles J., Paul Kay & Catherine O'Connor. 1988. Regularity and Idiomaticity in Grammatical Constructions: The Case of let alone. *Language* 64. 501–538.

Fillmore, Charles J. 1985. Frames and the semantics of understanding. *Quadernie di Semantica* 2(6). 222–254.

Fillmore, Charles J. & Colin Baker. 2010. A Frames Approach to Semantic Analysis. In B. Heine & H. Narrog (eds.), *The Oxford Handbook of Linguistic Analysis*, 313–340. Oxford: Oxford University Press.

Fillmore, Charles J. & Paul Kay. 1996. Construction Grammar Coursebook, unpublished manuscript, UC Berkeley.

Fillmore, Charles J., Russell R. Lee-Goldman & Russell Rhomieux. 2012. The FrameNet construction. In Hans Boas & Ivan A. Sag (eds.), *Sign-Based Construction Grammar*, 309–372. Stanford: CSLI Publications.

Fodor, Janet D. & William Sakas. 2001. The structural triggers learner. In *Language acquisition and learnability*, 172–233. Cambridge: Cambridge University Press.

Gazdar, Gerald. 1982. Phrase structure grammar. In Pauline Jacobson & Geoffrey K. Pullum (eds.), *The Nature of Syntactic Representation*, 131–186. Dordrecht, Netherlands: D. Reidel.

Ghomeshi, Jila, Ray Jackendoff, Nicole Rosen & Kevin Russell. 2004. Focus reduplication in English (the salad-salad paper). *Natural Language and Linguistic Theory* 22. 307–357.

Ginzburg, Jonathan & Ivan A. Sag. 2000. *Interrogative Investigations: the form, meaning and use of English interrogative constructions*. Stanford: CSLI Publications.

Goldberg, Adele E. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.

Goldberg, Adele E. 2006. *Constructions at Work: the nature of generalization in Language*. Oxford: Oxford University Press.

Goldberg, Adele E. 2009. Constructions Work. *Cognitive Linguistics* 20(1). 201–224.

Goldberg, Adele E. January 2019. *Explain me this: Creativity, Competition, and the Partial Productivity of argument structure constructions*. Princeton University Press.

Hagoort, P. 2003. Interplay between syntax and semantics during sentence comprehension: Erp effects of combining syntactic and semantic violations. *Journal of cognitive neuroscience* 15. 883–899.

Hagoort, P. 2005. On broca, brain, and binding: a new framework. *Trends in Cognitive Sciences* 9. 416–423.

Harris, Zellig S. 1951. *Methods in Structural Linguistics*.

Haspelmath, Martin. 2007. Pre-established categories don't exist: consequences for language description and typology. *Linguistic Typology* 11(1). 119–132.

Hawkins, John A. 2004. *Efficiency and Complexity in Grammars*. Oxford.

Hoffman, Thomas & Graeme Trousdale. 2013. *The Handbook of Construction Grammar*. Oxford: Oxford University Press.

Inkelas, Sharon & Cheryl Zoll. 2005. *Reduplication: Doubling in morphology*. Cambridge: Cambridge University Press.

Jackendoff, Ray. 2010. The Ecology of English Noun-Noun Compounds. In *Meaning and the Lexicon: The Parallel Architecture, 1975–2010*, 413–451. Oxford University Press.

Jackendoff, Ray & Jenny Audring. 2014. Relational Morphology in the Parallel Architecture. In *The Oxford Handbook of Morphological Theory*. Oxford University Press.

Jackendoff, Ray S. 2002. *Foundations of Language: Brain, meaning, grammar, evolution*. Oxford: Oxford University Press.

Johnson, David & Shalom Lappin. 1999. Local Constraints vs Economy (Stanford Monongraphs in Linguistics). Stanford: CSLI Publications.

Kathol, Andreas. 2000. *Linear Syntax*. Oxford University Press.

Kay, Paul. 2002. An informal sketch of a formal architecture for construction grammar. *Grammars* 5. 1–19.

Kay, Paul & Charles J. Fillmore. 1999. Grammatical Constructions and Linguistic Generalizations: the *What's X Doing Y?* Construction. *Language* 75(1). 1–33.

Kay, Paul & Laura A. Michaelis. 2016. A few words to do with multiword expressions. In C. Condoravdi & T. Holloway King (eds.), *Festschrift for Lauri Karttunen*, 87–118. Stanford: CSLI Publications.

Kay, Paul & Karl Zimmer. 1990. On the semantics of compounds and genitives in English. In S. L. Tsohatzidis (ed.), *Meanings and Prototypes*, 239–246. London: Routledge.

Koenig, J.-P. 1999. *Lexical Relations*. Stanford: CSLI Publications.

Koenig, Jean-Pierre & Anthony Davis. 2003. Semantically Transparent Linking in HPSG. In Stefan Müller (ed.), *Proceedings of the HPSG-2003 Conference, Michigan State University, East Lansing*, 222–235. Stanford: CSLI Publications.

Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar, Volume 1: Theoretical Prerequisites*. Stanford, CA: Stanford University Press.

Lee-Goldman, Russell. 2011. Context in Constructions. Ph.d. dissertation, University of California, Berkeley.

Liu, Xiling, Mehmet Somel, Lin Tang, Zheng Yan, Xi Jiang, Song Guo, Yuan Yuan, Liu He, Anna Oleksiak, Yan Zhang, abd Yuhui Hu, Na Li, Wei Chen, Zilong Qiu, Svante Pääbo & Philipp Khaitovich. 2012. Extension of cortical synaptic development distinguishes humans from chimpanzees and macaques. Genome Research 22(4). 611–622.

Marcus, Garry F. & Simon E. Fisher. 2003. FOXP2 in focus: What can genes tell us about speech and language? *Trends in Cognitive Sciences* 7. 257–262.

Marcus, Gary F. 2001. Plasticity and nativism: Towards a resolution of an apparent paradox. In S. Wermter, J. Austin & D. Willshaw (eds.), *Emergent neural computational architectures based on neuroscience*, 368–382. Berlin: Springer-Verlag.

Marr, David. 1982. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. New York: Freeman.

Masini, Francesca & Anna M. Thornton. 2008. Italian VeV lexical constructions. In Angela Ralli, Geert Booij, Sergio Scalise & Athanasios Karasimos (eds.), *On-line Proceedings of the 6th Mediterranean Morphology Meeting (MMM6)*, 146–186.

McCawley, James D. 1968. Concerning the base component of a transformational grammar. *Foundations of Language* 4(1). 55–81. Reprinted in *Meaning and Grammar*. 35–58. New York, NY: Academic Press (1976).

Michaelis, Laura. 2003. Headless Constructions and Coercion by Construction. In E. J. Francis & L. A. Michaelis (eds.), *Mismatch: Form-Function Incongruity and the Architecture of Grammar*, 259–310. Stanford: CSLI Publications.

Michaelis, Laura. 2012. Making the Case for Construction Grammar. In Hans Boas & Ivan A. Sag (eds.), *Sign-Based Construction Grammar*, 31–69. Stanford: CSLI Publications.

Müller, Stefan. 2005. Resultative Constructions: Syntax, World Knowledge, and Collocational Restrictions: Review of Hans C. Boas: A Constructional Approach to Resultatives. *Studies in Language* 29(3). 651–681.

Newmeyer, Frederick J. 2004. Against a parameter-setting approach to typological variation. In Pierre Pica, Johan Rooryck & Jeroen Van Craenenbroeck (eds.), *Linguistic Variation Yearbook*.

Newmeyer, Frederick J. 2005. *Possible and Probable Languages: A Generative Perspective on Linguistic Typology*. Oxford University Press.

Reape, Mike. 1994. Domain Union and Word Order Variation in German. In John Nerbonne, Klaus Netter & Carl Pollard (eds.), *German in Head-Driven Phrase Structure Grammar* (CSLI Lecture Notes 46), 151–197. Stanford: CSLI Publications.

Ruppenhofer, Josef & Laura A. Michaelis. 2014. Frames and the Interpretation of Omitted Arguments in English. In S. Katz Bourns & L. Myers (eds.), *Linguistic Perspectives on Structure and Context: Studies in Honor of Knud Lambrecht*, 57–86. Amsterdam: Benjamins.

Sag, Ivan A. 1997. English relative clause constructions. *Journal of Linguistics* 33(2). 431–484.

Sag, Ivan A. 2010. English filler-gap constructions. *Language* 86(3). 486–545.

Sag, Ivan A. 2012. Sign-Based Construction Grammar: An informal synopsis. In Hans Boas & Ivan A. Sag (eds.), *Sign-Based Construction Grammar*, 69–202. Stanford: CSLI Publications.

Shieber, Stuart M. 1986. *An Introduction to Unification-Based Approaches to Grammar* (CSLI Lecture Notes Series 4). Stanford, CA: Center for the Study of Language and Information.

Steels, Luc. 2011. *Design Patterns in Fluid Construction Grammar*. Amsterdam: John Benjamins.

Steels, Luc & Joachim De Beule. 2006. Unify and Merge in Fluid Construction Grammar. In P. Vogt et al. (eds.), *Symbol Grounding and Beyond: Proceedings of the Third International Workshop on the Emergence and Evolution of Linguistic Communication* (Lecture Notes in Computer Science), 197–223. Springer.

Stefanowitsch, Anatol & Stefan T. Gries. 2005. Covarying collexemes. *Corpus Linguistics and Linguistic Theory* 1(1). 1–43.

Traugott, Elizabeth Closs & Graeme Trousdale. 2014. *Constructionalization and Constructional Changes*. Oxford: Oxford University Press.

Trueswell, John C. & Albert E. Kim. 1998. How to prune a garden path by nipping it in the bud: Fast priming of verb argument structure. *Journal of Memory and Language* 39. 102–123.

Vargha-Khadem, F., K. Watkins, K. Alcock, P. Fletcher & R. Passingham. 1995. Praxic and nonverbal cognitive deficits in a large family with a genetically transmitted speech and language disorder. *Procedures of the National Academy of Science USA* 92. 930–933.

Wechsler, Stephen. 1995. Preposition Selection Outside the Lexicon. In Raul Aranovich, William Byrne, Susanne Preuss & Martha Senturia (eds.), *Proceedings of the Thirteenth West Coast Conference on Formal Linguistics*, 416–431. Stanford, CA: CSLI Publications.

Wetta, Andrew. 2014. Construction-based Approaches to Flexible Word Order. Ph. D. thesis, University at Buffalo.

Peter W. Culicover and Ray Jackendoff

4 Simpler Syntax

Abstract: Simpler Syntax originates in the stance that syntax needs to be only complex enough to be able to relate sound and meaning. Unlike mainstream generative grammar, Simpler Syntax views phonology, syntax, and semantics as independent structures, linked by non-homomorphic interfaces. This stance leads to very concrete syntactic analyses, devoid of covert underlying forms and movement. Much of the linguistic complexity treated as syntactic by mainstream generative grammar proves to be a consequence not of syntax per se, but of the interaction between syntax and semantics. The present chapter sketches how Simpler Syntax approaches phrase structure, meaningful constructions, grammatical functions, control, raising, the passive, A' constructions, and ellipsis. We further suggest that many allegedly grammatical constraints are better accounted for as consequences of processing complexity. At the same time, it is not possible to eliminate syntax entirely; a bare-bones syntax is still necessary to relate sound and meaning.

1 Goals

Simpler Syntax (Culicover and Jackendoff, 2005, henceforth SS), is a framework for syntactic analysis that builds on the overall mentalistic approach of the Parallel Architecture (PA: Jackendoff, 2002, Jackendoff and Audring, this volume). In particular, SS envisions syntax as a component of human cognition, and its theoretical constructs are to be understood as hypotheses about the information utilized by the brain in language comprehension and production.

The central goal of SS is to describe and explain a language user's ability to establish a correspondence between meaning and sound (or in the case of signed languages, gesture). SS seeks to characterize this correspondence rigorously, invoking syntax only when other components of language such as semantics, prosody, or processing are insufficient. This perspective is embodied in the Simpler Syntax Hypothesis.

> **Simpler Syntax Hypothesis (SSH): Syntactic structure is only as complex as it needs to be in order to establish the relationship between phonological structure and semantic interpretation.**

As preliminary motivation for the SSH, we observe that every word has a phonological structure complex enough to differentiate its pronunciation from that of all other words: at least a sequence of speech sounds, each consisting of a collection of distinctive features, plus syllable, foot, and stress structure. Moreover, every word has a fairly complex meaning, necessary for differentiating its meaning from that of all

https://doi.org/10.1515/9783110540253-004

other words. In contrast, a word's syntactic specification typically involves at most a few features like syntactic category, grammatical gender, count vs. mass, number, and syntactic argument structure (subcategorization). As far as syntax is concerned, *dog*, *armadillo*, *table*, *triangle*, and *principle* are indistinguishable: they are all count nouns. In other words, most of the information in a word's lexical entry is phonological and semantic; syntax plays only a relatively minor role. The SSH proposes that this balance of informational content is characteristic of language as a whole.

SS originated as a response to Mainstream Generative Grammar (MGG: Chomsky, 1965, 1981, 1995a). It seeks to account for the same empirical phenomena, and more where possible – but with less machinery. However, SS does not make a case for "No Syntax" – the position that syntax is entirely dependent on semantics, as in Cognitive Grammar (Langacker, 1987, 1991, Broccias, this volume), or the even more extreme position that hierarchical constituent structure is unnecessary (Frank et al., 2012). Rather, it acknowledges the need for syntax while attempting to minimize it.

An important goal of SS, shared with MGG, is to explain how a child acquires a language, under minimal assumptions about "universal grammar" – those characteristics of language and language acquisition that arise as a product of genetically coded specializations in the human brain. The existence of such characteristics depends on the language-specific biological evolution of the brain during the time period since the hominid line broke off from the apes. Again, SS is not committed to eliminating such domain-specific factors in language acquisition, as are Cognitive Grammar and much of Construction Grammar (Goldberg, 1995, 2006; Croft, 2001, Chaves, this volume) – only to minimize them, as encouraged by Occam's Razor.

In general, this perspective leads to very concrete syntactic analyses. SS takes the position that if the meaning of an observed structure can be computed directly on the basis of that structure, there is no reason to complicate the analysis by assuming that there is invisible structure, or that the observed linear ordering of constituents is derived by movement. For instance, SS does not posit covert syntactic structure to account for elliptical phenomena. Rather, it works out an approach in which meaning is assigned to sentence fragments directly (Section 3.6; Culicover and Jackendoff, 2012). Nor does SS assume invisible syntactic arguments as part of the representation of control (Section 3.3; Culicover and Jackendoff, 2001, 2006, Jackendoff and Culicover, 2003). However, the SSH does not rule out covert structure in principle: for each specific case, the need for such structure is an empirical question. We mention below some cases where invisible structure seems essential.

A final goal: SS strives for formal rigor in formulating hypotheses about linguistic structure. This is essential especially when comparing rival accounts of phenomena. On the other hand, we do not pursue formalism for its own sake. As will be seen below, our formalism is adapted to the nature of the phenomena it is to account for, and in practice we do our best to strike an appropriate balance between rigor and readability.

2 Data

As a component of the PA, SS is open to the full range of data from the languages of the world, both “core” and “peripheral,” and both synchronic and diachronic (for the latter, see Culicover, 2014, 2016, in preparation). It makes use of introspective judgments, corpus searches, and, notably, evidence from psycholinguistic and neuroscientific experimentation on normal and brain-damaged individuals, as well as evidence from computational modeling (see for instance Culicover and Nowak, 2003).

An important aspect of the SS approach is what evidence from these sources is evidence *for*. In particular, a consequence of minimizing the complexity of syntax is the way SS deals with unacceptability judgments. Conventionally, if a sentence is judged to make sense semantically but is still less than fully acceptable, this is taken as a reflection of syntactic ill-formedness, other things being equal. For instance, the reduced acceptability of extraction from extraposed relative clauses, as in (1b), is generally taken to show that the syntax must incorporate a constraint that blocks such extraction (Ross, 1967; Chomsky, 1973, 1977).

(1) a. Chris met a writer at the party [who has just published an interesting book].
    b. *This is the book that Chris met a writer at the party [who has just published *t*].

In contrast to conventional practice, SS does not automatically take the unacceptability of examples like (1b) to be an indication that they are syntactically ill-formed. Rather, unacceptability may as well be due to a variety of nonsyntactic factors, including not only semantic ill-formedness (2a) but also prosodic ill-formedness (2b) and processing complexity (Miller and Chomsky, 1963; Kluender, 1998; Hofmeister and Sag, 2010; see Section 4 below).

(2) a. *Colorless green ideas sleep furiously.
    b. *Chris looked a writer who has just published an interesting book up.

In principle, the insistence on syntax being as simple as possible distributes the burden of explanation more broadly, with significant implications for how we interpret the data.

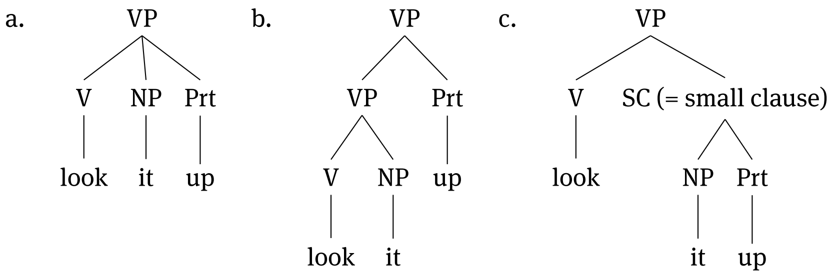
3 Tools

This section discusses the SS approach to a broad range of syntactic phenomena. The theory adapts analyses and formalisms from a variety of frameworks. For instance, SS adopts a constructional perspective, inspired by the PA and Construction Grammar (CxG; Fillmore et al., 1988; Goldberg, 1995, 2006; Sag, 2012; Chaves, this volume). This perspective permits accounts of correspondences between sound and meaning ranging from the very general, such as wh-questions, to the idiosyncratic, such as idioms, collocations, and a variety of "syntactic nuts" (Culicover, 1999; Culicover et al., 2017). We discuss this aspect of the theory further in Sections 3.2, 3.5, and 3.6.

Similarly, following Relational Grammar (Perlmutter, 1983; Perlmutter and Rosen, 1984), HPSG (Pollard and Sag, 1994; Müller and Machicao y Priemer, this volume), and LFG (Bresnan and Kaplan, 1982, Dalrymple and Findlay, this volume), SS takes relations such as Subject and Object to be grammatical primitives that play an essential role in determining the relation between semantic argument structure and its syntactic realization. Sections 3.3 and 3.4 sketch this Grammatical Function tier.

3.1 Phrase structure

We begin with the SS stance on phrase structure. There are two philosophies of what constitutes simplicity in phrase structure. Recent MGG (Chomsky, 1995a) argues that the simplest phrase structure (and therefore the *only* phrase structure) is uniformly binary branching. In contrast, SS posits that the simplest phrase structure is one with the fewest nodes, even at the cost of multiple branching. For instance, SS analyzes a VP like *look it up* as the structure in (3a), whereas MGG requires a structure with the geometry of either (3b) or (3c) (or something larger).

(3)


Trees (3b, 3c) contain more information than (3a): *look it* in (3b) and *it up* in (3c) form constituents. SS argues that without arguments for such constituency, a flat structure like (3a) is simpler. In fact, the absolute simplest structure would be one in which there are no internal constituents at all: a sentence would consist simply of a string of words. However, there is normally evidence for constituency of a relatively traditional sort, with constituents such as NPs, VPs, APs, PPs, and subordinate clauses. Section 5 mentions two cases where questions of constituency arise.

3.2 Constructions

Following the PA, SS assumes that linguistic structure as a whole involves a linkage among three independent levels or ***tiers***: phonology, syntax, and conceptual structure (semantics); a fourth tier, grammatical functions, is discussed in Sections 3.3 and 3.4. A well-formed sentence has well-formed structures in all the tiers, plus well-formed links between them.

Under this conception, a stereotypical word consists of a linkage between pieces of phonology, (morpho)syntax, and semantics. (4) illustrates the lexical entry for the verb *eat*; the links between the levels are notated by coindexation. The coindices can be thought of as marking the endpoints of association lines between the three structures.[1]

(4) Phonology: $/\text{it}/_1$
Syntax: $[_{\text{VP}}\ \text{V}_1\ \langle\underline{\text{NP}}_{\text{y}}\rangle\ ]$
Semantics: $[_{\text{Event}}\ \text{EAT}_1\ (\text{Agent: }\underline{\text{X}}, \text{Patient: }\underline{\text{Y}}_{\langle\text{y}\rangle}]$

In (4), the phonology $/\text{it}/_1$ is linked to the verb in syntax and to the semantic function EAT_1. The syntactic structure also encodes subcategorization: this verb optionally takes a direct object (where optionality is notated by angle brackets < >). The object position is underlined to indicate that this constituent is a variable: it must be instantiated by other material in order to be well-formed.

The semantic predicate has two arguments, denoted by *X* and *Y*. They are obligatory: an event can't be an event of eating unless there is an eater and something being eaten. The Agent variable *X* is canonically linked to subject position. The Patient argument is linked to the syntactic direct object (coindex y), if there is one. Because these constituents are variables, the link between them is also notated as a variable rather than a constant.

For a non-idiomatic VP, the phonological, syntactic, and semantic content of the direct object satisfies the verb's variables, and a general principle puts the pieces of phonology in linear order. (5) illustrates the structure of *eat the pizza*, the result of unifying the lexical entries of the three words.

(5) Phonology: $/\ /\text{it}/_1\ /\text{ðə}/_2\ /\text{pitsə}/_3\ /_4$
Syntax: $[_{\text{VP}}\ \text{V}_1\ [_{\text{NP}}\ \text{Det}_2\ \text{N}_3]_5\ ]_4$
Semantics: $[_{\text{Event}}\ \text{EAT}_1\ (\text{Agent: }\underline{\text{X}}, \text{Patient: }[\text{PIZZA}_3; \text{DEF}_2]_5\ )]_4$

Here, coindex 4 establishes the links among the entire phonological, syntactic, and semantic constituents. Coindex 1 links the three levels of the verb *eat*, as in lexical entry (4). The variable coindex *y* in the entry of *eat* unifies with the three levels of the

1 We ignore the treatment of inflection here. See Jackendoff and Audring (this volume) for regular inflection, and Jackendoff and Audring (forthcoming) for irregular inflection.
For reasons of simplicity and readability, our semantic representations are for the most part formulated in the notation of Conceptual Semantics (Jackendoff, 1990, 1997, 2002). These representations can alternatively be recast in terms of lambda notation, which is used in formal semantics to state functions over arbitrary variables.

direct object, resulting in coindex 5 for the NP as a whole. Finally, the syntactic combination of determiner and noun corresponds to a semantic structure in which PIZZA and definiteness codetermine the reference of the NP. This is all essentially equivalent to standard lexical insertion. (For details, see Jackendoff, 2002, chapter 12.)

Consider next an idiom such as *sell NP down the river*. This has internal syntactic structure that does not correspond one-to-one with semantic structure.

(6) Phonology: / /sɛl/$_6$ /$\underline{...}$/$_z$ /daʊn/$_7$ /ðə/$_2$ /rɪ-vɚ/$_8$ /$_w$
Syntax: [$_{VP}$ V$_6$ $\underline{\mathrm{NP}}_z$ [$_{PP}$ P$_7$ [$_{NP}$ Det$_2$ N$_8$]]]$_w$
Semantics: [$_{Event}$ BETRAY (Agent: $\underline{X}$, Patient: $\underline{Y}_z$)]$_w$

The verb, preposition, determiner, and noun in syntax are linked to corresponding phonology. But they are not linked to the meaning of the whole, which is unrelated to the meanings of the words. The direct object, however, *is* linked to a semantic argument in the normal way, and since it is a variable, it can be satisfied by any semantically appropriate NP. This structure is not derivable by MGG treatments of lexical insertion, in which syntactic structure is built up word by word and meaning is determined compositionally. SS, like HPSG and CxG, allows such structures to be stored in the lexicon as a whole.

Finally, SS allows for syntactic schemas consisting entirely of variables, without associated phonological structure, and with or without associated semantic structure. A case with associated semantics is English yes-no auxiliary inversion, in which the inverted order signals a question, for instance *Can Sue sing?* SS follows HPSG and CxG and introduces an inversion *construction*, which (for a first approximation) links a sentence with Aux preceding the subject to a semantic level containing a Question operator, notated as *Q*.

(7) Syntax: [$_S$ $\underline{\mathrm{Aux}}$ $\underline{\mathrm{NP}}$...]$_z$
Semantics: [Q [$_{Situation}$ $\underline{X}$]]$_z$

A case without associated semantics is the phrase structure rule for transitive VPs (8).

(8) Syntax: [$_{VP}$ $\underline{V}$ $\underline{\mathrm{NP}}$]

Here SS deviates from the sign-based versions of HPSG and CxG (Goldberg, 1995, Sag, 2012), which insist that every syntactic configuration is associated with a meaning. (8) is not a pairing of form and meaning; it just stipulates form. To be sure, direct objects stereotypically are Patients of an action (9a). But they can also have many other thematic roles (9b,c,d,e), or no thematic role at all with respect to the verb (9f,g,h); and the verb itself need not be associated with a meaning (9h).

(9) a. Sandy [$_{VP}$ ate the pizza]. [object is Patient]
b. Sandy [$_{VP}$ entered the room]. [object is Goal]
c. Sandy [$_{VP}$ fears chaos]. [object is Stimulus]

d. Sandy [$_{VP}$ impresses Pat]. [object is Experiencer]
e. Sandy [$_{VP}$ resembles Pat]. [object is Theme]
f. Sandy [$_{VP}$ believes Pat [$_{VP}$ to be smart]]. [object is Theme of embedded clause]
g. Sandy [$_{VP}$ took a walk]. [object is part of the predicate]
h. Sandy [$_{VP}$ kicked the bucket]. [verb and object are not individually linked to meaning]

SS therefore concludes that (8) has an independent existence in syntax over and above the many semantic uses to which it can be put.

For more on constructions, see Jackendoff and Audring (this volume), Chaves (this volume).

3.3 Grammatical functions; control and raising

Given SS's goal of eliminating abstract structure, it is somewhat unexpected that a proper linkage between syntactic and semantic argument structure requires a level of *grammatical functions* (GFs). However, unlike LFG and Relational Grammar, SS invokes grammatical functions only in a sharply delimited set of circumstances, so it is still relatively minimal. Its basic principles are the following (with some caveats to be set aside here; see Culicover and Jackendoff, 2005, chapter 6):

– GFs are assigned to *syntactic* arguments that are expressed by NPs, i. e. subjects, direct objects, indirect objects, and certain obliques, but not to arguments expressed by PPs, APs, or clauses, and not to modifiers.
– GFs are assigned to *semantic* arguments that are canonically expressed by NPs.
– Each clause has its own independent GF configuration (see (15)–(17) below).
– Within each clause, the GFs form a hierarchy from left to right. They are canonically linked to semantics according to a hierarchy of thematic roles: Agent > Patient/Recipient > Theme > other.[2]
– GFs are linked to syntax according to a hierarchy of syntactic roles, involving position, case-marking, and/or agreement.

In the stereotypical case, semantic arguments, GFs, and syntactic configuration are in alignment, as in (10): the subject links to the leftmost GF, which in turn links to the Agent, and the object links to the right-hand GF, which in turn links to the Patient. (For convenience, we collapse phonology and syntax into one line in (10).)

2 Psychological predicates have to stipulate the linking of GFs to thematic roles: the leftmost GF links to Stimulus in verbs like *frighten*, and to Experiencer in verbs like *fear* (see Jackendoff, 2007, chapter 6 for arguments for such a verb-specific mapping).

(10) Phonology/syntax: $[_{S} [_{NP} \text{Sandy}]_1 [_{VP} \text{ate}_2 [_{NP} \text{pizza}]_3]]_4$
GF: $[GF_1 > GF_3]_4$
Semantics: $[_{Event}\ \text{EAT}_2\ (\text{Agent: SANDY}_1, \text{Patient: PIZZA}_3)]_4$

However, GFs are also assigned to expletive subjects – NPs that do not correspond to a semantic argument (11). Such GFs are only syntactically linked.

(11) Phon/Syn: $[_{S} [_{NP} \text{it}]_1 [_{VP} \text{was drizzling}_2]]_3$
GF: $[\ GF_1]_3$
Semantics: $[_{Event}\ \text{DRIZZLE}_2]_3$

Other GFs that are linked only to syntax include those linked to meaningless idiom chunks, as in *kick the bucket*, and those linked to epenthetic reflexives with semantically one-argument verbs such as *behave oneself* and *perjure oneself*.

Cases in which a GF is only semantically linked arise in a variety of constructions such as infinitival and gerundive clauses, as well as clauses with null arguments in "pro-drop" languages. In such constructions, MGG posits a PRO or *pro*; SS (along with other constraint-based theories) treats them simply as VPs with no subject NP. Such a configuration looks like (12), in which coindex 1 appears only in GFs and semantics.

(12) Phon/Syn: $[_{VP} \text{to eat}_2\ \text{pizza}_3]_4$
GF: $[\ GF_1 > GF_3\ ]_4$
Semantics: $[\text{EAT}_2\ (\text{Agent: }\underline{X}_1, \text{Patient: PIZZA}_3)]_4$

The Agent role in the semantics of (12) is still a variable, which needs to be instantiated. There are various ways to do this, partly depending on the syntactic configuration. It can be understood as generic (what MGG calls PRO_{arb}), as in (13a). It can also be understood as the speaker, the addressee, or both, as in (13b). Or it can be taken as coreferential with some character in the discourse, rather like a pronoun, as in (13c).

(13) a. It's delightful to eat pizza.
b. To contradict myself/yourself/ourselves in public would be embarrassing.
c. To admit his mistake would embarrass Robin.

Some predicates require an infinitival complement whose Agent role is bound to the main verb's Agent (*Sandy tried to sleep*) or to the main verb's Patient/Recipient (*Sandy implored Robin to leave*). These are the standard cases of ***control***. (14) illustrates. Each clause gets its own GF-tier; binding is notated by the α superscript on the controller and the matching α in the embedded Agent position.

(14) Phon/Syn: $[\ \text{Sandy}_1\ \text{tried}_2\ [_{VP} \text{to sleep}_3\ ]_4\ ]_5$
GF: $[\ GF_1\ ]_5\ [\ GF_6\ ]_4$
Semantics: $[\text{TRY}_2\ (\text{Agent: SANDY}_1^{\alpha}, [\text{SLEEP}_3\ (\text{Agent: }\alpha_6\ )]_4\ )]_5$

So-called ***raising predicates*** such as *seem* and *believe* require an infinitival complement, and their syntactic argument structure provides an NP position that instantiates the infinitival's variable. This instantiation is indirect: the NP position is syntactically linked to a GF in the main clause. This GF is in turn linked to the leftmost GF in the subordinate clause, which may (15a) or may not (15b) be linked to the semantics of the infinitival.

(15) a. Pat seems to like Dana. [*Pat* is a semantic argument of *like*]
b. It seems to be drizzling. [*It* is not a semantic argument of *drizzle*]

The "raising" configuration looks like (16). (The connection between GFs is again notated by α superscripts.)

(16) Phon/Syn: $[\ \mathrm{Pat}_1\ \mathrm{seems}_2\ [_{\mathrm{VP}}\ \mathrm{to\ like}_3\ \mathrm{Dana}_4\ ]_5\ ]_6$
GF: $[\ \mathrm{GF}_1^{\alpha}\ ]_6\ [\ \mathrm{GF}_7^{\alpha} > \mathrm{GF}_4\ ]_5$
Semantics: $[\ \mathrm{SEEM}_2\ ([\mathrm{LIKE}_3\ (\mathrm{Exp{:}\ PAT}_7,\ \mathrm{Stim{:}\ DANA}_4)]_5\ )]_6$

In (16), the semantic argument PAT is linked to the leftmost GF in the subordinate clause (coindex 7). This GF is linked by the alphas to a GF in the main clause (coindex 1), which in turn is finally linked to *Pat* in the main clause in syntax.

The upshot in all these cases is that, in certain configurations, the leftmost GF need not appear in syntax if its associated semantics can be specified by other means. This treatment of control and raising parallels that in LFG (Dalrymple and Findlay, this volume) and HPSG (Müller and Machicao y Priemer, this volume).

3.4 Grammatical functions in the passive

The leftmost GF meets a different fate in the ***passive***. Three things are going on in the (English) passive. First, the verb has a special form which we provisionally notate as *Vpass*. Second, normally the most prominent thematic role in semantics is linked to the leftmost GF. However, in the passive, it is instead linked to a low-ranked GF, which may or may not be expressed in syntax as a *by*-phrase. (This corresponds to Relational Grammar's notion of a "chômeur": Perlmutter, 1983.) Third, the next most prominent thematic role links to the leftmost GF.[3]

(17) shows the configuration for the passive VP *eaten by the mice*.

(17) Phon/Syn: $[_{\mathrm{VP}}\ [_{\mathrm{Vpass}}\ \mathrm{eaten}]_1 \ldots [_{\mathrm{PP}}\ \mathrm{by}\ [_{\mathrm{NP}}\ \mathrm{the\ mice}]_2\ ]\ ]_3$
GF: $[\ \mathrm{GF}_4 > \mathrm{GF}_2\ ]_3$
Semantics: $[\mathrm{EAT}_1\ ([\mathrm{Agent{:}\ MICE}_2,\ \mathrm{Patient{:}}\ \underline{\mathrm{X}}_4\ )]_3$

If the *by*-phrase is absent, the oblique GF_2 is linked only to semantics, where the Agent is unspecified. Depending on context, it is instantiated semantically as a generic, a definite, or an indefinite.

3 NPs "raised" from subordinate clauses can also link to the leftmost GF, as in *Sam is believed to like gin*.

The structure in (17) is formulated in terms of a passive VP rather than a passive sentence like (18a), because there are many contexts where there is no (local) subject and the Patient is instantiated by other means, for instance (18b–i).

(18) a. The pizza was eaten (by the mice). [subject of tensed passive]
b. It's nice to be admired (by one's students) [generic unexpressed argument]
c. Being caught (by the cops) would be embarrassing. [speaker/hearer/both]
d. Being caught would be embarrassing for Pat. [character in discourse]
e. Pat tried not to be caught. [control]
f. Pat seems to have been caught. [raising]
g. Robin got/had Chris arrested by the cops. [causative *get*/*have* passive]
h. With Chris arrested by the cops, what next? [*with*-construction]
i. The protester arrested last month starved. [passive reduced relative]

Culicover and Jackendoff (2005) discuss many other details of passive constructions, as well as the role of GFs in other grammatical phenomena such as binding theory and Romance clitics.

3.5 A' constructions

So-called ***A' constructions*** (or ***long-distance dependencies***) have a gap in the canonical position of an argument or adjunct. The interpretation of the gap is supplied by material at the front of some clause that contains the gap. For instance, in *Who did you meet?*, the canonical direct object position is phonologically empty, and the A' constituent *who* is "understood" as specifying the semantic content of that position. A' constructions in English include direct and indirect wh-questions, where indirect wh-questions can be tensed or infinitival; relative clauses, both tensed and infinitival; free relatives (*whatever he says*); exclamatives (*what big teeth you have*); topicalization; and at least some cases of comparatives (Chomsky, 1977).

In the traditional analysis of A' constructions, the material at the front of the clause moves from an underlying position, leaving an unpronounced ***trace***, coindexed with the moved phrase.[4] Thus the sentence has two (or more) syntactic structures, such that the A' constituent is in situ in the first (or "underlying") structure and at the front in the last (or "surface") structure. For those A' constructions that have no clause-initial A' constituent, e. g. *the man to do the job*, either the A' constituent is deleted or it is posited to be an invisible operator.

4 We leave aside the details of movement as it has evolved in various versions of MGG.

SS (along with the other constraint-based approaches) takes a different and simpler approach. Rather than posit a sequence of structures, SS posits a single structure, the surface structure. If the construction in question has a clause-initial A' constituent, it is licensed there; if the construction does not have a clause-initial A' constituent, there is nothing there. The gap in the clause is occupied by a trace, as in the MGG treatment, but it is not a trace *of movement*, it is just an XP stipulated to be unlinked to phonology.[5]

The real action is in the semantics. The meaning of the clause (minus the A' constituent) can be conveniently expressed by a conventional lambda abstraction, whose variable corresponds to the trace, as in (19). (For convenience, we ignore tense.)[6]

(19) Phon/Syn: ... $[_S$ did Ozzie$_1$ $[_{VP}$ drink$_2$ t$_3$ $]]_4$
Semantics: λx.[DRINK$_2$ (Agent: OZZIE$_1$, Patient: x$_3$) $]_4$

The lambda expression in (19) can be read informally as 'thing such that Ozzie drank it.' In the wh-question as a whole, the wh-word is equated with this individual, and the whole is within the scope of a question operator Q that is bound to the wh-word. This might be notated as (20), roughly 'what is the thing such that Ozzie drank it?'

(20) Phon/Syn: $[_S$ what$_5$ $[_S$ did Ozzie$_1$ $[_{VP}$ drink$_2$ t$_3$ $]]_4]_6$
Semantics: $[Q^{\alpha}$ [THING$^{\alpha}_5$ = λx.[DRINK$_2$ (Agent: OZZIE$_1$, Patient: x$_3$) $]_4$ $]]_6$

The specifications of long-distance dependencies can be factored into two parts, following Sag 1997. One part, which they all have in common, is the lambda abstraction whose variable is bound to the gap. All the constraints on extraction pertain to this part, whether due to syntax, semantics (Erteschik, 1973, Kluender, 1992, Kuno and Takami, 1993, Van Valin, 1998) or processing difficulty (see Section 4). Since this part of the construction is (relatively) indifferent as to which kind of A' construction it is in (question, relative, topicalization, etc.), the same constraints obtain across all of them. This captures the fundamental observation by Ross (1967) and built on by Chomsky (1977), to the effect that all long-distance constructions obey (largely) the same constraints.

The second part of the specification of long-distance dependencies is the configuration of the left edge.

– Direct wh-questions begin with a range of possible wh-expressions – *who*, *what*, *which*, *when*, *where*, *how*, *why* (and any of these except *which* may be followed in casual speech by expressions like *the hell*) or more complex wh-phrases; and the

5 This is a case where SS requires a syntactic constituent with no phonological realization. Ideally, following the SSH, the gap ought to be a gap in syntax as well as phonology, but we have (so far) been unable to eliminate traces from syntax.

6 There are other possible formalizations, for instance Langacker's (1987) notion of 'profiling': picking out a character in an event, here the wh-phrase, as the locus of attention. Culicover and Jackendoff (2005) also formalize this configuration somewhat differently, tying it in with ellipsis.

auxiliary follows the wh-expression, resulting in inversion for non-subject wh-expressions.
- Indirect wh-questions begin the same way, but the auxiliary is not inverted.
- Tensed relative clauses have a different range of wh-expressions (**the wine what Ozzie drank*, **the way how he did it*), do not allow *the hell*, and do not display auxiliary inversion.
- Zero-relatives (*the wine Ozzie drank*) and *that*-relatives (*the wine that Ozzie drank*) lack wh-words in A' position.
- Exclamatives begin with *what* or *how*, and do not invert the auxiliary.
- Free relatives allow yet another range of wh-expressions, including *whoever* and all the other *-ever* forms.
- Infinitival indirect questions lack tense and require a wh-expression at the front (21a); infinitival relatives also lack tense but disallow a wh-expression (21b), though they allow preposition+wh-word (21c). The difference between these two constructions seems especially arbitrary (though see Culicover, 2011, 2013).

(21) a. Joe wonders *(who) to vote for *t*.
b. the man (*who) to vote for *t*.
c. the man for whom to vote *t*.

Similarly, each long-distance dependency construction has its own characteristic semantics surrounding the lambda-expression. Wh-questions target an entity whose identity is requested, based on the properties picked out by the lambda-expression. Relative clauses (with or without wh-words) bind the lambda-expression to the noun heading the relative clause. Topicalization marks the individual equated with the lambda-expression as topic. Other A' constructions have similar analyses.

The upshot is that long-distance dependencies can be treated as constructions that link a characteristic syntax and phonology to a characteristic semantics (however the semantics is formalized). They all make use of a clause containing a trace; this clause is linked to a lambda-expression whose variable is linked to the trace. These constructions differ in the syntactic and phonological properties of the left edge of the clause (up to and including the auxiliary or absence thereof), and they differ in the way the lambda-expression is integrated into the meaning of the sentence. There is no need for movement, for invisible interrogative complementizers, or for invisible operators that undergo movement. Hence, importantly, there is no underlying syntactic structure; aside from traces, the syntax is as simple as possible.

3.6 Ellipsis

Elliptical constructions are sentence fragments that have the interpretation of a full clause. In the classical approach, the interpretation is entirely filled out by fully specified syntactic structure surrounding the fragment in underlying form; this structure is either deleted in the course of derivation or invisible from the start (Wasow, 1979, Merchant, 2001, 2004, and many others). An alternative approach is that the unspoken parts of the interpretation are derived from the *semantic* structure of an antecedent in the linguistic and/or nonlinguistic context, with syntax playing a guiding role. Following the SSH, SS works out this interpretive alternative, attempting to minimize syntactic structure as much as possible (Culicover and Jackendoff, 2005 and especially Culicover and Jackendoff, 2012; see also Hardt, 1992, 1993, 1999, 2004, 2008; Hardt and Romero, 2004, Jackendoff, 1972, Kubota and Levine, 2013, 2014a,b,c, 2016).

Culicover and Jackendoff (2012) propose that the interpretation of elliptical constructions is based on the domain-general cognitive relation ***Same-Except***: Item 1 is the ***same*** as item 2, ***except*** for part P1 of item 1, which is different from the corresponding part P2 of item 2. This general type of relation was noticed by William James (1890, 528), who points out that the more two items are perceived as alike, the more the remaining differences between them pop out. He also observes that this relation can play a role in judging visual, auditory, and even gustatory percepts.

To see how Same-Except applies to ellipsis, consider perhaps the syntactically simplest elliptical construction, Bare Argument Ellipsis (BAE), illustrated in B's response in (22). In the interpretive account of BAE, the syntactic structure of (22B) is just the fragment $[_{\text{Utterance}}$ No, $[_{\text{NP}}$ scotch]].

(22) A: Ozzie's drinking bourbon again.
B: No, scotch.

The interpretation of (22B) is the same as (22A) – except for the part that is pronounced, namely *scotch*. *Scotch* contrasts with the parallel part of (22A), namely *bourbon*. This suggests a basic strategy for the analysis of elliptical constructions: any part of the interpretation that is not expressed overtly is understood to be the same as the interpretation of some antecedent; and the actual sentence fragment (or some particular part of it) is understood as an "Except," contrasting with the corresponding part of the antecedent.

In order to fill out the interpretation of an elliptical utterance, then, two steps are necessary. The first is to ***find*** a potential antecedent. In the case of (22B), the antecedent is obviously (22A). But the process is not always so simple: it depends on semantics and pragmatics as well as syntax, as can be seen from the dialogues in (23) and (24), which are syntactically identical.[7]

7 Normally the antecedent is in an immediately preceding clause. But this condition can be overridden in cases such as the following, due to Hardt (1993) (which has a sort of jokey quality):

(i) A: You never go swimming.
B: That's because we don't look good in a swimming costume. We might [~~go swimming~~] if we did [~~look good~~].

(23) A: I hear Ozzie's drinking bourbon again.
B: No, scotch.

[= 'Ozzie's drinking scotch again'; ≠ 'I/you hear Ozzie's drinking scotch again']

(24) A: I doubt Ozzie's drinking bourbon again.
B: *No, scotch. [uninterpretable]

The second step in interpreting the fragment is to ***align*** the relevant features of the fragment and the antecedent, and to ***identify differences***. If the fragment does not share properties with some part of the antecedent, there can be no match. Recognition of semantic as well as syntactic features is again crucial. Consider B's responses in (25), all of the form [$_{\mathrm{PP}}$ *on NP*].

(25) A: Ozzie drank his scotch on the porch.
B: i. No, on the patio.
ii. *No, on Tuesday.
iii. Yes, on Tuesday.

In (25Bi), B's *no* implies there is a contrast (or correction) to come. *On the patio* and *on the porch* are both locative PPs, so they align syntactically and semantically. Hence *on the patio* is a possible Except. In (25Bii), *no* similarly signals a contrast. But this time, the fragment is a time phrase, so it cannot align semantically with *on the porch*. Finally, in (25Biii), *yes* signals agreement, that is, acknowledgment of *on the porch*. This time the fragment offers supplementary information, an elaboration of A's statement, so it aligns with the time of Ozzie's drinking, which is unspecified in (25A). Overall, then, (23)–(25) show that the felicity of BAE depends on a complicated interplay of syntax, semantics, and implicature; it cannot be a purely syntactic deletion operation.

The tableau in (26) informally notates the semantics of the four terms involved in the Same-Except in (22).

(26)

	Antecedent	**Elliptical expression**
SAME	[DRINK (OZZIE, BOURBON)]	[$_{\text{Situation}}$…SCOTCH…]
EXCEPT	BOURBON	SCOTCH

The first row shows that the antecedent situation of Ozzie drinking bourbon is the same as some unspecified situation involving scotch. The second row shows the difference between the two situations: the elliptical expression has *scotch* where the antecedent has *bourbon*. To fill in the interpretation of the elliptical expression, one can think of "solving" for it inferentially: subtracting *bourbon* from the antecedent, replacing it with a variable, (yielding (27a), and then instantiating this variable with *scotch*, yielding the full interpretation (27b):[8]

(27) a. [DRINK (OZZIE, BOURBON)] – BOURBON = λx.[DRINK (OZZIE, x)]
b. λx.[DRINK (OZZIE, x)] + SCOTCH = [DRINK (OZZIE, SCOTCH)]

The same principles are in play in single-phrase answers to wh-questions, such as (28), in which the contrasting constituents are *what* and *scotch*.

(28) A: What did Ozzie drink?
B: Scotch.

Another type of ellipsis that closely resembles BAE is sluicing, exemplified by the underlined phrases in (29).

(29) a. Ozzie drank the scotch in five minutes, but I can't tell you <u>where</u>.
b. Ozzie's drinking again, but I don't know <u>what</u>.
c. Abby speaks the same language that some guy in this class speaks, but I'm not sure <u>who</u>.

Like BAE, sluicing requires alignment. In a case-marking language, syntactic alignment fails if the contrasting expressions do not match in case. This accounts for the well-known examples of sluicing in German (Ross, 1969), shown in (30).

(30) a. Er will jemandem schmeicheln, aber sie wissen nicht wem/
he wants someone-DAT flatter, but they know not who.DAT/
*wen.
who.ACC
b. Er will jemanden loben, aber sie wissen nicht wen/ *wem.
he wants someone-ACC praise, but they know not who.ACC/ who.DAT

For every example of sluicing, there exists a corresponding example of BAE. For instance, the sluicing examples in (29) are paralleled by the BAE examples in (31).

(31) a. A: Ozzie drank the scotch in five minutes.
B: i. (Yeah,) in the kitchen.
ii. Where?
b. A: Ozzie's drinking again.
B: i. Yeah, scotch.
ii. (Yeah, but) what?

8 (28) reflects a widely proposed approach to ellipsis in the literature, e. g. Dalrymple et al. (1991); Hardt (1993, 1999). Notice also that the lambda-abstraction in (28) is of exactly the same form as that proposed for long-distance dependencies. This cannot be an accident.

c. A: Abby speaks the same language that some guy in this class speaks.
B: i. Yeah, Charlie.
ii. Who?

This parallel suggests that BAE and sluicing are two variants of the same construction. If BAE cannot be formulated in terms of syntactic deletion, then sluicing should not be so formulated either, contra e. g. Ross (1969), Merchant (2001), Chung, Ladusaw and McCloskey (1995, 2011). To push the point further, (32) offers two examples of sluicing for which it is impossible to specify an underlying full syntactic form that reflects the interpretation of the elliptical clause.[9]

(32) a. Fred either ate garlic or forgot to brush his teeth, but I can't tell you which.
b. It seems we stood and talked like this before. We looked at each other in the same way then. But I can't remember where or when. (Rodgers and Hart, *Where or When*)

Another prominent type of ellipsis is verb phrase ellipsis. This construction has the form of a clause that lacks an overt VP, as in (33).

(33) a. Sandy sang, and Chris did too.
b. Sandy sang, but Chris doesn't want to.

For convenience we take the VP in the elliptical clauses to be VP_{pro}, that is, a constituent with null phonology. This VP gets its interpretation through the Same-Except relation: the interpretation of its clause is the same as that of its antecedent, except for any part of the elliptical clause that differs from its counterpart in the antecedent. For instance, in (33a), aligning and identifying differences marks the subjects as the only contrasting constituent, so the tableau comes out as (34).

(34)

	Antecedent	**Elliptical expression**
SAME	[SING (SANDY)]	[$_{Situation}$...CHRIS...]
EXCEPT	SANDY	CHRIS

(35) "solves" for the full interpretation of *Chris did too*.

(35) [SING (SANDY)] – SANDY = λx.[SING (x)]
λx.[SING (x)] + CHRIS = [SING (CHRIS)]

9 Dalrymple et al. (1991, 442) and Lappin (2005) recognize that inference is necessary for examples like these, but they do not draw any general conclusions. Hardt (1993) uses such cases to argue for inferential interpretation and against copying/deletion. On his account, the individual antecedent events or actions are combined in the semantics into a composite antecedent for the ellipsis, essentially the approach advocated in SS. (32a) is of a type cited by Webber 1978.

Culicover and Jackendoff (2012) apply this approach to a wide range of elliptical expressions, including gapping, pseudo-gapping, and *one*-anaphora, as well as anaphoric expressions such as *vice versa* and *same here* that do not lend themselves to an account in terms of covert syntax. They also show that the Same-Except relation is invoked in the interpretation of contrastive stress and, not surprisingly, in the semantics of words such as *same*, *likewise*, *except*, and *instead*.

4 Constraints that are not part of the grammar

One important implication of SS is the hypothesis that constraints such as subjacency, the complex NP constraint, and the subject condition are not necessarily grammatical phenomena in the narrow sense. In fact, a growing body of evidence suggests that many of them are consequences of processing complexity, arising from particular configurations that are otherwise well-formed (Arnon et al., 2005; Hofmeister et al., 2007; Hofmeister and Sag, 2010; Hofmeister et al., 2013; Kluender, 1991, 1992, 1998; Kluender and Kutas, 1993a,b; Sag et al., 2007; Hawkins, 2004, 2014; for counterarguments, see Phillips, 2013; Sprouse and Hornstein, 2013). For example, recent studies of 'subject island' extractions suggest that several factors bear on acceptability, including discourse properties of what is extracted, the lexical head of the subject, and the lexical properties of the predicate (Chaves and Dery, 2013; Clausen, 2011).

As with other issues discussed in Section 1, SS does not take the rigid position that processing considerations can replace *all* grammatical constraints as explanations for unacceptability judgments; the correct balance between the two is an empirical matter. This said, we review a representative phenomenon, 'freezing', in which the judgments appear to have much the same status as those that have been attributed to grammatical constraints, but where processing considerations offer a plausible alternative.

Ross (1967) observed that extraction from a PP in extraposed position is reduced in acceptability. The PP in (36a) is an extraposed modifier of the direct object, and the related question in (36b) is less acceptable. But if the PP is within the direct object, extraction encounters no difficulty (36c).

(36) a. You saw [a picture] yesterday [$_{PP}$ of Thomas Jefferson]
b. ?Who did you see [a picture] yesterday [of *t*]?
c. Who did you see [a picture [of *t*]] yesterday?

Ross proposed to deal with such examples with a constraint on syntactic derivation that blocks extraction from extraposed constituents. Subsequently, Wexler and Culicover (1980) proposed the Freezing Principle: if a structure is created transformationally and is not compatible with the base phrase structure rules of a language, it is frozen; nothing can be extracted from it. Both of these solutions rely on constraints on syntactic derivations, and in particular, on movement.

The SSH suggests that the syntax need not be responsible for explaining why extractions from 'frozen' contexts are less than fully acceptable – if a satisfactory explanation can be found in terms of processing complexity. Such an explanation is proposed by Hofmeister et al. (2015). They find that the unacceptability of (36b) depends on two interacting dependencies, one between the direct object and the extraposed PP, and one between the wh-word and the trace within the PP. The dependencies are shown in (37).

(37) ?who did you see a picture yesterday $\boxed{\text{of } t}$

The experimental evidence cited by Hofmeister et al. suggests that the unacceptability is an additive function of the length of the dependencies. This result strongly resembles findings elsewhere in the psycholinguistic literature on the effects of dependency locality (Gibson, 1998, 2000; Grodner and Gibson, 2005): in general, the longer the dependency, the lower the acceptability judgment. Konietzko et al. (2018) offers a similar account for the unacceptability of extraction from shifted heavy NPs.

This analysis of freezing phenomena illustrates the implications of SS for explaining acceptability judgments. Sentences like (36b) are grammatically well-formed, in that every local configuration conforms to the requirements of the grammar; yet they arguably create enough processing difficulty to be judged unacceptable. Precisely this argument was made by Miller and Chomsky (1963) concerning the unacceptability of multiple center-embedding: the grammar generates them, but nevertheless they cannot be readily processed. While processing complexity may not be the correct account in every case, in this case it offers a plausible account, taking some of the burden off the syntax.

5 Sample analysis

We now turn to some aspects of the structure of the sample sentence (38).

(38) After Mary introduced herself to the audience, she turned to a man that she had met before.

Figure 1 gives one version of the syntactic tree for the main clause. For convenience, the phonology is notated as terminal elements in the tree, connected by dashed lines, and the semantics is omitted; for a more detailed, "purer" version of the first clause, keeping the components separate, see Jackendoff and Audring (this volume).

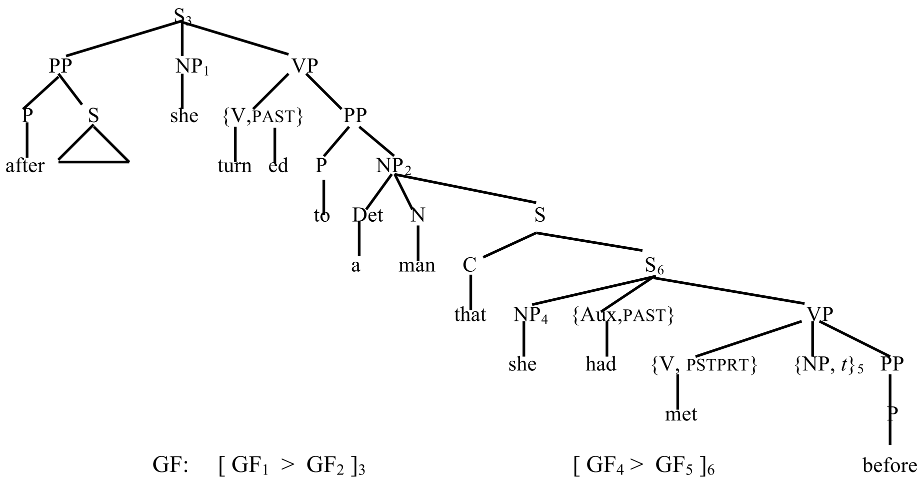


Figure 1: Syntax of the main clause of the sample sentence.

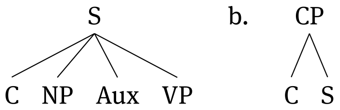
The syntactic analysis closely follows the apparent surface structure, rather in the style of Introductory Linguistics syntax, so we will mention only a few general points, plus points where there are possible variants.

– As advertised in Section 3.1, the phrase structure is not confined to binary branching. In particular, the main clause has ternary branching. There is no need, other than semantics, to adjoin the *after*-clause to the main clause, under a higher node.
– The NP *a man that*... also has ternary branching. Culicover and Jackendoff (2005) (chapter 4) champion this structure: there is no defensible reason either to group the Det and the N together or to group the N and the S together (or, alternatively, there are arguments both ways, and only flat structure can accommodate both).
– *After* is treated as a preposition which allows a clause as its complement.
– The pronoun *she* is treated as a full NP rather than as a noun dominated by an NP. Its coreference with *Mary* in the full sentence (38) is coded in the semantics rather than the syntax.
– In the relative clause, *before* is treated as a preposition that is the only daughter of PP. It is possible that in such cases, the X^0 category and the XP category coalesce.[10]
– *Before* has an implicit object (i. e. a semantic argument not expressed in syntax) which is anaphoric to the time of the main clause ('before the time that she introduced herself'). This anaphora is encoded in the semantics, not in the syntax.
– The complementizer *that* is treated as adjoined to S, the two forming an S. Two alternatives: (a) *that* is within the subordinate clause, as in (39a), or (b) it heads a

10 More generally, there may be no categorical distinction between a head and a maximal projection (Chomsky, 1995b).

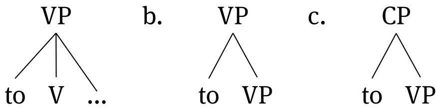
CP with the clause as its complement, as in (39b). SSH favors (39a), all else being equal (but all else is probably not equal).

(39) a.



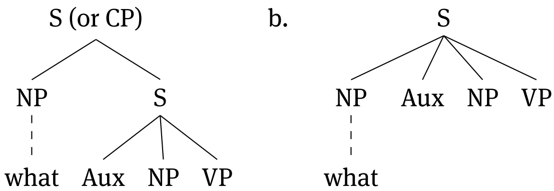
– The same issue arises with infinitival complements: which of the structures in (40) is correct? The SSH favors (40a), again all else being equal; this is what was shown in the discussion of infinitives in Section 3.3 (cf. (12) and (14)). (On the other hand, VP-ellipsis favors (40b) or (40c).)

(40) a.



– Similarly with long-distance dependency constructions: Does a wh-question, for instance, have the conventional structure (41a), or the flatter (41b) favored by the SSH? Section 3.5 assumed a structure like (41a) (cf. (20)), but it is an open question whether (41b) is able to capture all the relevant empirical phenomena.

(41) a.



– The GF tier in Figure 1 registers two clauses, the main clause and the relative clause (coindices 3 and 6 respectively). In each clause, the subject corresponds to a GF (coindices 1 and 4). In the main clause, *a man* is an oblique argument of the complex verb *turn to NP*, so it corresponds to a GF (coindex 2). In the relative clause, we tentatively link the trace to a GF (coindex 5).

6 Evaluation

The primary evaluative mechanism for SS is simplicity, understood as the application of Occam's Razor: does the analysis account for the data while minimizing assumptions? The assumptions in question cover the full range of formal universals proposed in various versions of generative grammar: categories, features, invisible structure, underlying forms, derivations, constraints, rule types, inheritance, grammatical functions, and so on. SS proposes to remove from syntactic theory any strictly syntactic mechanisms whose work can be accounted for by independently required non-syntactic mechanisms, such as prosody, semantics, pragmatics, and processing. In particular, if there is a choice between accounting for something in the syntax or in the semantics, prima facie the complication should be in semantics, because semantics is necessary anyway in order to explain inference and reference. Assuming this as a general principle, we have sketched here how syntax can be made very simple indeed – though by no means reduced to zero. The syntax is still responsible for such matters as where the verb belongs in the VP, what functional categories such as determiners, modals, and particles may appear and where they are positioned, what long-distance dependencies are available, what repertoire of cases is available and how cases are assigned, what has to agree with what, and so on.

7 Conclusion

Throughout our exposition we have contrasted the SS approach with Mainstream Generative Grammar, and we have highlighted similarities to and differences from other constraint-based theories, particularly LFG, HPSG, and CxG. Moreover, at the outset we stressed that SS is conceived as a theory of one faculty of the human mind, embedded in the PA's overarching outlook. SS's minimalist approach is quite different from that of the Minimalist Program; in particular it holds that not every aspect of a string of words is the responsibility of the syntax of a language. Rather, the burden of explanation for linguistic phenomena should be assigned as is appropriate to the various components of linguistic architecture (including syntax of course), to language processing, and to other mental capacities that are external to linguistic competence.

References

Arnon, Inbal, Philip Hofmeister, T. Florian Jaeger, Ivan A. Sag & Neal Snider. 2005. Rethinking superiority effects: a processing model. Poster presented at the CUNY Sentence Processing Conference.

Bresnan, Joan & Ronald M. Kaplan. 1982. Lexical-Functional Grammar: A Formal System for Grammatical Representation. In Joan Bresnan (ed.), *The Mental Representation of Grammatical Relations*, 173–281. Cambridge, MA: The MIT Press.

Broccias, Cristiano. *This volume*. Cognitive grammar.

Chaves, Rui. *This volume*. Construction grammar.

Chaves, Rui P. & Jeruen E. Dery. 2013. Which subject islands will the acceptability of improve with repeated exposure. In *31st West Coast Conference on Formal Linguistics*, Citeseer.

Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Chomsky, Noam. 1973. Conditions on transformations. In Stephen Anderson & Paul Kiparsky (eds.), *A Festschrift for Morris Halle*, 232–286. New York: Holt, Reinhart & Winston.

Chomsky, Noam. 1977. On *wh*-movement. In Peter W. Culicover, Thomas Wasow & Adrian Akmajian (eds.), *Formal syntax*, 71–132. New York: Academic Press.
Chomsky, Noam. 1981. *Lectures on government and binding*. Dordecht: Foris Publications.
Chomsky, Noam. 1995a. *The Minimalist Program*. Cambridge, MA: MIT Press.
Chomsky, Noam. 1995b. Bare phrase structure. In Gert Webelhuth (ed.), *Government binding theory and the minimalist program*, 383–439. Oxford: Oxford University Press.
Chung, Sandra, William Ladusaw & James McCloskey. 1995. Sluicing and Logical Form. *Natural Language Semantics* 3. 239–282.
Chung, Sandra, William Ladusaw & James McCloskey. 2011. Sluicing(:) Between Structure and Inference. In Rodrigo Gutiérrez-Bravo, Line Mikkelsen & Eric Potsdam (eds.), *Representing Language: Essays in Honor of Judith Aissen*, California Digital Library eScholarship Repository, 31–50. University of California, Santa Cruz: Linguistic Research Center.
Clausen, David R. 2011. Informativity and acceptability of complex subject islands. Poster presented at the 24th annual CUNY Sentence Processing Conference, Stanford, CA.
Croft, William. 2001. *Radical Construction Grammar*. Oxford: Oxford University Press.
Culicover, Peter W. 1999. *Syntactic nuts: hard cases, syntactic theory, and language acquisition*. Oxford: Oxford University Press.
Culicover, Peter W. 2011. A reconsideration of English relative clause constructions. *Constructions* 2.
Culicover, Peter W. 2013. *Grammar and complexity: Language at the intersection of competence and performance*. Oxford: Oxford University Press.
Culicover, Peter W. 2014. Constructions, complexity and word order variation. In Frederick J. Newmeyer & Laurel B. Preston (eds.), *Measuring linguistic complexity*, 148–178. Oxford: Oxford University Press.
Culicover, Peter W. 2016. Syntactic change in the Parallel Architecture: The case of parasitic gaps. *Cognitive Science* 41. 213–232.
Culicover, Peter W. in preparation. *The origin of languages*. Unpublished ms., The Ohio State University.
Culicover, Peter W. & Ray Jackendoff. 2001. Control is not movement. *Linguistic Inquiry* 32(3). 493–512.
Culicover, Peter W. & Ray Jackendoff. 2005. *Simpler syntax*. Oxford: Oxford University Press.
Culicover, Peter W. & Ray Jackendoff. 2006. Turn over control to the semantics!. *Syntax* 9(2). 131–152.
Culicover, Peter W. & Ray Jackendoff. 2012. Same-Except: A domain-general cognitive relation and how language expresses it. *Language* 82(2). 305–340.
Culicover, Peter W. & Andrzej Nowak. 2003. *Dynamical Grammar*. Oxford: Oxford University Press.
Culicover, Peter W., Ray Jackendoff & Jenny Audring. 2017. Multi-word constructions in the grammar. *Topics in Cognitive Science* 9. 1. doi:10.1111/tops.12255.
Dalrymple, Mary & Jamie Findlay. *This volume*. Lexical-functional grammar.
Dalrymple, Mary, Stuart M. Shieber & Fernando C. N. Pereira. 1991. Ellipsis and higher-order unification. *Linguistics and Philosophy* 14(4). 399–452.
Erteschik, Nomi. 1973. On the nature of island constraints. Doctoral dissertation, MIT.
Fillmore, Charles J., Paul Kay & Mary Catherine O'Connor. 1988. Regularity and idiomaticity in grammatical constructions. *Language* 64. 501–538.
Frank, Stefan L., Rens Bod & Morten H. Christiansen. 2012. How hierarchical is language use? *Proceedings of the Royal Society B: Biological Sciences* 297. 4522–4531.
Gibson, Edward. 1998. Linguistic complexity: Locality of syntactic dependencies. *Cognition* 68(1). 1–76.
Gibson, Edward. 2000. The dependency locality theory: a distance-based theory of linguistic complexity. In Alec Marantz, Yasushi Miyashita & Wayne O'Neil (eds.), *Image, language, brain*, 95–126. Cambridge: MIT Press.

Goldberg, Adele E. 1995. *Constructions: a construction grammar approach to argument structure*. Chicago: University of Chicago Press.

Goldberg, Adele E. 2006. *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.

Grodner, Daniel & Edward Gibson. 2005. Consequences of the serial nature of linguistic input for sentential complexity. *Cognitive Science* 29. 261–290.

Hardt, Daniel. 1992. VP ellipsis and semantic identity. In Chris Barker & David R. Dowty (eds.), *Proceedings of SALT 2*, 145–162. Cornell University, Ithaca, NY: CLC Publications.

Hardt, Daniel. 1993. Verb phrase ellipsis: Form, meaning and processing. Dissertation, University of Pennsylvania.

Hardt, Daniel. 1999. Dynamic interpretation of verb phrase ellipsis. *Linguistics and Philosophy* 22(2). 185–219.

Hardt, Daniel. 2004. Ellipsis resolution and inference. *Acta Linguistica Hafniensia* 36. 65–77.

Hardt, Daniel. 2008. VP ellipsis and constraints on interpretation. In Kyle Johnson (ed.), *Topics in ellipsis*, 15–29. Cambridge, United Kingdom: Cambridge University Press.

Hardt, Daniel & Maribel Romero. 2004. Ellipsis and the structure of discourse. *Journal of Semantics* 21(4). 375–414.

Hawkins, John A. 2004. *Efficiency and complexity in grammars*. Oxford: Oxford University Press.

Hawkins, John A. 2014. *Cross-linguistic variation and efficiency*. Oxford: Oxford University Press.

Hofmeister, Philip, Peter W. Culicover & Susanne Winkler. 2015. Effects of processing on the acceptability of frozen extraposed constituents. *Syntax* 18. 464–483.

Hofmeister, Philip, T. Florian Jaeger, Ivan A. Sag, Inbal Arnon & Neal Snider. 2007. Locality and accessibility in *wh*-questions. In Sam Featherston & Wolfgang Sternefeld (eds.), *Roots: Linguistics in search of its evidential base*, 185–206. Berlin: Mouton de Gruyter.

Hofmeister, Philip & Ivan A. Sag. 2010. Cognitive constraints and island effects. *Language* 86. 366–415.

Hofmeister, Philip, Laura Staum Casasanto & Ivan A. Sag. 2013. Islands in the grammar? Standards of evidence. In Jon Sprouse & Norbert Hornstein (eds.), *Experimental syntax and the islands debate*, 42–63. Cambridge: Cambridge University Press.

Jackendoff, Ray. 1972. *Semantic interpretation in generative grammar*. Cambridge, MA: MIT Press.

Jackendoff, Ray. 1990. *Semantic Structures*. Cambridge, MA: The MIT Press.

Jackendoff, Ray. 1997. *The Architecture of the Language Faculty*. Cambridge, MA: The MIT Press.

Jackendoff, Ray. 2002. *Foundations of language: Brain, meaning, grammar, evolution*. Oxford: Oxford University Press.

Jackendoff, Ray. 2007. *Language, Consciousness, Culture*. Cambridge, MA: MIT Press.

Jackendoff, Ray & Jenny Audring. *This volume*. The parallel architecture.

Jackendoff, Ray & Jenny Audring. Forthcoming. *The Texture of the Lexicon*. Oxford: Oxford University Press.

Jackendoff, Ray & Peter W. Culicover. 2003. The semantic basis of control in English. *Language* 79(3). 517–556.

James, William. 1890. *Principles of Psychology, vol. 1*, New York: Holt.

Kluender, Robert. 1991. *Cognitive constraints on variables in syntax*. La Jolla, CA: University of California, San Diego dissertation.

Kluender, Robert. 1992. Deriving islands constraints from principles of predication. In Helen Goodluck & Michael Rochemont (eds.), *Island constraints: Theory, acquisition and processing*, 223–258. Dordrecht: Kluwer.

Kluender, Robert. 1998. On the distinction between strong and weak islands: a processing perspective. In Peter W. Culicover & Louise McNally (eds.), *The Limits of Syntax* (Syntax and Semantics 29), 241–279. San Diego, CA: Academic Press.

Kluender, Robert & Marta Kutas. 1993a. Bridging the gap: evidence from ERPs on the processing of unbounded dependencies. *Journal of Cognitive Neuroscience* 5(2). 196–214.

Kluender, Robert & Marta Kutas. 1993b. Subjacency as a processing phenomenon. *Language and Cognitive Processes* 8(4). 573–633.

Konietzko, Andreas, Susanne Winkler & Peter W. Culicover. 2018. Heavy NP shift does not cause freezing. *Canadian Journal of Linguistics/Revue Canadienne de Linguistique*. 63(3). 454–464.

Kubota, Yusuke & Robert Levine. 2013. Coordination in hybrid type-logical categorial grammar. In *Ohio State University Working Papers in Linguistics*.

Kubota, Yusuke & Robert Levine. 2014a. Coordination as a tuple: A unified analysis of respective, symmetrical and summative predicates. Unpublished ms. Ohio State University.

Kubota, Yusuke & Robert Levine. 2014b. Pseudogapping as pseudo-VP ellipsis. In N. Asher & S. Soloviev (eds.), *Logical Aspects of Computational Linguistics. LACL 2014* (Lecture Notes in Computer Science 8535), 122–137. Berlin/Heidelberg: Springer.

Kubota, Yusuke & Robert Levine. 2014c. Unifying local and nonlocal modelling of respective and symmetrical predicates. In G. Morrill, R. Muskens, R. Osswald & F. Richter (eds.), *Formal Grammar. FG 2014* (Lecture Notes in Computer Science 8612), 104–120. Berlin/Heidelberg: Springer.

Kubota, Yusuke & Robert Levine. 2016. Gapping as hypothetical reasoning. *Natural Language & Linguistic Theory* 34(1). 107–156.

Kuno, Susumu & K.-I. Takami. 1993. *Grammar and Discourse Principles: Functional Syntax and GB Theory*. Chicago: University of Chicago Press.

Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar. Volume 1, Theoretical Prerequisites*. Stanford, California: Stanford University Press.

Langacker, Ronald W. 1991. *Foundations of Cognitive Grammar. Volume 2, Descriptive application*. Stanford, California: Stanford University Press.

Lappin, Shalom. 2005. A sequenced model of anaphora and ellipsis resolution. In *Anaphora Processing: Linguistic, Cognitive, and Computational Modelling*, 3–16. Amsterdam: John Benjamins.

Merchant, Jason. 2001. *The syntax of silence: sluicing, islands, and the theory of ellipsis*. Oxford: Oxford University Press.

Merchant, Jason. 2004. Fragments and ellipsis. *Linguistics and philosophy* 27(6). 661–738.

Miller, George & Noam Chomsky. 1963. Finitary models of language users. In R. Duncan Luce, Robert R. Bush & Eugene Galanter (eds.), *Handbook of mathematical psychology, vol. 2*, 419–492. New York: Wiley.

Müller, Stefan & Antonio Machicao y Priemer. *This volume*. Head-Driven Phrase Structure Grammar.

Perlmutter, David M. 1983. *Studies in relational grammar 1*. Chicago, Illinois: The University of Chicago Press.

Perlmutter, David M. & Carol Rosen. 1984. *Studies in relational grammar 2*. Chicago, Illinois: The University of Chicago Press.

Phillips, Colin. 2013. Some arguments and non-arguments for reductionist accounts of syntactic phenomena. *Language and Cognitive Processes* 28. 156–187.

Pollard, Carl & Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. Chicago, Illinois: University of Chicago Press and CSLI Publications.

Ross, John. 1967. Constraints on variables in syntax. Dissertation, Massachusetts Institute of Technology.

Ross, John Robert. 1969. Guess who. In Robert I. Binnick, Alice Davison, Georgia M. Green & Jerry L. Morgan (eds.), *Chicago linguistics society*, 252–286. Chicago, Illinois.

Sag, Ivan A. 1997. English relative clause constructions. *Journal of Linguistics* 33. 431–483.

Sag, Ivan A. 2012. Sign-based construction grammar – a synopsis. In Hans C. Boas & Ivan A. Sag (eds.), *Sign-based construction grammar*, 61–197. Stanford, CA: CSLI.
Sag, Ivan A., Philip Hofmeister & Neal Snider. 2007. Processing complexity in Subjacency violations: the complex noun phrase constraint. In *Proceedings of the 43 Annual Meeting of the Chicago Linguistic Society*, 215–229. Chicago: University of Chicago.
Sprouse, Jon & Norbert Hornstein. 2013. *Experimental syntax and island effects*. Cambridge: Cambridge University Press.
Van Valin, Robert D. Jr. 1998. The acquisition of wh-questions and the mechanisms of language acquisition. In Michael Tomasello (ed.), *The New Psychology of Language*, 221–249. Hillsdale, NJ: Erlbaum.
Wasow, Thomas. 1979. *Anaphora in generative grammar*. Dordrecht: John Benjamins Publishing Co.
Webber, Bonnie. 1978. A formal approach to discourse anaphora. Dissertation, Harvard University.
Wexler, Kenneth & Peter W. Culicover. 1980. *Formal principles of language acquisition*. Cambridge, MA: MIT Press.

Mary Dalrymple and Jamie Y. Findlay

5 Lexical Functional Grammar

Abstract: Lexical Functional Grammar is a linguistic theory which explores the various aspects of linguistic structure and how they are related. Like HPSG (Müller and Machicao y Priemer, 2018), Construction Grammar (Chaves, 2018), and the Parallel Architecture (Jackendoff and Audring, 2018), it is constraint-based and declarative, and does not assume that processes such as transformations are a part of linguistic theory. It was developed in the late 1970s by Joan Bresnan and Ron Kaplan as a theory of syntax, but has since been augmented by theories of other modules of linguistic structure and their relations to one another, including semantics and the syntax-semantics interface, argument structure, information structure, morphology, and prosody. These levels of structure are represented by separate grammatical modules which may be of very different formal character, connected to one another by means of functions relating parts of one structure to its corresponding parts in another structure.

Lexical Functional Grammar (LFG) is a declarative, constraint-based framework for analysing the various components of grammar, including, crucially, syntax. Although this chapter will introduce several parts of the formalism, it is not a comprehensive introduction to the theory, and the interested reader should consult one of a number of good book-length works which fill this role (Bresnan et al., 2016, and Falk, 2001, are textbooks, while Dalrymple, 2001 and Dalrymple et al., 2019 are reference works).

1 Data

LFG theory is built on a variety of types of linguistic evidence. In keeping with its origins in generative grammar, a common form of evidence is introspectively obtained judgements, either those of the linguist, or elicited judgements taken from others in more or less formalised (experimental) settings. However, no data type is ruled out as intrinsically irrelevant, and argumentation may make reference to experimental, corpus, or diachronic data. Furthermore, as in Construction Grammar (Chaves, 2018), no distinction is made between the "core" and the "periphery": the theory concerns itself with the analysis of the full range of constructions and phenomena of human language.

Acknowledgement: Thanks to Louise Mycock, Adam Przepiórkowski, Ron Kaplan, John Hewson, and the volume editors for helpful comments. Part of this work was conducted during the first author's fellowship at the Oslo Center for Advanced Study at the Norwegian Academy of Science and Letters.

https://doi.org/10.1515/9783110540253-005

In fact, LFG sprang in part from the dissatisfaction of Joan Bresnan and Ron Kaplan with what they saw as the “psychologically unrealistic” (Bresnan, 1978, 2) transformational theories developed by Chomsky and his colleagues during the 1950s, ’60s, and ’70s. In their efforts to build a mathematically well grounded linguistic theory that could underpin a psycholinguistically plausible model of human language processing, Kaplan and Bresnan adopt what they call the ‘Competence Hypothesis’:

> We assume that an explanatory model of human language performance will incorporate a theoretically justified representation of the native speaker’s linguistic knowledge (a *grammar*) as a component separate both from the computational mechanisms that operate on it (a *processor*) and from other nongrammatical processing parameters that might influence the processor’s behavior. To a certain extent the various components that we postulate can be studied independently, guided where appropriate by the well-established methods and evaluation standards of linguistics, computer science, and experimental psychology. However, the requirement that the various components ultimately must fit together in a consistent and coherent model imposes even stronger constraints on their structure and operation. (Kaplan and Bresnan, 1982, 173)

One of the aims of LFG is therefore to create a more psychologically plausible model of the grammar, one which takes seriously the role of processing in grammatical analysis. This is part of the motivation for LFG as a declarative/constraint-based framework.

Corpus data is also an important basis for theoretical claims in LFG. Indeed, corpora can serve as vital testing beds for evaluating computational implementations of the theory. This combination of theoretical and computational perspectives is crucial because large-scale computational grammars very quickly become far too complex to be assessed holistically by hand. Dyvik et al. (2009) and Patejuk and Przepiórkowski (2015) discuss large-scale annotated corpora and their use in the development and testing of computationally implemented LFG grammars.

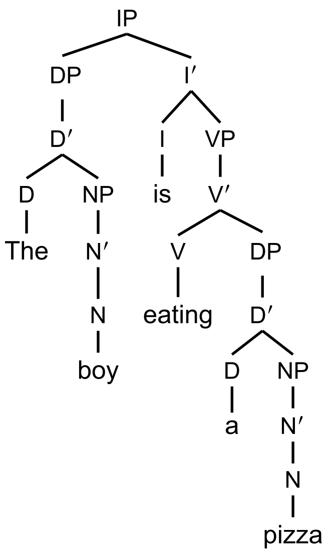
Theoretical claims are also made on the basis of historical data; Coppock and Wechsler (2010), for example, advocate an analysis of person agreement in the Uralic languages based on the historical path from incorporated pronoun to agreement marker. This makes use of work which describes grammaticalisation in terms of the loss of LFG features (Bresnan and Mchombo, 1987).

‘Performance’ data has also been important in the development of LFG-DOP, a combination of Data-Oriented Processing (Bod, 1992) with LFG. DOP models assume that “human language perception and production work with representations of concrete past language experiences, rather than with abstract grammatical rules” (Bod and Kaplan, 1998, page 145). In general, DOP works by taking a corpus of linguistic representations and decomposing them into fragments, which are then recomposed in the analysis of new utterances. The LFG-DOP model specialises the theory to a corpus of LFG syntactic representations. Although not part of ‘mainstream’ LFG work, LFG-DOP shows the possibility of combining LFG with other approaches, a theme which will reoccur in our discussion.

2 Goals

A fundamental assumption of LFG is that the language faculty is made up of multiple, inter-dependent modules, which exist in parallel and are mutually constraining. Language is not a unitary object, and the best way to describe and explain properties of, say, phonology, will not necessarily be the same as to explain syntax. This much is perhaps relatively uncontroversial (although it is at odds with the 'syntacto-centrism' of much generative linguistics, which sees other components, especially semantics, as ultimately parasitic on a syntactic level of representation—cf. the notion of Logical Form as a level of syntax: May 1985, Hornstein 1995). Even within syntax, though, we are not dealing with a single set of formally equivalent phenomena: the term 'syntax' is used to describe both superficial phenomena such as word order, as well as more abstract phenomena such as subjecthood. LFG therefore proposes to separate syntactic representation into the two levels of *constituent structure* (c-structure) and *functional structure* (f-structure), the former using a familiar phrase structure tree to represent linear order as well as hierarchical structure and constituency, the latter using a feature structure (also known as an attribute-value matrix) to represent abstract relational information about grammatical functions, binding, long-distance dependencies, etc.

(1) Constituent structure: Functional structure:



$$\begin{bmatrix} \text{PRED} & \text{`eat}\langle\text{SUBJ, OBJ}\rangle\text{'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{`boy'} \\ \text{DEF} & + \\ \text{PERS} & 3 \\ \text{NUM} & \text{SG} \end{bmatrix} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{`pizza'} \\ \text{DEF} & - \\ \text{PERS} & 3 \\ \text{NUM} & \text{SG} \end{bmatrix} \\ \text{TENSE} & \text{PRS} \\ \text{ASPECT} & \text{PROG} \end{bmatrix}$$

Bresnan sums up the view in this quotation:

> Semantic argument structure, constituent structure and functional structure are parallel information structures of very different formal character. They are related not by syntactic derivation, but by structural correspondences, as a melody is related to the words of a song. Semantic, structural and functional representations of a sentence can be superimposed, but they are independent planes of grammatical organisation. (Bresnan, 1993, 45)

The central challenge for this kind of approach is determining the modules relevant for linguistic analysis, their internal structure and constraints, and the relations between them. This is a large part of what modern work in LFG involves. However, in general the two syntactic modules, c- and f-structure, have remained central to LFG theorising since the beginning. Since the topic of the present volume is syntax, this is where our focus will lie as well.

2.1 Well-formedness criteria

LFG is, true to its generative roots, interested in describing linguistic competence as opposed to performance: that is, the knowledge that one possesses in knowing a language as opposed to what is required in order to deploy that knowledge in production or comprehension. For this reason, LFG as a grammatical theory does not encompass performance factors (although LFG-based theories of performance can be formulated, such as LFG-DOP, mentioned in Section 1), and so what are taken as data for analysis tend to be 'cleaned up', abstracting away from various performance 'noise' factors such as hesitations, repetitions, speech errors, etc.

In keeping with the usual approach in generative syntax, traditional LFG work treats well-formedness as categorical. That is, sentences (or, rather, linguistic descriptions) are either a part of the grammar or are not. There is no notion that some grammatical violations are 'better' or 'worse' than others. For example, a simple failure of agreement like (2) is just as ungrammatical as utter nonsense such as (3):

(2) * Kim see the painting.

(3) * Dog flarb the on ktaw.

Of course, there is nothing preventing us from devising other metrics, such as how many constraints a description violates, which can give us a derivative notion of grammatical gradience.

One strand of work which has sought to add a notion of gradience to the LFG understanding of well-formedness is Optimality-Theoretic LFG (OT-LFG: Bresnan, 2000, 2002; Kuhn, 2001), a variant of OT syntax where the output of the GEN component consists of pairs of c-structures and f-structures. In OT-LFG, in keeping with the general principles of OT (Prince and Smolensky, 2004; Legendre, 2018), the grammar consists of a set of possibly incompatible, violable constraints, where a linguistic description need not satisfy all of the constraints in order to be well-formed, but must merely be the 'least bad' candidate description. Such a system allows for a much more fine-grained analysis of well-formedness. For example, it makes it possible to describe levels of *un*grammaticality: a sub-optimal candidate can still be ranked above other sub-optimal candidates, by violating fewer highly-ranked constraints, and can therefore be 'less ungrammatical' in a well-defined sense. This can explain the reported observations that speakers are sensitive to distinctions even among clearly ungrammatical examples (Featherston, 2008).

2.2 Relations to other grammatical modules

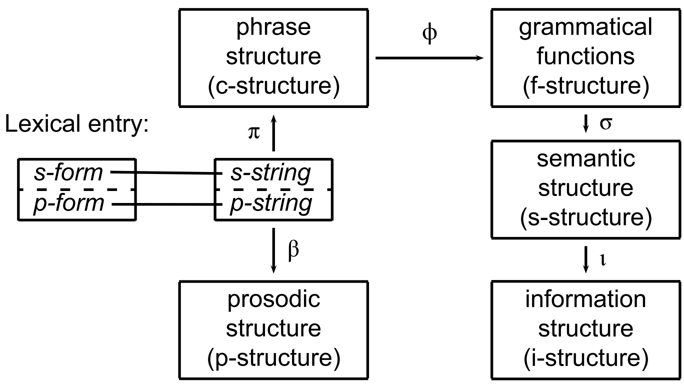
Given the central LFG assumption that language is modular and composed of multiple, internally complex subsystems, syntax is not, in a theoretical sense, privileged. In a practical sense, however, most work in LFG has been of a syntactic nature, and the framework did start life primarily as a model of syntax.

However, as we have seen, 'syntax' is not a single component of the grammar, but is rather sub-divided into two modules: c-structure, which represents word order and phrasal grouping, and f-structure, which represents grammatical functions and features. Kaplan (1987) proposed to generalise this view and extend it to other components of the grammar, such as semantics or phonology: different grammatical modules have their own internal structure and organisational principles, and are related via structural correspondences relating components of one module to components of other modules. Asudeh (2006) describes this *correspondence architecture* in more detail.

This means that the grammar as a whole—syntax, semantics, information structure, prosodic structure, etc.—is a "nearly decomposable system" in the sense of Simon (1962), where the internal organisation and behaviour of each component is largely independent, but relations among the components can be defined so that units of one component are related via correspondence to units of another. For example, just as the c-structure nodes of the subject of a sentence are related to the subject f-structure, in the same way an f-structure is related to the semantic structure corresponding to its meaning.

Current LFG work builds on this view, with an exploration of syntax (in its dual nature, c-structure vs. f-structure) and its relation to semantics, information structure, and phonology, and also defining the place of morphology in the overall architecture. A schematic version of the full correspondence architecture is given in (4):[1]

(4)



1 Some scholars assume an independent level of argument structure between c- and f-structure (Butt et al., 1997), but recent work has represented this information at semantic structure instead (Asudeh and Giorgolo, 2012; Findlay, 2016).

The morphological component generates lexical entries (Sadler and Spencer, 2001; Kaplan and Butt, 2002; Dalrymple, 2015). As shown in (4), a lexical entry encodes constraints on an s-form (the terminal node of the phrase structure tree and the grammatical and semantic constraints associated with it) and a p-form (which forms the basis of prosodic analysis). Thus, the lexicon is the locus of the relation between syntax/semantics and phonetics/phonology (Mycock and Lowe, 2013; Dalrymple et al., 2019). This allows the analysis of strings of words (sentences), which are connected to both a prosodic structure (Dalrymple and Mycock, 2011; Mycock and Lowe, 2013) and a syntactic constituent (c-)structure. Functional structure is related to c-structure via a function called ϕ (on which see Section 3.3), and f-structure is in turn related to semantic structure (Dalrymple, 1999; Asudeh, 2012) and information structure (Dalrymple and Nikolaeva, 2011; Dalrymple et al., 2019).

Owing to its modularity, LFG as a syntactic theory is agnostic about which particular theory one adopts for any of the other levels of structure (phonetics and phonology, semantics, morphology, etc.). In general, work on phonetics and phonology *per se* has been limited. Work on prosody has most often assumed an independent hierarchical prosodic structure, governed by the Prosodic Hierarchy of Prosodic Phonology (Selkirk, 1981; Nespor and Vogel, 1986).

Work on semantics is much more developed. The most common theory of the syntax-semantics interface in LFG is Glue Semantics (Dalrymple, 1999; Asudeh, 2012). In Glue, meaning composition is treated as deduction in a resource logic: the meaning of a sentence is assembled from the meaning of its parts via logical deduction. LFG+Glue remains agnostic about what particular 'meaning language' is used to actually express the natural language meanings themselves. Practitioners commonly use some variety of predicate calculus, but there is work which uses Discourse Representation Theory (van Genabith and Crouch, 1999; Bary and Haug, 2011; Lowe, 2012), Natural Semantic Metalanguage (Goddard and Wierzbicka, 2002; Andrews, 2006), and others.

2.3 Domain of analysis

The main focus of LFG, like other theories in the generative tradition, is the sentence level. However, work on larger domains is not excluded: for example, King and Zaenen (2004) and Dalrymple et al. (2017) offer analyses of discourse structure and intersentential anaphora within the LFG framework. Giorgolo and Asudeh (2011) also give a proposal for integrating gesture into the correspondence architecture, thus extending the coverage of LFG to other modalities beyond speech.

As well as larger, discourse-level issues, however, one must not neglect the importance of sub-sentential units. One of the strengths claimed for LFG from the start, as for other constraint-based,non-transformational theories including HPSG (Müller and Machicao y Priemer, 2018), is its ability to give an account of fragments and partial sentences (Kaplan and Bresnan, 1982), which are vitally important in terms of understanding acquisition and processing. Such a property emerges in LFG largely thanks to the local character of most constraints, unlike in Minimalism (Hornstein, 2018), for example, where features may not be checked until much higher up in the tree than they are introduced.

2.4 Language domain

Descriptions of individual languages' grammars are intended to be full and accurate accounts of a single, synchronic state of the language. However, when it comes to deeper questions of theory, cross-linguistic data is invaluable, as well as data about the historical evolution of languages.

LFG takes grammatical functions to be theoretical primitives, and assumes that the stock of such functions forms the basic vocabulary of functional structure. Cross-linguistic data has been crucial in reaching this conclusion: for example, Kroeger (1993) demonstrates the importance of the role of subject even in languages like Tagalog which have been argued not to make use of it (Schachter, 1976).[2]

Although most LFG work is synchronically oriented, an important strand of work relies on LFG assumptions in explanatory accounts of historical change, taking advantage of LFG's separation between phrasal structure and abstract functional structure. For example, Allen (1995, 2008), Bresnan and Mchombo (1987), Börjars et al. (1997) and Vincent (1999) discuss how grammaticalisation can be understood in terms of the loss of f-structure information.

The less restrictive LFG conception of phrase structure has also been important in historical work: Vincent (1999), Börjars et al. (2016) and Börjars and Vincent (2016), for example, take advantage of LFG's ability to incorporate not only X′-theoretic categories but also non-projecting categories and exocentric categories in the phrase structure in their analysis of the introduction and development of closed-class words such as prepositions and determiners, which may start out as non-X′-theoretic words, but which later evolve to full X′-theoretic status, appearing as the head of a phrase with specifiers and complements. For further examples of LFG-based diachronic work, see the papers collected in Butt and King (2001).

2.5 Applications

LFG as a linguistic theory is not specifically aimed at solving practical problems, but it has been used as the basis or grammar component of several systems and theories which have a more practical application.

2 See Falk (2006) for a contrasting view on the primacy of SUBJ as a theoretical primitive.

Computational implementation has been an important aspect of LFG research from the start. One of the earliest LFG implementations was the Grammar Writer's Workbench (Kaplan and Maxwell, 1996), originally implemented in the early 1980s. In 1993, the team at the Xerox Palo Alto Research Center (Xerox PARC) began work on a new implementation, which ultimately became the Xerox Linguistic Environment (XLE: Crouch et al. 2008; Maxwell 2015).

Within XLE, a major focus has been cross-linguistic application. The Parallel Grammar Project (PARGRAM: Butt et al. 1999, Butt et al. 2002) is a prime example of this. The project started in 1994 with grammars of English, French, and German, and later grew to include grammars of Norwegian, Japanese, Urdu, Turkish, Hungarian, Georgian, Tigrinya, Wolof, Indonesian, Welsh, Malagasy, Mandarin Chinese, Arabic, Vietnamese, Polish, and Northern Sotho.

In general, the availability of computational implementations of LFG, and in particular the XLE, has led to productive research on combining LFG grammars with other computational tools to increase parsing efficiency, improve parsing results, or produce more useful language-based applications.

A different kind of practical application for LFG is found in the domain of second language acquisition. Processability Theory (Pienemann, 1998, 2005; Bettoni and Di Biase, 2015) takes LFG as its formal model, and makes crucial use of the division among grammatical modules assumed by LFG in its treatment of second language acquisition.

3 Tools

LFG is a formalised framework, and this means that all levels of representation have well-defined mathematical properties, and are subject to explicit well-formedness conditions. In this section, we introduce the two syntactic levels assumed in all LFG work, constituent structure (c-structure) and functional structure (f-structure).

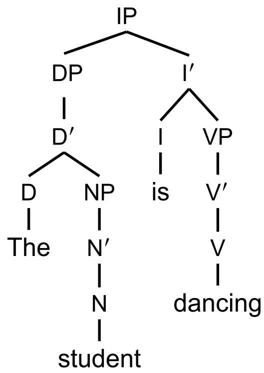
3.1 C-structure

LFG departs from Chomskyan approaches in seeing the locus of cross-linguistic similarity not as c-structure but as f-structure (for more on which, see Section 3.4 below), taking traditional grammatical terms such as *subject* and *object* not as derivable from constituent structure, but rather as theoretical primitives in their own right. This means that no constituent structure is privileged as underlying or 'deep' in any sense: surface word orders are purely a matter of language-specific phrase structure rules.

For this reason, when it comes to describing the c-structure of a language, the focus is on analyzing just those phenomena which are best represented explicitly by phrase structure trees, notably word order and hierarchical structure, but also syntactic category, since we assume that the nodes in such a tree are labelled. Constituency relations are also expressed here, given that constituency can be understood as exhaustive dominance.

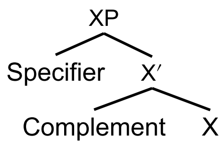
An example of a c-structure is given in (5):

(5)



Formally, c-structures are described via a simple context-free grammar, where the phrase structure rules are understood not as rewrite rules but as *node admissibility conditions* (McCawley, 1968), which describe the set of trees admitted by the theory. LFG subscribes loosely to *X′ theory* (Jackendoff, 1977), so that there are three levels of projection described with respect to the head of a phrase, X: X^0 (or simply X; the head itself), XP (the maximal, phrasal projection), and X′ (all intermediate levels). This provides the following schematic positions in relation to X (ignoring linear order):

(6)



In configurational languages like English, phrases filling these positions have particular functional roles (as suggested by their names), and this is no doubt in part why some theories conflate the abstract properties of syntax with the overtly manifested ones. For example, such theories might say that all subjects appear in the specifier of IP. However, such a correlation between position and function is far from necessary, and in reality only represents one end of a spectrum of possibilities in which "morphology competes with syntax" (Bresnan, 1998). That is, in a language such as Latin or Warlpiri, where relational information is indicated by case, word order tends to be governed by other communicative concerns (such as information structure), and thus is a poor guide to grammatical relations. In transformationalist theories, this leads to the claim that there is a confluence at some other point in a derivation—perhaps there is a 'deep' or underlying structure where grammatical relations are manifested configurationally, or perhaps such correspondences do not emerge until later on in the derivation. In contrast, LFG simply assumes that such functional information is best represented at a different level of structure, since it is not inherently configurational.

Although we do not assume that functional information is defined configurationally, there do undoubtedly exist correlations between certain phrase structure positions and certain grammatical functions (Bresnan et al., 2016). Nevertheless, languages vary widely in the kinds of syntactic configurations they permit, and thus LFG allows them to vary relatively freely in the kinds of phrase structure rules they contain, and thereby the kinds of trees they admit. C-structure constraints are thus much less stringent in LFG than in some other formalisms; LFG adheres only loosely to X′ theory.

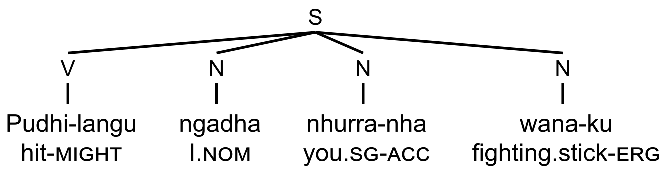
Similarly, we do not require the presence of empty categories in the c-structure tree. For example, while LFG is compatible with a theory of long-distance dependencies that posits traces (Kaplan and Bresnan, 1982; Bresnan, 1995, 1998), they are not required (Kaplan et al., 1987; Kaplan and Zaenen, 1989; Dalrymple et al., 2007), nor are other unpronounced elements such as *pro*/*PRO*. Instead, such elements are represented at f-structure when required.

What is more, where there is no strong evidence for the existence of a particular hierarchical structure, we do not feel constrained to nonetheless posit it. For example, in 'free word order' languages, LFG makes use of the exocentric category S, permitting a flat structure when there is no evidence for a more articulated one:

(7) Tharrkari (Austin and Bresnan, 1996, 248):

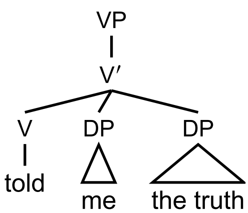
Pudhi-langu ngadha nhurra-nha wana-ku.
hit-MIGHT I.NOM you.SG-ACC fighting.stick-ERG

'I might hit you with a fighting stick'



In addition, where a verb has more than one complement, trees do not have to be binary branching:

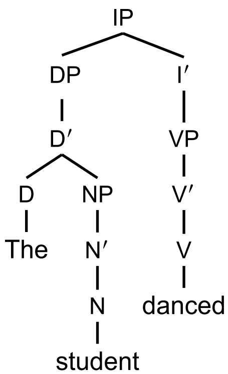
(8)



This avoids the requirements of heavily articulated trees with unpronounced functional elements which a theory restricted to purely binary branching necessitates (Kayne, 1984). C-structure represents words and their grouping into constituents, not abstract functional features and relations.

Another way in which LFG c-structures differ from familiar X′-theoretic trees is in the optionality of all positions. That is, there can be phrases without heads (and, in some versions of the theory, heads without phrases—the so-called non-projecting categories of Toivonen, 2003). For example, in an English sentence without an auxiliary verb to occupy I, we do not need to posit an unpronounced feature appearing in the tree which expresses tense, nor do we need to separate the main verb so that its tense ending appears at I and ‘lowers’ to the main verb, since in such cases tense is expressed by the finite verb in V. Instead, we have the structure in (9), where the head simply does not appear:

(9)



Finally, some practitioners subscribe to what is called the principle of *Economy of Expression* (Bresnan et al., 2016, 89–94), whereby all nodes are omitted unless motivated by some functional requirement. Under this approach, in (9) there would be no N′ and no V′, since there is no evidence for their existence. Where the structural configuration provides important information, though, the node remains: thus, the I′ would remain in our example because the fact that the DP *the student* appears as its sister is what tells us, in English, that this DP is the subject (for more on which see Section 3.3). For further discussion of Economy of Expression, see Dalrymple et al. (2015).

The set of syntactic categories generally assumed is fairly conservative. Almost all LFG practitioners agree on the following inventory of categories and their projections, along with the exocentric category S:

(10) Lexical categories: N, V, P, Adj, Adv
Functional categories: C, I, D

Other categories are used occasionally, such as K for case (Butt and King, 2004), but the extended functional projections of other theories are not appealed to, since functional information is represented separately, in f-structure.

3.2 F-structure

The point just made once again illustrates the separation of levels held to be central in LFG: although different levels constrain and impact on each other, so that, for example, in configurational languages, word order and grammatical functions are correlated, nevertheless, *internally*, each level represents only the material which is pertinent to that level. Thus, inherently ordered, configurational properties such as word order and constituency are represented by a formal object that expresses such properties, namely the c-structure tree. More abstract, and inherently unordered, properties like grammatical relations and other functional phenomena are expressed elsewhere, namely at f-structure, which has different properties more conducive to this task. In essence, LFG does not force the data to fit the properties of a particular mathematical object, but rather chooses data structures on the basis of how well their properties fit with the observed phenomena.

An f-structure for sentence (5) is given in (11):

(11) *The student is dancing.*

$$\begin{bmatrix} \text{PRED} & \text{'dance}\langle\text{SUBJ}\rangle\text{'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'student'} \\ \text{DEF} & + \\ \text{NUM} & \text{SG} \\ \text{PERS} & 3 \end{bmatrix} \\ \text{TENSE} & \text{PRS} \\ \text{ASPECT} & \text{PROG} \end{bmatrix}$$

F-structures are attribute-value matrices, which formally are understood as functions from their attributes (the left hand column, e. g. TENSE) to their values (the right hand column, e. g. PRS). Given a set-theoretic understanding of the notion of function, this means that f-structures are sets of pairs, the left-hand member of each pair being an attribute, and the right-hand member being a value (which can itself be an f-structure, as in the value of SUBJ in example 11).[3] Since sets are unordered, this means that the ordering of attribute-value pairs with respect to one another is irrelevant; in other words, the f-structure in (12) is identical to the f-structure in (11):

3 HPSG signs appear similar to LFG f-structures, since both are represented as attribute-value matrices. They are formally quite different, however: one difference is that HPSG signs are typed, and another is that HPSG incorporates a type-token distinction which is precluded in the set-theoretic setting of LFG f-structures. See Müller and Machicao y Priemer (2018) for more discussion.

(12) $$\begin{bmatrix} \text{TENSE} & \text{PRS} \\ \text{SUBJ} & \begin{bmatrix} \text{DEF} & + \\ \text{PERS} & 3 \\ \text{PRED} & \text{‘student’} \\ \text{NUM} & \text{SG} \end{bmatrix} \\ \text{ASPECT} & \text{PROG} \\ \text{PRED} & \text{‘dance}\langle\text{SUBJ}\rangle\text{’} \end{bmatrix}$$

Thus, we are not forced to impose an ordering on properties where order seems not to matter: the fact that *the student* is the subject of the sentence is not in any sense ‘prior’ to the fact that the sentence is present tense; nor, in a transitive sentence like *the student read the book*, is the fact that *the student* is the subject in any sense prior to the fact that *the book* is the object. Of course in a language like English where subjects (generally) linearly precede objects, there does seem to be a sense in which this relation of priority holds. But this is really a fact about the c-structure encodings of f-structure relationships. We wish to say that the exact same relations of subjecthood and objecthood obtain in languages where word order is different: that is, a subject is just as much of a subject in a VOS language like Malagasy as it is in an SVO language like English, the only difference being in how it is realised configurationally. In this sense, there is no ordering between different pieces of functional information, and so an unordered data structure such as the attribute-value matrix of f-structure is a more appropriate way of modelling these facts than a phrase-structure tree which is inherently ordered.

Some of the most important attributes used in LFG are the *grammatical functions* (GFs). These are abstract categories used to characterise the relations between different elements of a sentence, many of which are familiar from traditional grammars. They are taken to be theoretical primitives, that is, they are not derivable from some other property like position in the tree.[4] A list of the most frequently assumed grammatical functions is given in Table 1 (see also Dalrymple, 2001, 8–28, Asudeh and Toivonen, 2015). It is conventional to define the abbreviation GF as a meta-category which represents a disjunction over all the possible attributes listed in Table 1, which allows for easy reference to the full set of grammatical functions:

(13) $\text{GF} \equiv \{\text{SUBJ} \mid \text{OBJ} \mid \text{OBJ}_\theta \mid \text{OBL}_\theta \mid \text{COMP} \mid \text{XCOMP} \mid \text{ADJ} \mid \text{XADJ}\}$

Apart from the GFs, functional structure contains other attributes such as NUMber or PERSon, which are proposed on an empirical basis and motivated by f-structurally defined syntactic patterns and processes such as agreement. Agreement is handled

4 In certain strands of LFG research on argument structure and diathesis alternations (e. g. Bresnan and Kanerva, 1989; Butt, 1995; Kibort, 2004), the GFs are further decomposed into the features [*r*] (for ‘semantically *r*estricted’) and [*o*] (for ‘*o*bjective’).

Table 1: Grammatical functions in LFG.

LFG abbreviation	Grammatical function
SUBJ	Subject
OBJ	Object
OBJ_θ	Restricted/secondary object (indexed by thematic role)
OBL_θ	Oblique (indexed by thematic role)
COMP	Closed sentential complement
XCOMP	Open sentential complement
ADJ	Adjunct
XADJ	External (open) adjunct

at f-structure rather than c-structure because it makes reference to f-structure properties, such as subject and object, rather than c-structure ones, such as 'leftmost phrase'. That is, 'subject agreement', making reference to functional information, is common, but 'sentence-initial NP agreement' or similar, making reference to configurational information, is not.

Other features such as DEFiniteness are commonly assumed to be represented at f-structure, but it is a matter of ongoing debate to what extent such properties have purely syntactic effects, outside of e. g. semantic ones, and thus to what extent they deserve to be encoded at a level of syntactic representation like f-structure at all.

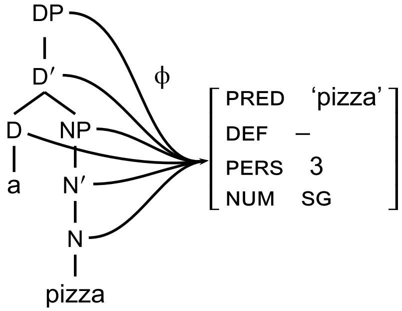
Because f-structure is where LFG represents abstract syntactic information, it is also where a lot of the heavy lifting of syntactic theory occurs; as we will see in Section 3.3, it is where, for example, agreement, 'raising', and long-distance dependencies are represented. An important corollary of this is the impact which f-structure has on the computational complexity of LFG. Although c-structure is only context-free, the presence of f-structure pushes grammars in the LFG framework into the context-sensitive space (the recognition task for LFG languages is, in the worst case, likely NP-complete: see Berwick, 1982).[5]

3.3 Connecting c-structure and f-structure

C-structure and f-structure are related by a function ϕ, from c-structure nodes to f-structures. ϕ is a potentially many-to-one correspondence. That is, more than one

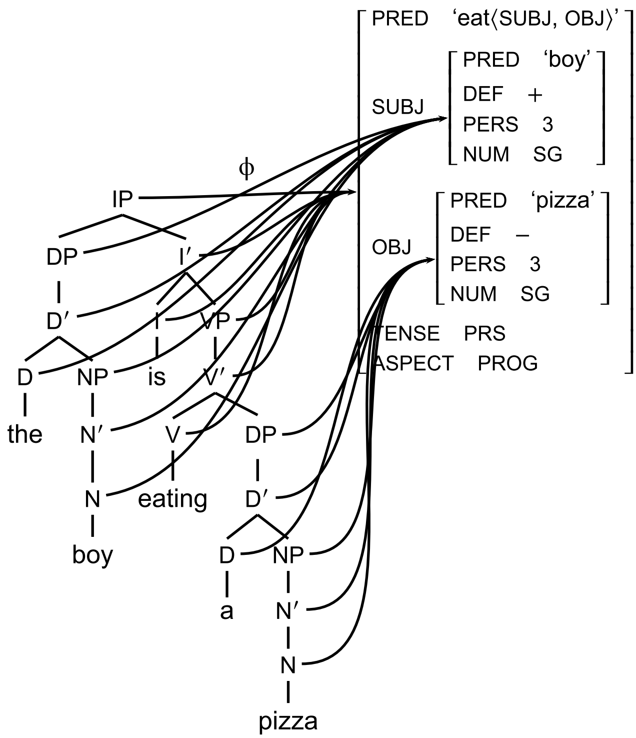
5 This is true of the formalism itself, which can be used to encode grammars with properties that are not required for describing human languages. LFG grammars for natural languages may be more tractable if they are constrained in accordance with linguistic data and principles. For example, Wedekind and Kaplan (2019) characterize a "k-bounded" proper subclass of LFG grammars for which the number of c-structure nodes that map to a given f-structure for any derivation of any sentence is guaranteed to be less than a grammar dependent constant k. An LFG grammar with this property is only mildly context-sensitive, which is the level of expressive power generally considered necessary for the description of natural language syntax (Joshi, 1985).

c-structure node can correspond to the same f-structure:

(14)


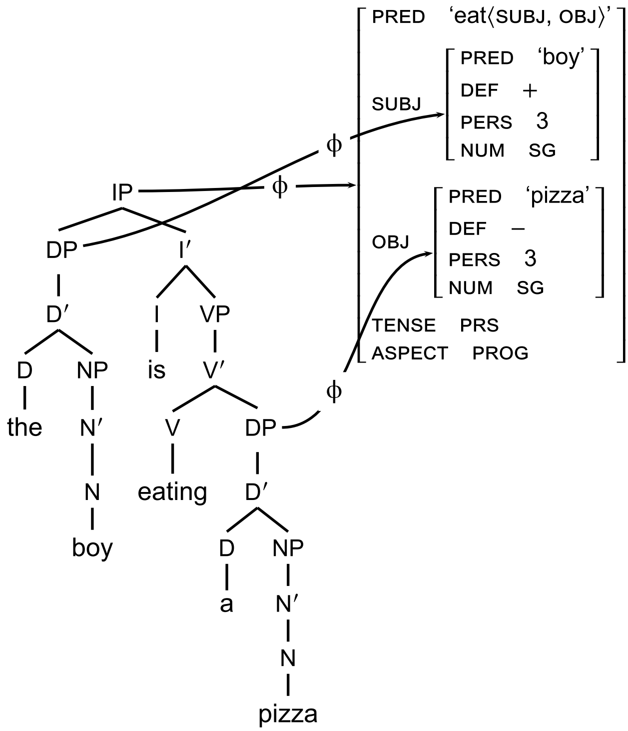
As shown in (14), the DP, D′, and D nodes in each DP correspond to a single f-structure, and the projections of N also correspond to this same f-structure (we say that the NP is an *f-structure co-head* of the DP).

Similarly, in the whole sentence, all of the verbal projections, *viz*. IP, I′, I, VP, V′, and V, correspond to the outer f-structure:

(15)


However, this quickly becomes difficult to read, and so usually we will present diagrams as in (16), where only the maximal projections have their ϕ correspondence indicated:

(16)

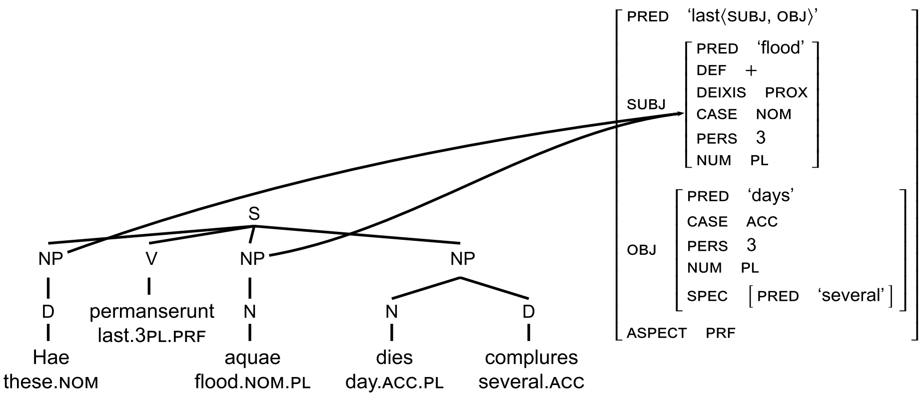


The relation between c-structure and f-structure is constrained by the function ϕ: each c-structure node can correspond at most to a single f-structure. Thus, although structures in LFG are taken to be simultaneously present and mutually constraining—c-structures do not ‘generate’ f-structures in any sense—it is nonetheless true that there is a certain directionality encoded in the fact that such correspondences are functional. It is an empirical claim that c-structure projects f-structure, for example, rather than vice versa, based on the fact that more than one separate c-structure constituent can correspond to the same grammatical function. In the Latin example in (17), the two parts of *hae...aquae* are both part of the SUBJ, even though they appear discontinuously at c-structure, with the determiner nonadjacent to the noun.

(17) From Snijders (2015), citing Caes. Civ. 1.50.1, via Spevak (2010, 24):

Hae	permanserunt	aquae	dies	complures.
these.NOM	last.3PL.PERF	flood.NOM.PL	day.ACC.PL	several.ACC

‘These floods lasted for several days.’



This means that more than one c-structure node can correspond to the same f-structure, but we do not have evidence of the opposite: say, a single word giving rise to two different f-structures serving as both subject and object to the same predicate. Since the correspondence relations are functions, this motivates a grammatical architecture where c-structure is mapped to f-structure, and not vice versa.

To determine the mapping which ϕ describes, we annotate c-structure nodes with equations describing the relation they bear to f-structure; the f-structure which corresponds to a given c-structure is then the smallest f-structure which satisfies all of the equations. In writing these equations, we make use of the following conventions:

(18) Variables over c-structure nodes:
  a. $*$ refers to the current node (the node hosting the annotation).
  b. $\hat{*}$ refers to the mother of the current node.

(19) Meta-variables over f-structures:
  a. $\downarrow \equiv \phi(*)$
  b. $\uparrow \equiv \phi(\hat{*})$

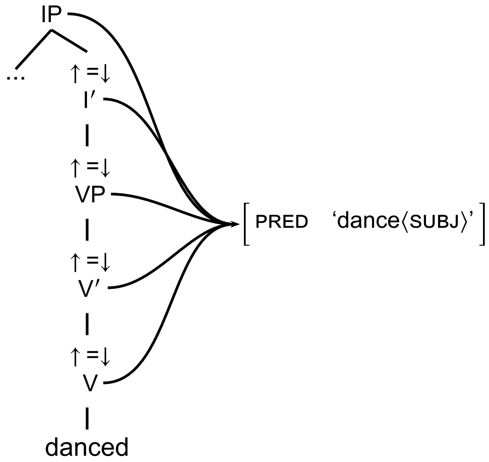
The asterisks refer to c-structure nodes, and the arrows refer to f-structures. In less formal terms, $\downarrow$ refers to 'my f-structure', the f-structure that corresponds (via ϕ) to the node bearing this annotation; $\uparrow$ refers to 'my mother's f-structure', the f-structure that corresponds to the mother of the node bearing this annotation. The simplest equation we can write, then, is (20):

(20) $\uparrow = \downarrow$

This says that my mother corresponds to the same f-structure as myself; this is what we use to pass information up along a non-branching c-structure, for example:[6]

6 Annotations are written above c-structure nodes, with the intention of giving the meta-variables a kind of iconicity: the arrows point to the nodes whose f-structures they stand for. Nevertheless, many

(21)



By convention, nodes that are unannotated are assumed to be annotated with the $\uparrow = \downarrow$ equation, which enables us to reduce clutter in c-structures.

Aside from simple equality of f-structures, we can also say things about the values of particular attributes; for example that the value of the SUBJ attribute is to be found at the f-structure of the current node:

(22) $(\uparrow \text{SUBJ}) = \downarrow$

This equation says that the f-structure corresponding to the node that bears it ($\downarrow$) is the value of the SUBJ attribute of the f-structure corresponding to its mother's node ($\uparrow$). It would be used, for instance, to annotate the phrase structure rule corresponding to the specifier of IP in English, which is where subjects often appear:

(23) $$\text{IP} \quad \rightarrow \quad \underset{\text{DP}}{(\uparrow \text{SUBJ}) = \downarrow} \quad \underset{\text{I}'}{\uparrow = \downarrow}$$

Such equations can also be contributed by lexical entries, and this is where idiosyncratic information such as the value of PRED attributes is encoded. Lexical entries in LFG contribute three things: a word form, a category, and a functional description, which is a set of equations describing f-structure. An example is given in (24) for past tense *danced*:

(24) *danced* V $(\uparrow \text{PRED}) = \text{'dance}\langle\text{SUBJ}\rangle\text{'}$
$(\uparrow \text{TENSE}) = \text{PST}$

An LFG grammar consists of a set of annotated phrase structure rules, describing the c-structure of a language and its relationship to f-structure, and a set of lexical entries

researchers write such annotations underneath the c-structure nodes instead; especially when multiple annotations are required, this can aid readability.

(which can in fact be thought of as phrase structure rules which simply expand preterminal category symbols).

Now that we have some more tools in place, we are in a position to explain how various syntactic phenomena are analysed in LFG. We start with agreement. Agreement is achieved via multiple specification: the source and target of agreement both specify values for the same features, with the result that these specifications must agree or else the structures will be ruled out as illicit by the functional nature of f-structure (which means that each attribute can only have one value). For example, a singular noun like *Alex* will specify that its number is singular and that it is third person. A plural noun like *caterpillars*, on the other hand, will specify that it is plural and third person. The third person singular verb form in English, e. g. *sings*, meanwhile specifies that its subject is third person and singular. If *Alex* is its subject, the specifications from the two items can combine successfully into a single f-structure, and thus *Alex sings* is grammatical. If it is combined similarly with *caterpillars*, though, there will be a clash when it comes to the NUMber attribute, since this feature is simultaneously stipulated to be singular (by the verb) and plural (by the noun). Thus, **Caterpillars sings* is not grammatical.

The (simplified) lexical entries for these three words are given below:

(25) *sings* V (↑ PRED) = 'sing⟨SUBJ⟩'
(↑ SUBJ NUM) = SG
(↑ SUBJ PERS) = 3

(26) *Alex* N (↑ PRED) = 'Alex'
(↑ NUM) = SG
(↑ PERS) = 3

(27) *caterpillars* N (↑ PRED) = 'caterpillar'
(↑ NUM) = PL
(↑ PERS) = 3

As we can see, when *Alex* is the subject of *sings*, the assignment of a value to the subject's NUM attribute proceeds without any problem, since both the verb and the noun specify SG as the value of that feature. When *caterpillars* is the subject, though, there will be a clash, since the noun specifies PL, while the verb calls for SG.

'Raising' and long-distance dependencies are represented via f-structure sharing: the same f-structure can be the value of multiple attributes. We are not merely talking about type identity here, where there are two distinct f-structures of the same form, but rather token identity, where the f-structures are one and the same. For this reason, such structure sharing is often represented using a line to connect one instance of the f-structure with its other positions—in this way, each unique f-structure is only ever represented on the page once.

The f-structure for *Alex seemed to yawn* is given in (28) (recall that XCOMP is the GF assigned to open complement clauses):[7]

(28)
$$\begin{bmatrix} \text{PRED} & \text{'seem}\langle\text{XCOMP}\rangle\text{SUBJ'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'Alex'} \end{bmatrix} \\ \text{XCOMP} & \begin{bmatrix} \text{PRED} & \text{'yawn}\langle\text{SUBJ}\rangle\text{'} \\ \text{SUBJ} & \end{bmatrix} \\ \text{TENSE} & \text{PST} \end{bmatrix}$$

The parallel with the transformational raising analysis is clear: rather than saying that the subject of the lower clause has moved and now appears in the higher clause, the structure sharing analysis simply says that the two positions are identified with one another. In this case, the sharing is via lexical specification: it is a property of the verb *seem* that its subject is identified with the subject of its complement clause.

Long-distance dependencies are also handled at f-structure, without the need for traces or other empty nodes at c-structure. The f-structure for *What did Kim say you stole?* is given in (29):[8]

(29)
$$\begin{bmatrix} \text{DIS} & \begin{bmatrix} \text{PRED} & \text{'pro'} \\ \text{PRONTYPE} & \text{WH} \end{bmatrix} \\ \text{PRED} & \text{'say}\langle\text{SUBJ, COMP}\rangle\text{'} \\ \text{SUBJ} & [\text{"Kim"}] \\ \text{COMP} & \begin{bmatrix} \text{PRED} & \text{'steal}\langle\text{SUBJ, OBJ}\rangle\text{'} \\ \text{SUBJ} & [\text{"you"}] \\ \text{OBJ} & \\ \text{TENSE} & \text{PST} \end{bmatrix} \\ \text{TENSE} & \text{PST} \end{bmatrix}$$

The fronted element, in this case a *wh*-proform, contributes the value of the attribute DIS (for 'displaced element') as well as of its *in situ* grammatical function. A special attribute DIS in the main clause is employed for two reasons: firstly, there is evidence that the displaced element plays a grammatical role in the main clause as well as in the subordinate one where it fills a gap (for example, in binding reflexives

7 Note that because it is the *same* f-structure which appears as the value of both SUBJ and XCOMP SUBJ, no issue arises regarding the functional nature of ϕ: the nodes are not projecting two different f-structures, one in each position, but rather a single f-structure which appears in both.

8 The contents of f-structures can be abbreviated by enclosing the words that make them up in double inverted commas, just as we can conceal the internal structure of part of a phrase structure tree using a triangle.

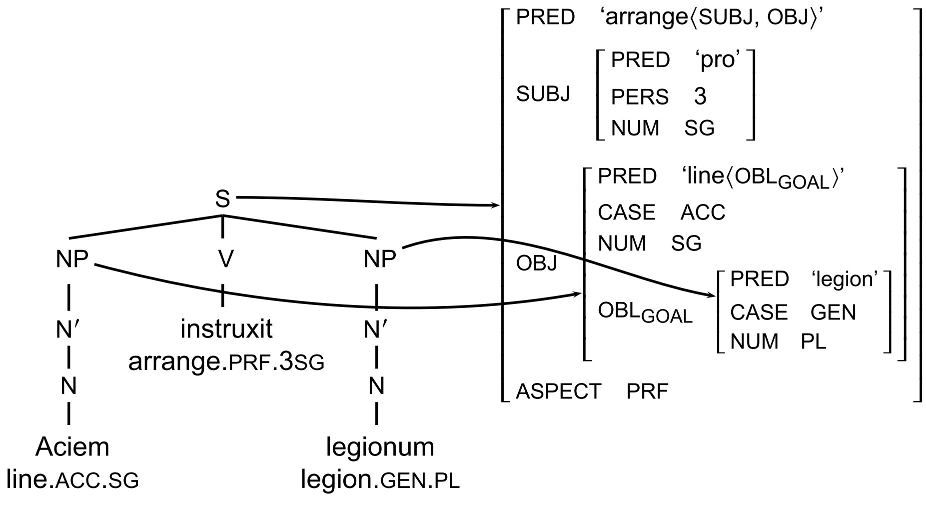
in so-called ‘picture noun phrases’); secondly, by identifying the displaced element with a special attribute at f-structure, we allow for it to have a special role at other levels of representation: we license its fronting at c-structure by associating a specific phrase-structure position with the expression of the DIS attribute, for example. We also license its special role at information structure (i-structure), in this case ‘focus’, by associating the value of DIS, but not OBJects in general, with a special discourse function.

3.4 Crosslinguistic similarities and differences

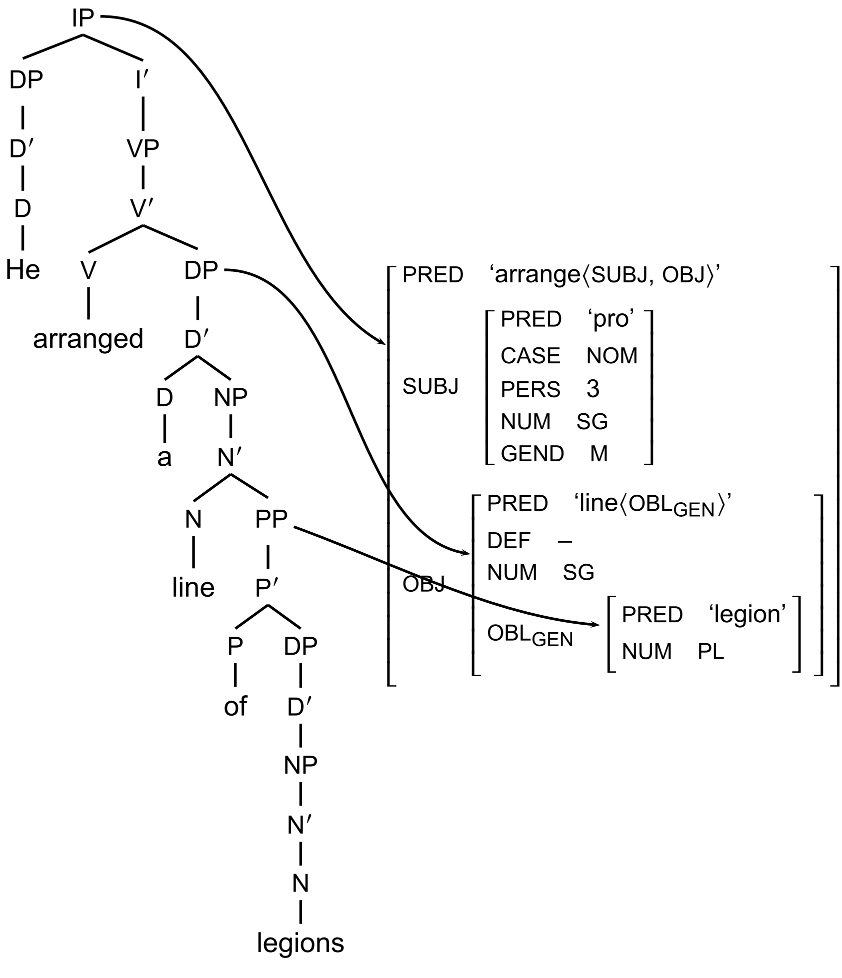
As mentioned, LFG takes the level of representation which is most constant across languages to be f-structure. C-structure can vary widely, as attested by, among other things, the different word orders of the world’s languages. The nature of functional information is taken to be (largely) universal, whereas the means of mapping strings to that functional information is language-specific, albeit constrained by general factors: for example, the less morphology tells us about grammatical functions, the more constituent structure is likely to (and vice versa).

For the purposes of illustration, we give an example analysis of a sentence in Latin (30), which has relatively ‘free’ word order, and English (31), which is more configurational:

(30) Aciem instruxit legionum.
line.ACC.SG arrange.PRF.3SG legions.GEN.PL
‘He arranged a line of legions.’
(simplified from Caes. Gal. 1.24.2; for the full example, see Haug 2017, 124)



(31) ‘He arranged a line of legions.’



4 Evaluation

As in most linguistic theories, LFG accounts of linguistic phenomena are valued to the extent that they provide clear insight into the nature and properties of the phenomena. Evaluation of the relative merits of alternative accounts depends on their success at providing an accurate account of the data at all levels of structure. There is no requirement that constructions with similar meanings must have the same syntactic analysis crosslinguistically, but there is an expectation that there is a basic set of concepts and theoretical vocabulary within each module that is useful for the analysis of all human languages. For example, all languages make use of grammatical functions drawn from the inventory given in Table 1. Similarly, X′ theory is a part of the inventory of c-structure constraints available to all languages. However, languages need not have exclusively X′-theoretic categories, but may instead make extensive use of the exocentric category S, and may use only a subset of universally available c-structure categories.

The treatment of copula sentences offers an example of a point of theoretical debate within LFG. Dalrymple et al. (2004) argue that copular constructions can have different f-structure representations across languages and across constructions within the same language—they can be biclausal or monoclausal (often, but not always, conditioned by the presence or absence of an explicit copula verb), and if biclausal, can take an open or closed complement clause (XCOMP vs. PREDLINK). Attia (2008), on the other hand, argues for a unified analysis, whereby all copula constructions crosslinguistically are treated as involving PREDLINKs. He claims that the approach of Dalrymple et al. (2004) misses the underlying functional similarity, and incorrectly encodes c-structural variation in f-structure. The framework itself does not impose one analysis or the other, but is compatible with various theoretical treatments, which must be decided between based on other criteria such as empirical coverage or analytic efficacy.

Another example is the debate over the existence of traces. For transformationalist theories, something must occupy the canonical c-structure position of a displaced element, since this is how its within-clause grammatical function is ascertained (whether this ‘something’ is a trace *per se* or another kind of object, such as a subsequently deleted copy of the moved phrase). Since LFG separates out such functional information from the phrasal configuration, no such requirement is present in an LFG analysis: thus, c-structure terminals can be much closer to the observed linguistic data, and if something is not pronounced, it need not appear at c-structure. Nonetheless, traces can and have been used in LFG analyses of long-distance dependencies, as mentioned in Section 3.1 (e. g. Bresnan, 1995). Once again, therefore, the LFG framework itself remains agnostic about this particular theoretical question, and empirical factors, compatability with analyses of related phenomena, and/or questions of parsimony will have to be the final arbiters.

4.1 Incrementality

It is a hallmark of constraint-based, nontransformational approaches like LFG that all grammatical constraints associated with the relevant words and constructions must be satisfied. Grammatical representations do not have different properties at different stages of a derivation, as in a transformational or other movement-based approach.

Instead, each constraint adds a piece of information about a structure or the relation between structures. As Halvorsen (1983) observes, “just as in a jigsaw puzzle, what piece is found at what time is inconsequential to the final outcome”. As a corollary of this, we can impose independent psycholinguistic restrictions on the order in which constraints are evaluated, without affecting the resulting analysis: for example, we can build up a structure incrementally, as the words of a sentence are encountered, or we can adopt a head-driven approach, building up the structure of heads first and then their arguments.

4.2 Simplicity

LFG factors grammatical analysis into modules: c-structure, f-structure, semantic structure, information structure, prosodic structure, and so on. The internal structure of each module is relatively simple, since it represents only one aspect of the structure of an utterance. This means that it is easy to isolate a module and explore only one aspect of its linguistic structure—c-structure, for example, or f-structure. Although modules do often interact, proposed changes in analysis that are confined to one module often leave the other modules unaffected, so that, for example, advances in our understanding of the semantics of a construction need not entail a revision to the theory of its syntax, and vice versa.

4.3 Mismatches

LFG assumes that different grammatical modules represent different aspects of grammatical structure, and mismatches between levels are common. For example, c-structure and f-structure have different units, motivated differently and representing different aspects of linguistic structure, and a linguistic unit at f-structure need not correspond to a c-structure unit: as shown in (17), the subject noun phrase translated as ‘these floods’ is represented as a unit at f-structure, but at c-structure the two parts of the phrase are separated, and do not form a unit. What is a unit at one level need not form a unit at all levels of representation.

5 Sample analysis

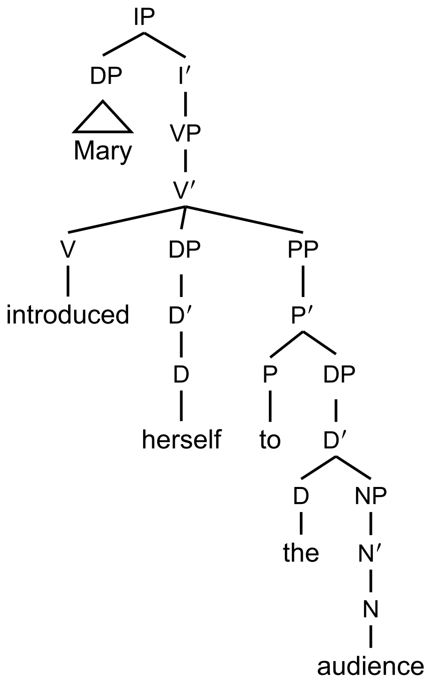
In this section, we provide a sample analysis of the sentence in (32):

(32) After Mary introduced herself to the audience, she turned to a man that she had met before.

We first break the example sentence into two smaller sentences and provide detailed analyses of each of these, before combining them to give an analysis of the whole example in (35), which abbreviates some of the detail introduced in (33–34).

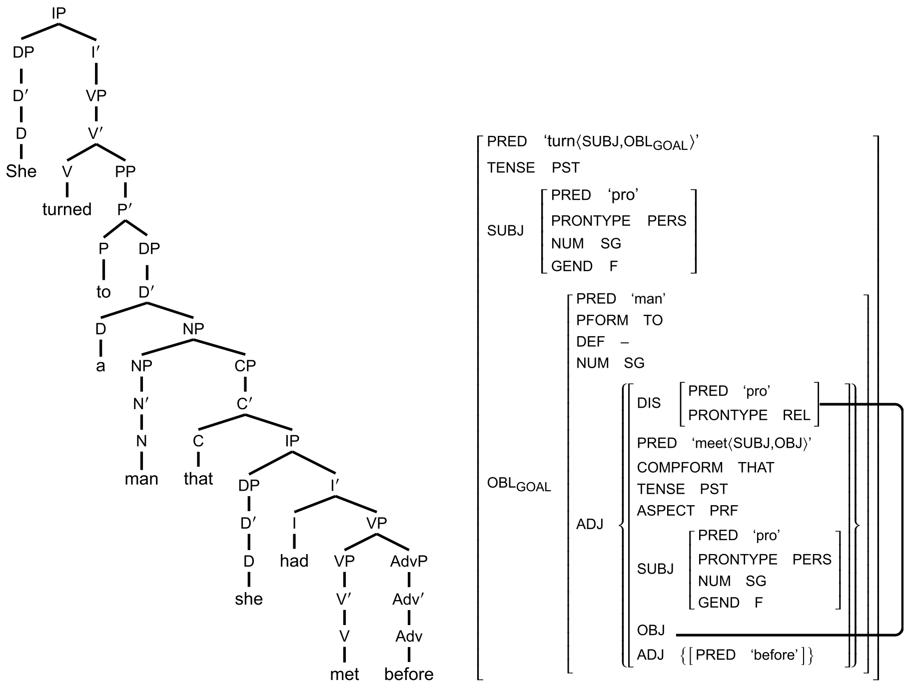
Two points should be noted in what follows. Firstly, in standard LFG analyses, intersentential and intrasentential coreference and binding are not represented at c-structure or f-structure, but instead in the semantics. Thus, the binding of the reflexive *herself* and the potential coreference of *Mary* and *she* are not represented in our analysis. Secondly, we are assuming a theory of adjunction where 'like adjoins to like' (Toivonen, 2003)—that is, zero-level categories can adjoin to other zero-level categories, and maximal projections to maximal projections. Since relative clauses are maximal projections, we therefore adjoin the CP 'that she had met before' at the level of NP (rather than N′, for example) in (34), and since the temporal adjunct 'After Mary introduced herself to the audience' is a CP, this is adjoined at the level of IP in (35).

(33) Mary introduced herself to the audience.

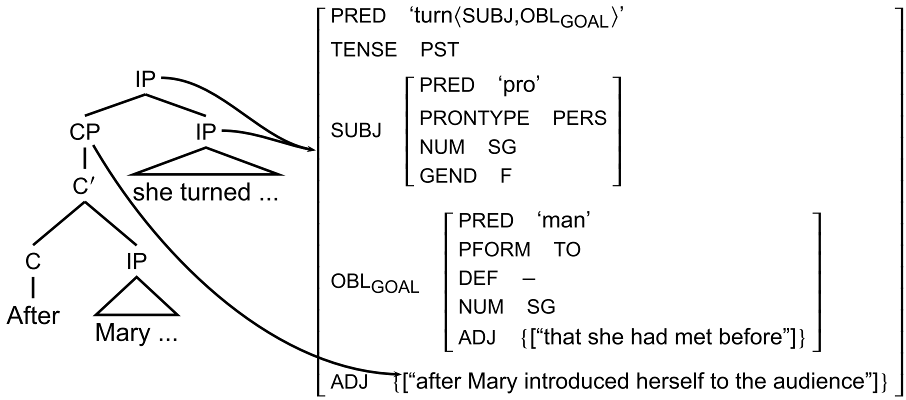


$$\begin{bmatrix} \text{PRED} & \text{`introduce}\langle\text{SUBJ, OBJ, OBL}_{\text{GOAL}}\rangle\text{'} \\ \text{TENSE} & \text{PST} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{`Mary'} \\ \text{NUM} & \text{SG} \end{bmatrix} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{`pro'} \\ \text{PRONTYPE} & \text{REFL} \\ \text{NUM} & \text{SG} \\ \text{GEND} & \text{F} \end{bmatrix} \\ \text{OBL}_{\text{GOAL}} & \begin{bmatrix} \text{PRED} & \text{`audience'} \\ \text{PFORM} & \text{TO} \\ \text{DEF} & + \\ \text{NUM} & \text{SG} \end{bmatrix} \end{bmatrix}$$

(34) She turned to a man that she had met before.



(35) After Mary introduced herself to the audience, she turned to a man that she had met before.



6 Conclusion

LFG has a number of appealing characteristics both in terms of theory and practice. Its mathematically explicit formalisation (exemplified in its successful computational implementation) lends a precision to analyses which can be lost in other theories through appeals to metaphor or other intuitive devices. Its modularity enables parsimonious and accurate description of different phenomena using different formal tools, and this has the added practical advantage of allowing researchers to focus on one particular module without worrying unduly about all of the other components of the grammar simultaneously. This also makes it well-suited for use as a grammatical framework for other areas than purely theoretical syntax, such as in language acquisition in the form of LFG-DOP, or computational grammar development. Finally, its focus on accurate and precise description has made it possible to analyse a number of diverse languages without insisting on similarity where none is otherwise motivated (making non-configurational languages after all configurational at some level of representation, for example).

References

Allen, Cynthia. 1995. *Case marking and reanalysis: Grammatical relations from Old to Early Modern English*. Oxford: Clarendon Press.

Allen, Cynthia. 2008. *Genitives in early English: Typology and evidence*. Oxford: Oxford University Press.

Andrews, Avery D. 2006. Semantic composition for NSM, using LFG+Glue. In *Proceedings of the 2005 Conference of the Australian Linguistic Society*. URL http://www.als.asn.au/proceedings/als2005/andrews-semantic.pdf.

Asudeh, Ash. 2006. Direct compositionality and the architecture of LFG. In Miriam Butt, Mary Dalrymple & Tracy Holloway King (eds.), *Intelligent Linguistic Architectures: Variations on Themes by Ronald M. Kaplan*, 363–387. Stanford: CSLI Publications.

Asudeh, Ash. 2012. *The logic of pronominal resumption*. Oxford: Oxford University Press.

Asudeh, Ash & Gianluca Giorgolo. 2012. Flexible composition for optional and derived arguments. In Miriam Butt & Tracy Holloway King (eds.), *On-Line Proceedings of the LFG2012 Conference*, 64–84. Stanford: CSLI Publications. URL http://web.stanford.edu/group/cslipublications/cslipublications/LFG/17/lfg12.html.

Asudeh, Ash & Ida Toivonen. 2015. Lexical-Functional Grammar. In Bernd Heine & Heiko Narrog (eds.), *The Oxford Handbook of Linguistic Analysis*. 2nd edn. Oxford: Oxford University Press.

Attia, Mohammed. 2008. A unified analysis of copula constructions in LFG. In Miriam Butt & Tracy Holloway King (eds.), *On-Line Proceedings of the LFG2008 Conference*, 89–108. Stanford: CSLI Publications. URL http://web.stanford.edu/group/cslipublications/cslipublications/LFG/13/lfg08.html.

Austin, Peter & Joan Bresnan. 1996. Non-configurationality in Australian Aboriginal languages. *Natural Language and Linguistic Theory* 14(2). 215–268.

Bary, Corien & Dag T. T. Haug. 2011. Temporal anaphora across and inside sentences: The function of participles. *Semantics and Pragmatics*. 4. article 8. 1–56.

Berwick, Robert C. 1982. Computational complexity and Lexical-Functional Grammar. *American Journal of Computational Linguistics* 8. 97–109.

Bettoni, Camilla & Bruno Di Biase (eds.). 2015. *Grammatical development in second languages: Exploring the boundaries of Processability Theory*. Italy: European Second Language Association.

Bod, Rens. 1992. A computational model of language performance: Data Oriented Parsing. In *Proceedings of the 15th International Conference on Computational Linguistics* (Project notes with demonstrations 3), 855–859. International Committee on Computational Linguistics, Nantes, France: Association for Computational Linguistics. URL http://aclweb.org/anthology/C92-3126.

Bod, Rens & Ronald M. Kaplan. 1998. A probabilistic corpus-driven model for Lexical-Functional analysis. In *COLING/ACL98: Joint Meeting of the 36th Annual Meeting of the ACL and the 17th International Conference on Computational Linguistics*, 145–151. Montréal: International Committee on Computational Linguistics.

Börjars, Kersti, Pauline Harries & Nigel Vincent. 2016. Growing syntax: The development of a DP in Northern Germanic. *Language* 92(1). e1–e37.

Börjars, Kersti & Nigel Vincent. 2016. Lexical-functional grammar. In Adam Ledgeway & Ian Roberts (eds.), *The Cambridge Handbook of Historical Syntax*. Cambridge, UK: Cambridge University Press.

Börjars, Kersti, Nigel Vincent & Carol Chapman. 1997. Paradigms, periphrases and pronominal inflection: A feature-based account. In Geert Booij & Jaap van Marle (eds.), *Yearbook of Morphology*, 155–180. Dordrecht: Kluwer.

Bresnan, Joan. 1978. A realistic transformational grammar. In Morris Halle, Joan Bresnan & George A. Miller (eds.), *Linguistic Theory and Psychological Reality*, 1–59. Cambridge, MA: The MIT Press.

Bresnan, Joan. 1993. Interaction between grammar and discourse in Chicheŵa (Bantu). In William A. Foley (ed.), *The Role of Theory in Language Description* (Trends in Linguistics: Studies and Monographs), 45–60. Berlin: Mouton de Gruyter.

Bresnan, Joan. 1995. Linear order, syntactic rank, and empty categories: On weak crossover. In Mary Dalrymple, Ronald M. Kaplan, John T. Maxwell III & Annie Zaenen (eds.), *Formal Issues in Lexical-Functional Grammar*, 241–274. Stanford: CSLI Publications.

Bresnan, Joan. 1998. Morphology competes with syntax: Explaining typological variation in weak crossover effects. In Pilar Barbosa, Danny Fox, Paul Hagstrom, Martha McGinnis & David Pesetsky (eds.), *Is the Best Good Enough? Optimality and Competition in Syntax*, 59–92. MA, Cambridge: The MIT Press and MIT Working Papers in Linguistics. URL https://web.stanford.edu/~bresnan/final-pub-rev.ps. Also presented at the MIT Workshop on Optimality in Syntax, May 1995.

Bresnan, Joan. 2000. Optimal syntax. In Joost Dekkers, Frank van der Leeuw & Jeroen van de Weijer (eds.), *Optimality Theory: Phonology, Syntax and Acquisition*, 334–385. Oxford: Oxford University Press. URL http://www.stanford.edu/ProtectedTildeSymbolbresnan/pt3.pdf.

Bresnan, Joan. 2002. The lexicon in Optimality Theory. In Suzanne Stevenson & Paola Merlo (eds.), *The Lexical Basis of Syntactic Processing: Formal, Computational and Experimental Issues*, 39–58. Amsterdam: John Benjamins.

Bresnan, Joan, Ash Asudeh, Ida Toivonen & Stephen Wechsler. 2016. *Lexical-Functional Syntax*. 2nd edn. Oxford: Wiley Blackwell.

Bresnan, Joan & Jonni M. Kanerva. 1989. Locative inversion in Chicheŵa: A case study of factorization in grammar. *Linguistic Inquiry* 20(1). 1–50. Reprinted in Stowell and Wehrli (1992).

Bresnan, Joan & Sam A. Mchombo. 1987. Topic, pronoun, and agreement in Chicheŵa. *Language* 63(4). 741–782.

Butt, Miriam. 1995. *The structure of complex predicates in Urdu* (Dissertations in Linguistics). Stanford: CSLI Publications. Revised and corrected version of 1993 Stanford University dissertation.

Butt, Miriam, Mary Dalrymple & Anette Frank. 1997. An architecture for linking theory in LFG. In Miriam Butt & Tracy Holloway King (eds.), *On-Line Proceedings of the LFG97 Conference*.

Stanford: CSLI Publications. URL http://web.stanford.edu/group/cslipublications/cslipublications/LFG/LFG2-1997/.

Butt, Miriam & Tracy Holloway King (eds.). 2001. *Time over matter: Diachronic perspectives on morphosyntax*. Stanford: CSLI Publications.

Butt, Miriam & Tracy Holloway King. 2004. The status of case. In Veneeta Dayal & Anoop Mahajan (eds.), *Clause Structure in South Asian Languages*, 153–198. Dordrecht: Springer.

Butt, Miriam, Tracy Holloway King, Hiroshi Masuichi & Christian Rohrer. 2002. The parallel grammar project. In J. Carroll, N. Oostijk & R. Sutcliffe (eds.), *Proceedings of the Workshop on Grammar Engineering and Evaluation*, 1–7. COLING02.

Butt, Miriam, Tracy Holloway King, María-Eugenia Niño & Frédérique Segond. 1999. *A grammar writer's cookbook*. Stanford: CSLI Publications.

Chaves, Rui P. 2018. Construction grammar. In András Kertész, Edith Moravcsik & Csilla Rákosi (eds.), *Current Approaches to Syntax: A Comparative Handbook*. De Gruyter Mouton.

Coppock, Elizabeth & Stephen Wechsler. 2010. Less-travelled paths from pronoun to agreement: The case of the Uralic objective conjugations. In Miriam Butt & Tracy Holloway King (eds.), *On-Line Proceedings of the LFG2010 Conference*. Stanford: CSLI Publications. URL http://web.stanford.edu/group/cslipublications/cslipublications/LFG/15/lfg10.html.

Crouch, Dick, Mary Dalrymple, Ronald M. Kaplan, Tracy Holloway King, John T. Maxwell III & Paula S. Newman. 2008. *XLE documentation*. Palo Alto Research Center, Palo Alto, CA. URL http://www2.parc.com/isl/groups/nltt/xle/doc/xle_toc.html.

Dalrymple, Mary (ed.). 1999. *Semantics and syntax in Lexical Functional Grammar: The resource logic approach*. Cambridge, MA: The MIT Press.

Dalrymple, Mary. 2001. *Lexical Functional Grammar* (Syntax and Semantics, 34). New York: Academic Press.

Dalrymple, Mary. 2015. Morphology in the LFG architecture. In Miriam Butt & Tracy Holloway King (eds.), *On-Line Proceedings of the LFG2015 Conference*. Stanford: CSLI Publications. URL http://csli-publications.stanford.edu/LFG/20/lfg15.html.

Dalrymple, Mary, Helge Dyvik & Tracy Holloway King. 2004. Copular complements: Closed or open? In Miriam Butt & Tracy Holloway King (eds.), *On-Line Proceedings of the LFG2004 Conference*. Stanford: CSLI Publications. URL http://csli-publications.stanford.edu/LFG/9/lfg04.html.

Dalrymple, Mary, Dag T. T. Haug & John J. Lowe. 2017. Integrating LFG's binding theory with PCDRT. In Miriam Butt & Tracy Holloway King (eds.), *On-Line Proceedings of the LFG2017 Conference*. Stanford: CSLI Publications. URL http://csli-publications.stanford.edu/LFG/22/lfg17.html.

Dalrymple, Mary, Ronald M. Kaplan & Tracy Holloway King. 2007. The absence of traces: Evidence from weak crossover. In Annie Zaenen, Jane Simpson, Tracy Holloway King, Jane Grimshaw, Joan Maling & Christopher Manning (eds.), *Architectures, Rules, and Preferences: Variations on Themes by Joan W. Bresnan*. Stanford: CSLI Publications.

Dalrymple, Mary, Ronald M. Kaplan & Tracy Holloway King. 2015. Economy of expression as a principle of syntax. *Journal of Language Modelling* 2(3). 377–412. URL http://jlm.ipipan.waw.pl/index.php/JLM/article/view/82.

Dalrymple, Mary, Ronald M. Kaplan, John T. Maxwell III & Annie Zaenen (eds.). 1995. *Formal issues in Lexical-Functional Grammar*. Stanford: CSLI Publications.

Dalrymple, Mary, John J. Lowe & Louise Mycock. 2019. *The Oxford Reference Guide to Lexical Functional Grammar*. Oxford: Oxford University Press.

Dalrymple, Mary & Louise Mycock. 2011. The prosody-syntax interface. In Miriam Butt & Tracy Holloway King (eds.), *On-Line Proceedings of the LFG2011 Conference*. Stanford: CSLI Publications. URL http://web.stanford.edu/group/cslipublications/cslipublications/LFG/16/lfg11.html.

Dalrymple, Mary & Irina Nikolaeva. 2011. *Objects and information structure*. Cambridge, UK: Cambridge University Press.

Dyvik, Helge, Paul Meurer, Victoria Rosén & Koenraad De Smedt. 2009. Linguistically motivated parallel parsebanks. In *Proceedings of the Eighth International Workshop on Treebanks and Linguistic Theories*. Milan: EDUCatt.

Falk, Yehuda N. 2001. *Lexical-Functional Grammar: An introduction to parallel constraint-based syntax*. Stanford: CSLI Publications.

Falk, Yehuda N. 2006. *Subjects and universal grammar: An explanatory theory*. Cambridge, UK: Cambridge University Press.

Featherston, Sam. 2008. Thermometer judgements as linguistic evidence. In C. M. Riehl & A. Rothe (eds.), *Was ist linguistische Evidenz?* Aachen: Shaker Verlag.

Findlay, Jamie Y. 2016. Mapping theory without argument structure. *Journal of Language Modelling* 4(2). 293–338. URL http://jlm.ipipan.waw.pl/index.php/JLM/article/view/171/0.

Giorgolo, Gianluca & Ash Asudeh. 2011. Multimodal communication in LFG: Gestures and the correspondence architecture. In Miriam Butt & Tracy Holloway King (eds.), *On-Line Proceedings of the LFG2011 Conference*, 257–277. Stanford: CSLI Publications. URL http://web.stanford.edu/group/cslipublications/cslipublications/LFG/16/lfg11.html.

Goddard, Cliff & Anna Wierzbicka. 2002. *Meaning and universal grammar: Theory and empirical findings*. John Benjamins. Two volumes.

Halvorsen, Per-Kristian. 1983. Semantics for Lexical Functional Grammar. *Linguistic Inquiry* 14(4). 567–615.

Haug, Dag T. T. 2017. Backward control in Ancient Greek and Latin participial adjuncts. *Natural Language and Linguistic Theory* 35(1). 99–159.

Hornstein, Norbert. 1995. *Logical Form: From GB to Minimalism*. Oxford: Blackwell.

Hornstein, Norbert. 2018. The minimalist programme. In András Kertész, Edith Moravcsik & Csilla Rákosi (eds.), *Current Approaches to Syntax: A Comparative Handbook*. De Gruyter Mouton.

Jackendoff, Ray. 1977. *$\bar{X}$ syntax: A study of phrase structure*. Cambridge, MA: The MIT Press.

Jackendoff, Ray & Jenny Audring. 2018. The parallel architecture. In András Kertész, Edith Moravcsik & Csilla Rákosi (eds.), *Current Approaches to Syntax: A Comparative Handbook*. De Gruyter Mouton.

Joshi, Aravind K. 1985. Tree adjoining grammars: How much context-sensitivity is required to provide reasonable structural descriptions? In David R. Dowty, Lauri Karttunen & Arnold M. Zwicky (eds.), *Natural Language Parsing*, 206–250. Cambridge University Press.

Kaplan, Ronald M. 1987. Three seductions of computational psycholinguistics. In Peter Whitelock, Mary McGee Wood, Harold L. Somers, Rod Johnson & Paul Bennett (eds.), *Linguistic Theory and Computer Applications*, 149–188. London: Academic Press. Also published as CCL/UMIST Report No. 86.2: Alvey/ICL Workshop on Linguistic Theory and Computer Applications: Transcripts of Presentations and Discussions. Center for Computational Linguistics, University of Manchester. Reprinted in Dalrymple et al. (1995, 337–367).

Kaplan, Ronald M. & Joan Bresnan. 1982. Lexical-Functional Grammar: A formal system for grammatical representation. In Joan Bresnan (ed.), *The Mental Representation of Grammatical Relations*, 173–281. Cambridge, MA: The MIT Press. Reprinted in Dalrymple et al. (1995, 29–130).

Kaplan, Ronald M. & Miriam Butt. 2002. The morphology-syntax interface in LFG. URL http://cslipublications.stanford.edu/LFG/7/lfg02.html. Abstract in *On-Line Proceedings of the LFG2002 Conference*, Miriam Butt & Tracy Holloway King (eds.).

Kaplan, Ronald M. & John T. Maxwell III. 1996. LFG grammar writer's workbench. Technical report, Xerox Palo Alto Research Center, Palo Alto, CA. URL http://www2.parc.com/isl/groups/nltt/medley/.

Kaplan, Ronald M., John T. Maxwell III & Annie Zaenen. 1987. Functional uncertainty. In *CSLI Publications Monthly Newsletter*. Stanford University.
Kaplan, Ronald M. & Annie Zaenen. 1989. Long-distance dependencies, constituent structure, and functional uncertainty. In Mark Baltin & Anthony S. Kroch (eds.), *Alternative Conceptions of Phrase Structure*, 17–42. Chicago: University of Chicago Press. Reprinted in Dalrymple et al. (1995, 137–165).
Kayne, Richard S. 1984. *Connectedness and binary branching*. Dordrecht: Foris.
Kibort, Anna. 2004. Passive and passive-like constructions in English and Polish. Ph.D. thesis, University of Cambridge.
King, Tracy Holloway & Annie Zaenen. 2004. F-structures, information structure, and discourse structure. Christchurch, New Zealand: University of Canterbury. Presented at the *Winter School in LFG and Computational Linguistics*. URL http://cslipublications.stanford.edu/LFG/9/lfg04kingzaenen-abs.html.
Kroeger, Paul. 1993. *Phrase structure and grammatical relations in Tagalog* (Dissertations in Linguistics). Stanford: CSLI Publications. Revised and corrected version of 1991 Stanford University dissertation.
Kuhn, Jonas. 2001. Formal and computational aspects of Optimality-Theoretic syntax. Ph.D. thesis, Universität Stuttgart.
Legendre, Géraldine. 2018. Optimality theoretical syntax. In András Kertész, Edith Moravcsik & Csilla Rákosi (eds.), *Current Approaches to Syntax: A Comparative Handbook*. De Gruyter Mouton.
Lowe, John J. 2012. The syntax and semantics of tense-aspect stem participles in Early Ṛgvedic Sanskrit. D.Phil. thesis. University of Oxford.
Maxwell, John T. III. 2015. XLE. URL http://www2.parc.com/isl/groups/nltt/xle/.
May, Robert. 1985. *Logical form: Its structure and derivation*. Cambridge, MA: The MIT Press.
McCawley, James D. 1968. Concerning the base component of a transformational grammar. *Foundations of Language* 4. 243–269. Reprinted in McCawley (1973).
McCawley, James D. 1973. *Grammar and meaning: Papers on syntactic and semantic topics*. Tokyo: Taishukan.
Müller, Stefan & Antonio Machicao y Priemer. 2018. Head-driven phrase structure grammar. In András Kertész, Edith Moravcsik & Csilla Rákosi (eds.), *Current Approaches to Syntax: A Comparative Handbook*. De Gruyter Mouton.
Mycock, Louise & John J. Lowe. 2013. The prosodic marking of discourse functions. In Miriam Butt & Tracy Holloway King (eds.), *On-Line Proceedings of the LFG2013 Conference*. Stanford: CSLI Publications. URL http://web.stanford.edu/group/cslipublications/cslipublications/LFG/18/lfg13.html.
Nespor, Marina & Irene Vogel. 1986. *Prosodic phonology*. Dordrecht: Foris.
Patejuk, Agnieszka & Adam Przepiórkowski. 2015. Parallel development of linguistic resources: Towards a structure bank of Polish. *Prace Filologiczne* 65. 255–270. URL https://www.ceeol.com/search/viewpdf?id=307372.
Pienemann, Manfred. 1998. *Language processing and second language development: Processability theory*. Amsterdam: John Benjamins.
Pienemann, Manfred (ed.). 2005. *Cross-linguistic aspects of Processability Theory* (Studies in Bilingualism, 30). Amsterdam: John Benjamins.
Prince, Alan & Paul Smolensky. 2004. *Optimality Theory: Constraint interaction in generative grammar*. Oxford: Blackwell.
Sadler, Louisa & Andrew Spencer. 2001. Syntax as an exponent of morphological features. In Geert Booij & Jaap van Marle (eds.), *Yearbook of Morphology*, 71–96. Dordrecht: Kluwer.
Schachter, Paul. 1976. The subject in Philippine languages: Topic, actor, actor-topic, or none of the above. In Charles N. Li (ed.), *Subject and Topic*. New York: Academic Press.

Selkirk, Elisabeth O. 1981. On the nature of phonological representation. In Terry Myers, John Laver & John Anderson (eds.), *The Cognitive Representation of Speech*, 379–388. Amsterdam: North Holland.

Simon, Herbert A. 1962. The architecture of complexity. *Proceedings of the American Philosophical Society* 106(6). 467–482.

Snijders, Liselotte. 2015. The nature of configurationality in LFG. D.Phil. thesis. University of Oxford.

Spevak, Olga. 2010. *Constituent order in classical Latin prose* (Studies in Language Companion Series, 117). John Benjamins.

Stowell, Timothy & Eric Wehrli (eds.). 1992. *Syntax and the lexicon* (Syntax and Semantics, 26). San Diego: Academic Press.

Toivonen, Ida. 2003. *Non-projecting words: A case study of Swedish verbal particles*. Dordrecht: Kluwer.

van Genabith, Josef & Richard Crouch. 1999. Dynamic and underspecified semantics for LFG. In Mary Dalrymple (ed.), *Semantics and Syntax in Lexical Functional Grammar: The Resource Logic Approach*. Cambridge, MA: The MIT Press.

Vincent, Nigel. 1999. The evolution of c-structure: Prepositions and PPs from Indo-European to Romance. *Linguistics* 37(6). 1111–1153.

Wedekind, Jürgen & Ronald M. Kaplan. 2019. Tractable Lexical-Functional Grammar. Submitted.

Sam Featherston

6 The Decathlon Model

Abstract: The Decathlon Model is an exploration of what changes would be necessary to standard models of generative grammar in order to make them more empirically adequate. Following Chomsky (1965), generative grammars are taken to be models of human introspective intuitions of linguistics well-formedness. The experimental results show that only one major change is required, namely the linguistics constraints which make up the grammar need to have a violation cost assigned to them. Put differently, it is not enough for a grammar say that a phenomenon is not permitted to occur in a language, the grammar must also specify what cost in perceived well-formedness results if the phenomenon does indeed occur. This 'crime and punishment' model of the architecture of the grammar is argued to account for many so far poorly understood aspects of the way that sentence grammars operate.

1 Introduction

The Decathlon Model is a development from the Chomskyan tradition of grammars, though it is not generative in the technical sense. Fundamentally, it attempts to reflect the features of the available data basis more faithfully in the assumptions about how the grammar works. It is motivated by three basic assumptions about the nature of an *empirically adequate* grammar.

- The grammar should reflect the features of the primary language data.
- It can have features which go beyond the data basis, but it will not have features which are in conflict with what the data shows.
- The quality of a grammar can never exceed the quality of its data basis.

This leads naturally to the question what the primary language data is and what features it exhibits. We discuss this in detail below, but we shall note two major findings that set the Decathlon Model off from many other models. The first is an empirical observation: well-formedness is gradient. While this has often been perceived, it is not generally reflected in the architecture of the grammar; the Decathlon Model is an attempt to do exactly that. Our second finding is that this requires one major addition to fairly standard assumptions about how the grammar works: the 'rules' of the grammar – we shall call them here *constraints* – are not prohibitions, rather they are conditional statements of the form: *If a structure violates constraint X, it incurs the penalty in well-*

Acknowledgement: The development of this model took place within the collaborative research centres SFB 441 *Linguistic Evidence* and SFB 833 *Constructing Meaning*, both supported by the Deutsche Forschungsgemeinschaft. Many thanks are due to the staff of both. Thanks too to Katharina Salkowski and Miriam Gänzle for the use of experiment data.

https://doi.org/10.1515/9783110540253-006

formedness x. The grammar must therefore specify not only the crime, but also the punishment.

2 Data

2.1 Empirical motivation

The question of the most appropriate data sources for grammatical theory building is of central importance to the Decathlon Model. There are two reasons for this: first, because this model was developed because of a feeling that the evidential base in data employed in syntactic analysis and theory construction in recent years were insufficient for the purpose. Second, because the model is based upon a reassessment of the nature of certain data types and the relationships between these data types. In the light of this thinking, the model attempts to reflect the data faithfully, even when it seems to contradict long-held assumptions.

The mismatch between data and theory that the Decathlon Model is responding to has several aspects. First, it is a fairly common event for a reader of a syntax article to disagree with the judgements, or find the pattern of judgements claimed in the article to be an incomplete picture of perceptible differences. However, this perception has often not been regarded as a relevant concern by many syntacticians; instead they regard it merely as an issue of idiolectal variation of no consequence (e. g. Wasow, 2002, xiv). On other occasions, linguists interpret a given phenomenon as having a unique and homogeneous cause, ignoring the fact that parts of the effect that they are responding to can be identified in other related structures (e. g. Superiority effect below for an example). Third, the criteria for the evaluation of syntactic analyses have sometimes been ill-defined. This is particularly the case for factors such as 'theoretical elegance' and 'economy', which surely cannot contribute towards an assessment of analytical success without specification of how such qualities are to be measured. A fourth concern is replicability: sometimes reference is made just to the intuition of the individual linguist and their idiolect, so that the claims must be taken on trust (see Featherston, 2007 for further discussion).

The Decathlon Model is an attempt to construct a grammar in the generative tradition but with a clear commitment to respecting the data of language. To the extent that they are compatible with the evidence, the standard features of grammars in the Chomskyan tradition are adopted and maintained. Those features which turn out not to be represented in the data, however, are replaced with the empirically motivated alternatives.

This should not be seen as a rejection of the previous tradition, merely as a suggested new direction for further development. There were quite rational reasons why a generation of syntacticians grew up for whom the finer details of the primary data were of secondary interest. The advances made in the sixties and seventies had led to a consensus that the linguist's own judgements were a sufficient source of hypotheses and empirical confirmation, so there was no immediate need to adduce other evidence. Furthermore, the lack of simple correspondence between other more complex data types and the theoretical models were thought to show that current grammatical models had not yet attained the degree of abstractness at which they were represented in the human mind (Chomsky, 1986).

It is interesting to consider the role that Chomsky's own writing played in this. He famously notes in *Abstracts* (Chomsky, 1965, p. 19) that he considers the search for new data types a matter of small concern, since the introspective judgements available provide ample evidence for phenomena that require explanation. But it is clear from his text that he is aware that this is an abstraction from the primary data, a simplification in order for short-term progress to be made.

> Like acceptability, grammaticalness is, no doubt, a matter of degree (Chomsky, 1955, 1957, 1961), [...]. (*Aspects*, 11)

He limits his chosen low-data approach to "clear cases", to the "masses of evidence that are hardly open to serious question" (*Aspects* 19). He also calls upon linguists who feel that they have a source of evidence about the nature of the grammar which will yield a fuller picture to make their case. That is what the Decathlon Model attempts to do (cf. also Schütze, 1996, 26), and why it can be seen as an extension to the Chomskyan grammatical tradition. This may sound like a radical break, but in fact only one large change needs to be made, namely in the conception of well-formedness.

2.2 Types of judgements

Grammars in the Chomskyan tradition are fundamentally grammars of well-formedness as instantiated in introspective judgements:

> [...] there is no way to avoid the traditional assumption that the speaker-hearer linguistic intuition is the ultimate standard that determines the accuracy of any proposed grammar, [...]. (*Aspects*; 21)

This holds not because Chomsky said it, but because it has been so widely accepted since. A grammar which does not accept it is fundamentally attempting to model something else. An important factor which locates the Decathlon Model as a grammar in the generative tradition is that it adopts this position too. Other data types are of course permissible and useful in investigations of grammar, but these other data types must be related to intuitions of acceptability in order to be clearly interpretable evidence. The Decathlon Model attempts to do exactly this, as it relates judgement data and occurrence frequencies to one another systematically.

The first step in this is to establish our assumptions about what lies behind the data types, starting with judgements. First of all, let us note that when we talk of a structure here, we are referring to a form-meaning pair. We cannot judge a sentence without first interpreting it, which involves assigning a grammatical structure to the surface form. So if we gather judgements of an example such as (1a), where the non-finite verb is understood to be active, we may well receive very different results from those of (1b), where the non-finite verb will be understood to be passive. The surface forms are apparently the same, but we will assign them different structures. These structures are likely to trigger different impressions of well-formedness, based upon their differing interpretations.

(1) a. The guests are ready to eat.
    b. The pizzas are ready to eat.

The opposite case, where two forms have an identical meaning, is perhaps rarer but nevertheless not impossible. The pair in (2) could receive different acceptability ratings in spite of meaning exactly the same thing (my judgement).

(2) a. I'll fax it you.
    b. I'll fax you it.

We thus take the object of well-formedness judgements to be a form-meaning pair.

As a next step we must distinguish the different types of judgements to be found in the literature (see Featherston, 2007). First, there are *string-technical judgements*. The question that these answer is perhaps *Does this example break any known rule?* or *Would we predict this example to be judged to be good or bad?* We invite the reader to give a string-technical judgement of (3).

(3) Did John tell you how many of his exams that he failed?

The standard view of (3) would be that it violates the Doubly Filled Comp Filter, so the string-technical judgement would be that it contains a violation (cf. Zwicky, 2002). Since in fact (3) is quite acceptable, we can see that string-technical judgements are not themselves intuitions of well-formedness; rather, they depend upon the linguist's conscious knowledge of the grammar. They cannot therefore serve as a data basis for the building of a grammar, since this would be circular.

The next type of judgements are *categorical judgements*. These answer a question like *Is this example good enough to occur?* or *Can I imagine hearing this?* The standard judgements of the examples in (4) would be that the first is possible, the second is not.

(4) a. I'd love to know who danced with her at the party.
    b. *I'd love to know who that danced with her at the party.

The third sort of judgements are *relative judgements*. These answer the question: *How natural does this sound in this meaning?* or *How acceptable is this form in this meaning?* It will be clear that it is necessary to provide some sort of comparison point for these judgements. This can be a scale of well-formedness which has either descriptions of (some of) the points on it or, much better, examples which anchor the scale points by providing exemplars of their values.

These last two judgement types are the most common in the literature and are often not distinguished. I think many syntacticians would subscribe to the idea that grammaticality is fundamentally binary in the sense that there are such statuses as 'absolutely grammatical' and 'absolutely ungrammatical', perhaps with the proviso that performance factors make the boundaries fuzzy. Nevertheless, syntax papers routinely differentiate between examples merely on the basis of degrees of acceptability (giving one example a question mark ?, for example, and another an asterisk *) and treat such contrasts as sufficient motivation for a grammatical account. Syntacticians thus often allow themselves a degree of vagueness between a binary model of well-formedness and a gradient model.

I would argue that relative judgements are the most basic form of the intuitions of well-formedness: categorical judgements are thus derivative. They build upon relative judgements but add a threshold of acceptability, which is used as a decision criterion (see Bader and Häussler, 2010 for work on this). Since all other information is discarded, it will be clear that categorical judgements contain *less information* than relative judgements.

One piece of evidence for the non-primary status of categorical judgements is that informants often disagree widely about where to draw the line in a binary choice task (see Schütze, 1996). By contrast, experiments gathering relative judgements show that informants place examples relative to other examples with considerable consistency. Another piece of evidence for the primacy of relative judgements is that informants can distinguish far more than just acceptable and unacceptable; even a single individual can distinguish at least ten gradations (e. g. Featherston, 2009). This is the additional information which is discarded when employing a binary model of well-formedness.

In fact it seems likely that relative judgements are related to processing complexity. Humans have a conscious awareness of the mental effort involved in a task. We perceive the sum 7+13 to be easy, 7x13 to be more difficult, and 7÷13 to be more difficult still. It seems reasonable to assume that a related perception of difficulty is available for linguistic computation, and it is this that provides the basis for intuitive judgements of relative well-formedness. We are therefore assuming that judgements can be utilized as a measure of computational effort or processing difficulty, just as for example reading time is. If a structure in a particular meaning is easier to process than another, we perceive that form-meaning pair to be more well-formed. Let us hasten to add that relative judgements are not a magic bullet and they are of course the result of a complex series of task-related mental functions which we cannot even identify, much less distinguish. But they are fundamentally a psycholinguistic measure among others.

Notice that this means that we cannot gather reliable judgements of either ambiguous forms with equally accessible meanings or forms which have no single uniquely identifiable structure/meaning. Thus neither (5a) nor (5b) can provide usable relative judgements without some clarification of what they should mean.

(5) a. The doctor pointed to the patient with a burnt finger.
b. The woman are working.

2.3 The characteristics of well-formedness

We have said that the Decathlon Model is a model of introspective well-formedness judgements. We have further argued that relative judgements are the most useful representation of these intuitions. In this section we describe what this data type shows us. We shall illustrate this with a specific experiment gathering relative well-formedness judgements, but we find these features systematically.

This experiment was about multiple wh-questions in German; the aim was to quantify the effects of Superiority and 'discourse linking' (Pesetsky, 1987) by varying the forms of the wh-items in preposed and in situ locations. The experiment had eight conditions deriving from three binary factors: Extracted Element (WH-SUBJECT, WH-OBJECT), Type of Subject wh-item (BARE WH-WORD, WH-PHRASE), and Type of Object wh-item (BARE WH-WORD, WH-PHRASE). In the graphic, bare wh-words are coded with *wh-* and wh-phrases are coded as *wx-*, so *wh-subj* indicates the bare wh-item *wer* ('who'), and wx-DO indicates a wh-phrase like *welche Arznei* ('which medicine'), for details see Featherston (2005).

Typical sentences had forms such as those in (6) and (7). In (6) the wh-subject has been preposed, the object left in situ. Both wh-items are wh-phrases. In (7), by contrast, the wh-object has been preposed, and both wh-items are bare wh-words.

(6) Welcher Arzt hat dem Patienten welche Arznei gegeben?
which doctor has to.the patient which medicine given
'Which doctor gave the patient which medicine?'

(7) Was hat wer dem Patienten gegeben?
what has who to.the patient given
'What did who give to the patient?'

In this experiment we used the method Thermometer Judgements (Featherston, 2009), which permits us to capture a large proportion of informants' intuitions of well-formedness with little distortion. Space does not allow us to discuss the method in detail here, for our purposes here it will suffice to say that informants were instructed to give judgements of examples on a numerical scale, and that the scale is anchored with example sentences for comparison. The question put to the informants is of the type: *How natural does this example sound relative to the anchor examples?* This approach aims to allow informants to express the full range of the varying degrees of well-formedness that they perceive.

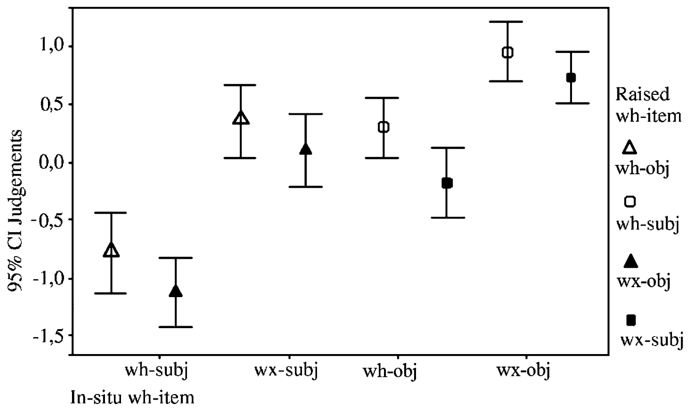


Figure 1: This experiment had eight conditions deriving from three binary factors: Extracted Element (WH-SUBJECT, WH-OBJECT), Type of Subject wh-item (BARE WH-WORD, WH-PHRASE), and Type of Object wh-item (BARE WH-WORD, WH-PHRASE).

Figure 1 shows the judgements of the eight experimental conditions, which are distinguished on the horizontal axis. The judgements are quantified on the vertical axis; higher scores indicate 'more natural' judgements, lower scores 'less natural' judgements. Note that the judgements have been transformed to normalized z-scores in order to reduce inter-subject differences. The scores by conditions are shown by error bars: the marker at the middle of the bar indicates the mean value, while the length of the bar shows the 95 % confidence interval of the mean.

While the results of this experiment were interesting, it is the overall pattern of the judgements that concerns us here. We are interested in what the pattern reveals to us about the nature and effect of constraint violations. We can identify four features that we find in this data, and in relative judgements more generally.

To make these features clear we present the identical data set again but in a stylized form in Figure 2. In the revised form of the graphic, the three binary parameters distinguishing the conditions are simplified to a 0 or 1 on the *x* axis of the graph. The pair of conditions 000 and 001 thus form a minimal pair, as do conditions 010 and 110 and the conditions 110 and 111. Notice also that the factor which distinguishes condition 000 from condition 001 is the same factor which distinguishes 110 from 111; we will therefore refer to this contrast as the xx0:xx1 constraint. In Figure 2 each minimal pair is joined by a line to its minimal pair partner. Since the conditions vary on three binary parameters, each condition has three minimal pair partners. Since one of the minimal pair partners is always judged weaker than the other, we may think of these minimal pairs as illustrating the effect of a constraint violation, with the 0 in a position showing the constraint violation, the 1 showing the constraint satisfaction.

Let us note that we find these features fairly consistently across studies, not just in this example. The figure thus illustrates the systematic relationships between

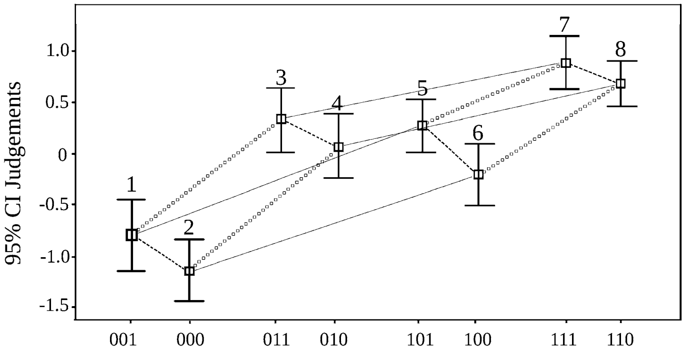


Figure 2: This is the same experiment result as above, but with the minimal pairs joined by lines to illustrate the effect of each constraint.

constraint-fulfilling and constraint-violating structures in terms of perceived well-formedness. Since our grammar is a model of well-formedness, we offer this study as evidence how our mental grammar works.

Constraint violations are thus:

1 quantifiable
2 constraint-specific
3 automatic (i. e. unconditional)
4 cumulative

We shall take these in turn. Constraint violations are *quantifiable* because however good an example is, its minimal pair with the violation is worse by about the same quantity in terms of well-formedness. In the experiment result graphic this is recognizable in the fact that the lines joining the same minimal pairs are roughly parallel. The drop in well-formedness between all conditions 0xx and 1xx is about the same. Compare for example the pairs of bars numbered 1 and 5 with those numbered 2 and 6, and again 3 and 7, and 4 and 8. We can therefore say that the violation cost is quantifiable, given a suitable scale (see Tools below).

Constraint violations are also *constraint-specific* because not all the three constraint violations in this study have the same violation costs. The constraint x0x:x1x causes the largest drop (that is from error bar 3 to error bar 1, 4 to 2, 7 to 5, or 8 to 6), followed by the 0xx:1xx constraint (5:1, 6:2, 7:3, 8:4); and lastly the xx0:xx1 constraint (1:2, 3:4, 5:6, 7:8).

This could be paraphrased as there being stronger and weaker constraints. But the 'strength' of a constraint has nothing to do with the probability that it is violated, or the absoluteness of the prohibition, rather it concerns what happens when it is violated. We can see this as being similar to the way the criminal law functions. The difference between the law forbidding copying DVDs and the law forbidding murder is not the strength of the prohibition: both are forbidden absolutely. The difference lies in the severity of punishment of offenders. Relative judgement data shows us that this holds for grammatical constraints too. Grammatical models often model constraint strength in terms of violation probabilities, or rule ordering; but this data shows neither of these. What it does show is variation in the cost in well-formedness caused by violating the constraint.

Thirdly, violation costs are *automatic*, by which we mean *automatically applied*. A grammar employing the traditional binary grammaticality contrast may sometimes show no effect of a constraint violation. Fairly weak violations whose costs are not sufficient to trigger ungrammaticality are essentially invisible, as are additional constraint violations in an example deemed already ungrammatical. Data from studies gathering relative judgements do not exhibit this pattern, however. Normally speaking, any additional constraint violation will make a structure worse by the specific cost in well-formedness of the violated constraint.[1]

Last, violation costs are *cumulative*. This means that multiple fairly minor constraint violations can make a structure as bad or indeed worse than one major violation does (see Keller, 2000 for discussion).

Intuitions of well-formedness can thus be seen as representing a measurable quantity. Experience with other studies reveals that the effects found in studies of grammatical constraints do not differ in kind or in behaviour from those referred to as representing processing complexity.[2] We therefore argue that our intuitions of well-formedness are a form of proprioceptive measure of the cognitive effort involved processing language structures (cf. Bornkessel-Schlesewsky and Schlesewsky, 2007). This is to be welcomed, because it ties grammars based on judgements in to the wider field of psycholinguistics.

This should come as no surprise, because the generative enterprise, with its emphasis on explanatory adequacy deriving from language acquisition, was originally psycholinguistic, though this aspect became less focused on in later years. The position that we are representing here is exactly what Chomsky says about the nature of well-formedness in *Aspects*:

> Obviously, acceptability will be a matter of degree, along several dimensions. [...] it is clear that we can characterize unacceptable sentences only in terms of some 'global' property of derivations and the structures they define – a property which is attributable, not to a particular rule, but rather to the way in which rules interact [...] (*Aspects*; 10)

1 Let us note here a caveat. The features we describe here apply to introspective judgements within a certain field of acceptability, roughly as far as we interpret and process an example as an interpretable sentence. They also require us to be able to analyse what is wrong; these features do not apply to strings of word salad, because the violations are not clearly identifiable.

2 This means: if we test constraints which are thought of as grammatical and those thought of as performance, we see no systematic difference between them. So-called 'grammatical' effects will be typically stronger, but there is no difference in kind or behaviour. See also Section 4.2.

2.4 The relationship between well-formedness and frequency

Having characterized judgement data as related to processing complexity, we now contrast this with corpus data, which provides a measure of frequency of occurrence. Our basic assumption is that there is a causal link from well-formedness to occurrence. Specifically, we see this as the result of a competition: all other things being equal, the easiest or most accessible structure is selected to instantiate a given message (in production) or as the parse of an input string in receptive processing.

It seems plausible that this competition depends upon the same aspects of language processing as the weight effect of Behaghel (1909/1910), Hawkins (1994) and many others in between: short and/or simple constituents are encoded faster than long and/or complex constituents, so they tend to appear earlier if their relative order is not grammatically critical (race model of processing McRoy and Hirst, 1990). Similarly, well-formed/more easily processed structures can be planned and encoded faster than less well-formed ones, so the more well-formed ones get to be output and appear in your corpus.[3]

The claim that occurrence frequency is dependent on well-formedness makes a clear prediction, namely that corpus data will distinguish only between forms which are all good enough to occur, while judgements will distinguish even between forms which are not good enough to occur. We see this confirmed in an example in Figure 3, which shows the frequencies and introspective ratings of a range of structures which

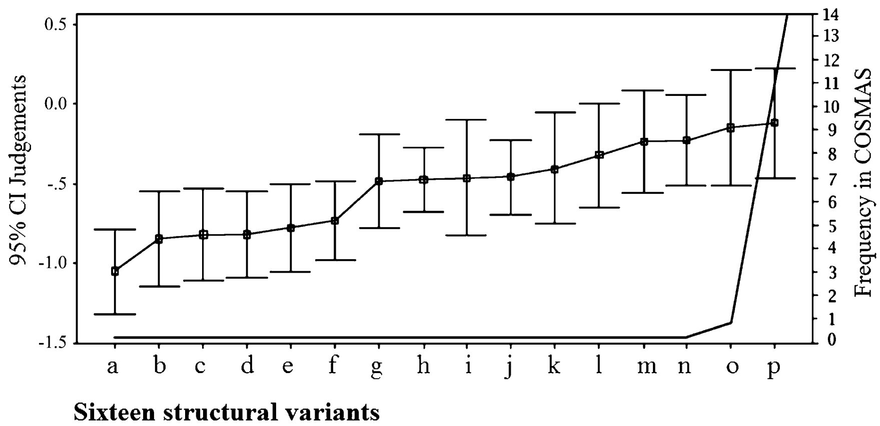


Figure 3: Well-formedness is the product of (the ease of) linguistic computation. Only the most easily computed structures will in practice occur. Occurrence frequency is thus a result of well-formedness.

3 Note that this is not supposed to be a full model of processing; many other factors affect output or parsing preferences too.

all encode the same meaning contents (from Featherston and Sternefeld, 2003). Only the strongest candidates appear in the corpus data, since they reliably win the competition to be output. The corpus frequencies are thus zero for all of the weaker candidates, whereas the judgements distinguish them clearly.[4]

It follows from these discussions that for a model of grammar based upon a merely binary opposition of *grammatical* and *ungrammatical*, both data types — binary judgements and corpus frequencies — are equivalently appropriate. However, for a model of grammar employing a gradient well-formedness construct, neither binary judgements nor corpus frequencies are optimal, because they do not provide interval scale data over all candidates.

2.5 Linguistic data in the Decathlon Model

We have suggested that both judgements and corpus data can tell us something about the way that we deal with language computation and production. We are now in a position to show how these can be combined in the Decathlon Model. The model owes its name to the athletic discipline Decathlon because the scoring system works rather in the same way as the interaction of linguistic constraints and output selection operate. Competitors in the decathlon participate in ten separate events which are scored separately. They are not at this stage competing directly with each other, rather they are trying to obtain their best possible performance on a scale which assigns point scores for times run, distances jumped and so on. It does not therefore matter how their performance in an individual event relates to those of the others; rather they need to perform at their personal best so as to gain the maximum number of points. The competition with the other athletes takes place at the level of the points gained over the ten sub-disciplines. In the end, the person with the highest number of points wins.

The grammar and production systems work in a similar way. Possible candidate structures are subject to the applicable grammatical constraints, and all constraint violations come at a cost. Put simply, if a structure breaks a rule, it causes processing load and the structure seems less acceptable. This perceived acceptability is the equivalent of the decathletes' points scores. The alternative structures compete to be selected for output on the basis of their well-formedness in the same way as athletes compete to win the gold medal on the basis of the points they have gathered in the individual events.

To continue the sports analogy, the architecture of generative grammars resembles the slalom, in that the candidates must pass through all the gates to win; missing

4 We refer to this as *the Iceberg Phenomenon*: only the best candidates are visible in frequency data, all others and the differences between them remain invisible because they do not cross the occurrence threshold (Featherston, 2008).

even one gate (violating even one restriction) causes categorical exclusion. Optimality Theory is like the high jump: the bar is put at a certain level and all competitors try to jump over it. All who fail are excluded, and the bar is put higher. When only one candidate remains, s/he becomes the optimal high jumper.

The Decathlon Model thus assumes two separate stages within what is commonly thought to be the unitary grammar and it is this distinction which causes judgements and occurrence frequencies to produce data patterns which are different, while generally corresponding in their indications of what the best structural alternatives for a given meaning output are. The key distinction is between well-formedness and occurrence. These are the output of two separate functions: Constraint Application and Output Selection.

The Constraint Application function applies grammatical constraints to candidate structures.[5] Constraint violations do not exclude candidates, rather they result in a violation cost in terms of apparent well-formedness being applied to the violating structure. Structures which violate multiple constraints are perceived to be less well-formed than those that violate fewer constraints. The result of this process is that each potential structure has a well-formedness value, which can be accessed introspectively as a (relative) judgement.

The next function is called Output Selection. Here realisations compete for output on the basis of their well-formedness ratings. It is plausible to think of this competition as being something like a race. That structure which is easiest to process is both introspectively perceived to be most well-formed, but also the earliest to finish its computation; it is therefore most likely to be output as linguistic production. We illustrate this two-part model in Figure 4.

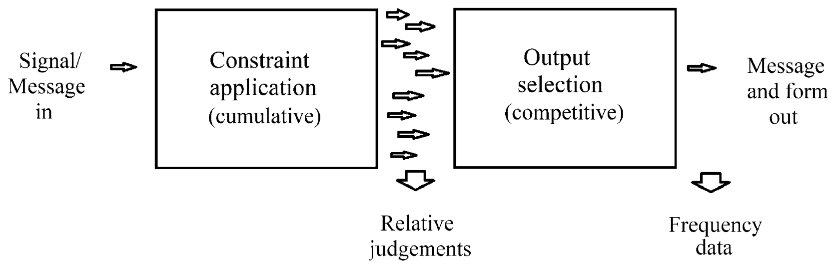


Figure 4: The Decathlon Model consists of two functions: Constraint Application and Output Selection. Our intuitions of well-formedness access the output of the first; occurrence frequencies are the result of the competition in the second.

5 This no doubt happens incrementally not globally, but we abstract away from this here.

We sum up the characteristics of the two functions in (8) and (9).

(8) *Constraint Application*
  a. applies rules
  b. registers constraint violations
  c. applies violation costs
  d. which collect cumulatively
  e. operates blindly and exceptionlessly.

(9) *Output Selection*
  a. selects structure for output
  b. on the basis of well-formedness
  c. competitively and probabilistically.

This model accounts for some otherwise puzzling facts. First, we can distinguish between degrees of well-formedness for sets of examples which are too unnatural ever to occur. Second, a structure can be judged to be fairly poor, but nevertheless appear in the language because it is the best of the set of available alternatives. An example might be (10). English doesn't have an easy inanimate alternative, so people sometimes use *whose*, although it seems poorly acceptable. The obvious alternatives to (10), *of which the leg* and *the leg of which* seem no better.

(10) ?? This is the chair whose leg broke.

2.6 Comparing architectures

The Decathlon Model is essentially an architecture of a grammar, rather than a grammar itself. For this reason it is helpful to compare it with other familiar models in order to appreciate its specificities. We shall compare it here with two simple cases, a standard generative grammar of the Principles and Parameters period and Optimality Theory (Prince and Smolensky, 1993). Readers can supply further comparisons for themselves when they have once recognized the variables under discussion. The factors which distinguish the grammar models are:

1. Constraint application
2. Violation cost application
3. Violation costs
4. Output selection

We shall clarify these headings by using them to describe a standard generative grammar model that all readers will be familiar with.

Generative grammars

Constraint application
In generative grammars, all constraints apply to all structures, blindly and exceptionlessly.

Violation cost application
All violations trigger the violation cost.

Violation cost
Generative grammar knows only one violation cost: ungrammaticality.

Output selection
Traditional generative grammar would predict that all grammatical structures can be selected for output and thus occur. Differences in frequency are due to independent factors irrelevant to the nature of the grammar (Chomsky, 1965).

This is a fairly simple grammar architecture. There are rules, which are binding, so any structure which breaks a rule is defined as ungrammatical, and is predicted to be inacceptable and thus excluded from occurrence.

The differences between models become apparent when we compare this with Optimality Theory (OT). For simplicity's sake we refer here to only the original OT model of Prince and Smolensky (1993).[6]

Optimality Theory

Constraint application
In OT, constraints are ranked, which effectively means that constraint application is ordered. Lower-ranked constraints may not be applied when there is only one single surviving candidate.

Violation cost application
Violation costs are applied only conditionally. If a given constraint does not distinguish between the multiple surviving constraints, then its violation cost is effectively not applied. If there is only one surviving candidate structure, either the constraint is not applied or else its violation cost is not applied.

Violation cost
OT knows only one violation cost: ungrammaticality.

Output selection
OT limits the number of candidates down to the most optimal one, which is therefore automatically selected for output.

6 Note that things change in the Minimalist Program, since this adds Economy as a major criterion. Economy is relative, not absolute. The Minimalist Program can thus be seen as a hybrid architecture between the first model and the OT model.

OT thus contains some significant differences to standard generative theory. In particular, it employs an additional parameter of variation, the constraint ranking in terms of dominance, which allows OT to capture more sophisticated constraint interaction behaviour than the basic generative model. The price of this is the inclusion of an additional parameter of constraint ranking.

OT differs from the standard generative model in several ways. First, OT is a competition model, in that candidate forms compete for output, while generative grammar is a decision model, in which candidate forms either are or are not grammatical. Second, while standard generative grammar applies all constraints to all structures automatically, in OT this is conditional upon the number of candidates remaining and whether the constraint differentiates the candidates. Third, generative grammar has a conception of well-formedness that is strictly distinct from occurrence, but OT does not clearly distinguish these. It is not obvious that OT makes any predictions about the introspective well-formedness of any competitor structures except for the optimal one (Keller, 2000). In the terms of our discussion of the relationship between judgements and frequencies as data above, it is plausible to suggest that OT is best seen as a model of what occurs (= corpus data), rather than what is well-formed (= judgements).

We now turn to the Decathlon Model.

Decathlon Model

Constraint application

Constraint application is simple. All constraints apply to all structures.

Violation cost application

All violations trigger the violation cost.

Violation cost

The violation costs vary between constraints: there are greater and lesser costs. If there are multiple costs, they add up. These costs are not directly linked to output.

Output selection

Output selection takes place in a competition over ratings of well-formedness. The most well-formed version is output. But there is a probabilistic element to this, so that good, but sub-optimal forms are sometimes output.

It will be seen that the Decathlon Model too introduces additional mechanisms compared to the basic generative model. The Decathlon Model chiefly varies from generative grammar in that it relies upon a gradient model of well-formedness. This more or less entails the interactions between constraint violation costs that we have argued for above, such as quantifiability and cumulativity. Output Selection as a competition over well-formedness is an additional aspect which depends upon a gradient model of well-fomedness. Apart from gradient grammaticality and its implications, the differences between the two architectures are not great.

But the Decathlon Model also has similarities with OT, above all, a parameter of something like constraint strength. But the detail of this is quite different in the two cases: OT's model of constraint ranking is essentially an ordinal scale, with strict domination of higher-ranked constraints over lower ones (though see e. g. Boersma and Hayes, 2001 for a tweaked version). The scale in the Decathlon Model on the other hand is an interval scale. Constraints do not dominate each other, instead their violation costs are added together. This is a very different type of interaction to that of Optimality Theory (but see 'Harmonic Grammar', a proto-form of Optimality Theory (Smolensky and Legendre, 2006)).

OT achieves its descriptive goals at the price of an additional parameter of variation, the constraint ranking. This plays a similar role to the parameter of violation costs in the Decathlon Model. But unlike constraint rankings in OT, these violation costs are obtained empirically, since they are precisely the values that informants assign them. This factor contributes to our characterization of the Decathlon Model as a tightly constrained account of the grammar.

3 Goals

The goals of the Decathlon Model are essentially the same as those of the generative grammar tradition, but they are perhaps weighted differently. Additionally, some common working assumptions of generative grammar are questioned and some further emphases set.

a. Descriptive adequacy
b. Explanatory adequacy
c. Empirical adequacy
d. Psychological reality
e. Capacity to accommodate variation

The first aim of a grammar must be *descriptive adequacy*; in this the Decathlon Model differs not one bit from mainstream generative thinking (Chomsky, 1965). The further aim of attaining *explanatory adequacy* is also shared. It is worth noting here that it is this ambition to account not only for the patterns of language observed but also for human language acquisition and development which marks Chomskyan grammar out as essentially a psycholinguistic enterprise.

The third goal is rather different. The term *empirical adequacy* is intended to convey two things. First, the data basis on which claims are made must be independent, objectively observable, replicable, and reliable. It will be clear that the intuitions of individuals do not in themselves fulfil this, though they are a very useful source of hypotheses (Featherston, 2011). Second, the grammar should correspond to what the empirical data shows. If carefully gathered data systematically exhibits a particular feature, our grammar should include this feature unless there are very solid reasons why not. One example of a feature which is not obviously compatible with the observable facts of language production and comprehension is the computation of a sentence structure from the bottom (and/or end) up, as in the merger model of the Minimalist Program. Language is generally understood and meanings are encoded incrementally (e. g. Phillips, 1996). This analysis would thus not find favour in the Decathlon Model.

The goal of *psychological reality* is an extension to the criterion of explanatory adequacy. A mentalistic grammar can only account for the observable facts of language acquisition if it is a model which is compatible with our wider understanding of how the mind works. Hard rules are an example of something which is not cognitively realistic. An example may make this plain: it belongs to the definition of a goat that it is a quadruped. If we imagine a goat with five or six legs, however, it doesn't stop being a goat. We are similarly fault-tolerant in language. My language doesn't permit multiple finite verbs in a single clause, it doesn't permit a verbal inflection *-s* in the first person, and it doesn't permit accusative subjects. But if someone says to me *Me wants eats an ice cream*, I can parse and interpret it. I conclude that grammar doesn't work with prohibitions and hard rules, for if it did, the sentence should crash my grammar.

The final goal is the capacity to model *variation*, both synchronous and diachronic. Both of these are fundamental features of language and should be feasible in our grammar architecture. Essentially, this boils down to an additional argument for gradient well-formedness, because these gradient models permit slow shifts from one form to another and periods of optionality between them.

4 Tools

There are two tools which I would like to address in this section, but they are of very different natures. The first is an empirical tool: the standard scale of well-formedness (Featherston, 2009; Gerbrich et al., 2016). The second is an analytical tool: the gradient well-formedness construct. We shall deal with these in turn.

4.1 Standard items

Gathering relative judgements yields data of fine detail, but it has the complicating effect that it does not deliver a judgement in absolute terms. Gradient data tells us how much better or worse each example is than the others, but not if they are finally so good that they might occur or so bad that they never would. Syntacticians want this additional information; they want to know where a given example or set of examples should be located relative to the broadly consensual categories of acceptable and unacceptable. And linguistic intuitions of well-formedness do contain this additional information. We can say, for example, that example A is better than example B, but that they are both quite unnatural.

The traditional method of doing this has been to refer to a single threshold point on the scale of well-formedness. Those items above the line are regarded as 'grammatical', those below it are labelled 'ungrammatical'. Since there are many examples whose placement relative to this line is known to all linguists and consensual, this threshold has a degree of intersubject validity. Nevertheless, there are plenty of discussions about individual examples and not infrequent disagreements and contradictions. Furthermore, this criterion offers us no help in distinguishing good from better or bad from worse.

To capture this part of our intuitions we have developed a tool which we refer to as *standard items*. These offer five anchor points along the scale of perceived well-formedness, relative to which other judgements can be located. These points are anchored by the use of multiple examples which give a value to each scale point. The anchoring items have been carefully selected and tested; they thus provide a basis for comparison so that the judgement score of a given sentence can more nearly be given an absolute value.

All linguists are familiar with Daniel Jones' cardinal vowels (Jones, 1967). These form a system of reference points relative to which other vowels are located. Their precise identity is not important; rather, they function as local reference points which permit us to give more exact judgements. Our perception of temperature is a good example of this. If you ask how warm it is outside, people can usually guess it to within few degrees. We know that above 20° it is a warm day, but below 17° you need a pullover, and below 12° you need a coat. Since these points are familiar, we can relate today's temperature to them with some accuracy. If on the other hand we present people with water between 40° and 60° and ask them to estimate its temperature, their guesses will be much less exact. People know that boiling water will scald them, they know that 37° is body temperature, but they have no points of reference between these, so they will judge much less accurately.

The system of standard items provides a set of *cardinal well-formedness values* like the cardinal vowels (see Gerbrich et al., 2016 for their selection and testing). These distinguish five degrees of well-formedness, from A (good) to E (bad). Each value is anchored by three example sentences, giving a total of fifteen examples, which we list in (11):

(11) Standard items representing cardinal well-formedness values

Naturalness value A

The patient fooled the dentist by pretending to be in pain.
There's a statue in the middle of the square.
The winter is very harsh in the North.

Naturalness value B

Before every lesson the teacher must prepare their materials.

Jack doesn't boast about his being elected chairman.
Jack cleaned his motorbike with which cleaning cloth?

Naturalness value C
Hannah hates but Linda loves eating popcorn in the cinema.
Most people like very much a cup of tea in the morning.
The striker must have fouled deliberately the goalkeeper.

Naturalness value D
Who did he whisper that had unfairly condemned the prisoner?
The old fisherman took her pipe out of mouth and began story.
Which professor did you claim that the student really admires him?

Naturalness value E
Historians wondering what cause is disappear civilization.
Old man he work garden grow many flowers and vegetable.
Student must read much book for they become clever.

The use of these standard items as known fixed points for introspective judgements provides multiple comparison points, which makes discussions about well-formedness easier and should improve accuracy. Syntactician A might say "I find this one a bit degraded, perhaps a C or a little better", while syntactician B could reply "I think you are being a bit harsh. I would make it a B minus". And each would know exactly what the other meant.

These items can thus provide something like absolute well-formedness values for relative judgements. Some people like to think of them as representing the values that syntacticians have traditionally given example sentences, as in (12).

(12) Naturalness value A: ✓✓ (very natural)
Naturalness value B: ✓ (natural)
Naturalness value C: ?
Naturalness value D: ??
Naturalness value E: *

The standard items are also useful as filler items in judgement experiments, so that we obtain both relative and (nearly) absolute values. We provide an example of their use in an experiment with Miriam Gänzle (Gänzle, 2017), looking at whether the insertion of a resumptive pronoun can prevent or ameliorate a violation of the Empty Category Principle (ECP) (Lasnik and Saito, 1984). The study therefore tested conditions with subject vs object extractions, from an interrogative vs from a declarative complement clause, with vs without resumptive pronouns at the trace site of the long movement – (13). This experiment built upon the work of Alexopoulou and Keller (2007) and Hofmeister and Norcliffe (2013), who investigated whether a resumptive pronoun could have the function of 'saving' an illicit extraction structure.

(13) a. *Declarative wh-subject > wh-object*
Which student$_j$ do you say he/t$_j$ should read which book?

b. *Declarative wh-object > wh-subject*
Which book$_i$ do you say which student should read it/t$_i$?

c. *Interrogative wh-subject > wh-object*
Which student$_j$ do you wonder which book$_i$ he/t$_j$ will read t$_i$?

d. *Interrogative wh-object > wh-subject*
Which book$_i$ do you wonder which student$_j$ t$_j$ will read it/t$_i$?

We assume the following violations for the conditions without resumptive pronouns (Lasnik and Saito, 1984):

Condition a: No violation
Condition b: Subjacency violation
Condition c: ECP violation
Condition d: Subjacency and ECP violations

Looking at the results chart in Figure 5, we see that the extractions with empty gaps become increasingly unacceptable as subjacency and the ECP predict. Interestingly, the examples with resumptive pronouns do not get worse, so that the very worst condition with a resumptive pronoun is no worse than its minimal pair without a resumptive pronoun, which may help to account for the occurrence of resumptive pronouns in such cases. This is a very interesting data set, but our concern here is with the stan-

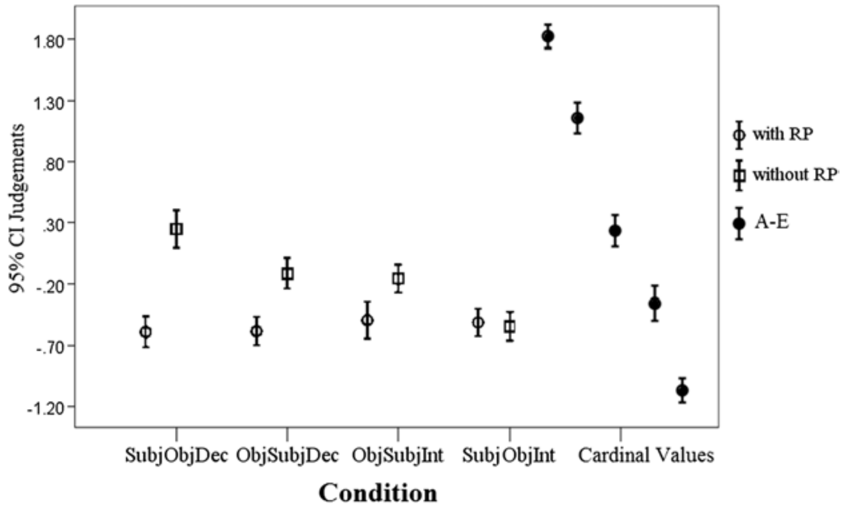


Figure 5: Results of experiment on the ECP and resumptive pronouns. The experimental conditions are Extraction {WH-SUBJ, WH-OBJ}, Clause Type {DECLARATIVE, INTERROGATIVE}, Gap Behaviour {RESUMPTIVE, NO RESUMPTIVE}.

dard items, whose results are seen on the right-hand side of the chart. Their results show a spread of well-formedness values.

The standard items show that the experimental conditions are of very marginal well-formedness. Even the very best multiple wh-question in the experiment: ‘Which student do you say should read which book?’, which would standardly be regarded as grammatical, is no better than the C value on the scale of well-formedness. The others descend to the D level and below. The examples with resumptive pronouns are never better than a D. At this level, there can be no question of their ‘saving’ examples with illegal extractions (as e. g. Alexopoulou and Keller (2007) and Hofmeister and Norcliffe (2013) discuss), since you cannot ‘save’ a structure to a sub-D level.

However, the flat pattern of the conditions with resumptive pronouns suggests that they are able to neutralize extraction violations such as Subjacency and the ECP. While their insertion comes at a significant cost in perceived well-formedness, the overt nominals at gap positions do appear to make a difference to the extraction structures, stopping the additional violation effects. One interpretation might be that the structures with resumptive pronouns are no longer extractions, so that movement constraints are not applicable. Another approach would be to see the resumptive pronoun as affecting the visibility of the base position of the dependency for syntactic processes, in a similar way that lexical government makes a gap position visible within the ECP.

We cannot continue this discussion here, but we can summarize that example studies like these can show the value of having an experimentally-determined set of well-formedness values, provided by the use of the standard items.

4.2 Gradient well-formedness

The adoption of a gradient construct of well-formedness is a very powerful tool which has far-reaching implications for our understanding of the grammar. We may distinguish two groups of effects: those relating to the data and those relating to the analytical possibilities.

The first effect of a gradient model of well-formedness is that it allows us to treat as ‘grammatical’ lots of phenomena which were previously excluded because they did not by themselves alone cause a structure to be so bad that it was effectively excluded from the language. A categorical model of well-formedness assumes only two values: grammatical and ungrammatical. Any constraint whose violation does not clearly shift an example from the one group into the other group cannot be accommodated in such a model; such effects must be accounted for as processing, or performance effects, or style, or markedness, since these modules of explanation could deal with non-categorical phenomena.

The adoption of a gradient model of well-formedness means that we need not attribute these non-categorical phenomena to external, non-grammatical mechanisms and processes. Instead they can remain located within the grammar, which thus has its coverage extended. Since exactly this gradient model is what we find represented in the primary data anyway, we are able to encompass more of the data and account for more phenomena simply by discarding the unmotivated assumption that violation costs must be fatal. This is thus a win-win step.

4.2.1 Apparent grammaticality differences

There are also analytical advantages of a gradient model of grammar, some of which we will illustrate here. The first example concerns some phenomena in German which seem to involve clausal boundaries being transgressed. Bech (1983) termed this *Kohärenz* ('coherence'), because two verbal forms seem to cohere and form a complex predicate.

The first phenomenon is 'long passive'. This can (marginally) occur in structures where a verb takes an infinitival complement clause and the two verbs can be interpreted as one complex predicate. If the matrix clause is passivized, then the object of the embedded clause apparently becomes the subject of a single merged passivized clause – (14).

(14) Der Lehrling$_i$ wurde versucht t$_i$ auszunutzen.
the$_{nom}$ apprentice was tried t to.exploit
'(It) was tried to exploit the apprentice.'

The other phenomenon is known as the 'third construction'. If a verb has a clausal complement consisting of an infinitive verb and its object, then this complement could theoretically appear in two positions: either inside its mother clause in the normally subcategorized position left-adjacent to the verbal head, or extraposed outside the clause. Both positions are feasible – (15).

(15) a. … weil der Meister den Lehrling auszunutzen versucht.
… because [the master]$_{nom}$ [the apprentice]$_{acc}$ to.exploit tried
'… because the master tried to exploit the apprentice'
b. … weil der Meister versucht, den Lehrling auszunutzen.
… because [the master]$_{nom}$ tried [the apprentice]$_{acc}$ to.exploit
'… because the master tried to exploit the apprentice'

The third construction is said to occur when the object of the infinitive verb seems to remain within the matrix clause, while the infinitive verb itself seems to be extraposed – (16).

(16) … weil der Meister den Lehrling versucht auszunutzen.
… because [the master]$_{nom}$ [the apprentice]$_{acc}$ tried to.exploit
'… because the master tried to exploit the apprentice'

An important point of debate about these marked constructions is whether they can be licensed by the same structural mechanism, *Kohärenz*. An convincing argument against this is the observation that there appear to be matrix verbs which license one but not the other. This would seem to militate against these structures responding to a single factor.

We tested this finding in a judgement study (for details Sternefeld and Featherston, 2004). Informants provided relative judgements of these two German constructions embedded under a range of different matrix verbs. The results are presented in Figure 6. The vertical axis shows the judgements of perceived naturalness, higher scores meaning more acceptable. Along the horizontal axis we see the various conditions tested. On the left-hand side we see the results of the standard items from A (best) to E (worst). Then follow our experimental conditions with the sixteen matrix verbs, ordered by their perceived well-formedness.

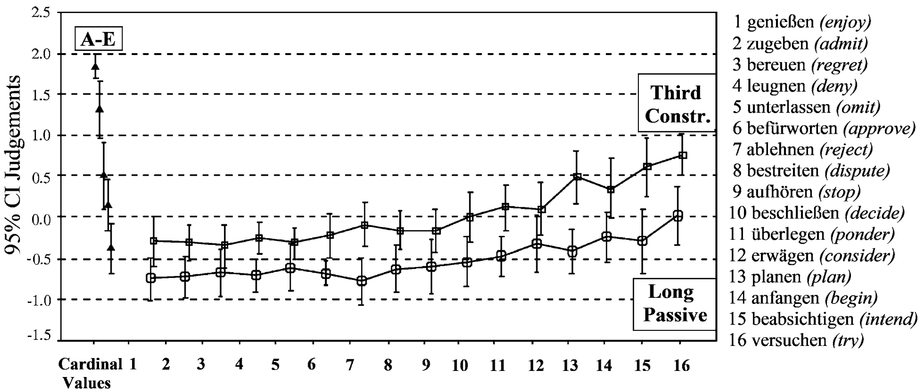


Figure 6: Pattern matching as evidence for grammatical structure.

Various things are apparent. First, none of our experimental conditions are very good – they vary from below E to about C on the scale provided by the standard items. Second, the third construction is consistently better than the long passive. Third, the matrix verb affects the perception of well-formedness of these marked structures: only certain verbs such as *planen* ('plan'), *anfangen* ('begin'), *beabsichtigen* ('intend'), *versuchen* ('try') permit our two structures to be remotely acceptable.

But here we can also see that the two structures respond to the lexical features of the verbs in the same way. Those verbs which produce bad scores for the one also produce bad scores for the other. This makes it immediately plausible that long passive and the third construction are licensed by the identical lexical factor, the ability of the verb to merge into a 'coherent' verbal cluster.

We can also see why linguists might not have recognized this. The assumption of a binary construct of well-formedness is like drawing a horizontal line across the data chart; all conditions above the line are regarded as undifferentially good, and all structures below the line are regarded as undifferentially bad. This will result in the third construction being good and the long passive being bad whenever the line falls between the values for these conditions for a given verb. This yields an apparent, but on closer inspection plainly delusive impression that there are verbs for which the licensing of long passive and third construction do not match.

We can thus see that the assumption of a binary well-formedness construct will cause us to draw an erroneous conclusion. A gradient well-formedness construct is thus a valuable tool to permit us to carry out grammatical analyses in greater detail and more accurately because it can capture the full complexity of the evidence.

4.2.2 Multiple effects

Our next example of the analytical benefit of gradient well-formedness relates to a cross-linguistic contrast in the Superiority effect (Chomsky, 1973). In an example with multiple wh-items, the structurally superior of them must raise, not the inferior one. This is fairly clear in English – (17).

(17) a. Who kissed who(m) at the party last night?
b. *Who(m) did who kiss at the party?

In German, on the other hand, the superiority-violating structure seems less bad – (18). On this basis, the consensus has been that this constraint does not apply in German or applies only to a limited degree (e. g. Grewendorf, 1988; Haider, 1993).

(18) a. Wer hat wen gestern Abend auf der Party geküsst?
Who has whom yesterday evening at the party kissed
'Who kissed who at the party last night?'
b. ?Wen hat wer gestern Abend auf der Party geküsst?
Whom has who yesterday evening at the party kissed
'Whom did who kiss last night at the party?'

We carried out a series of experimental judgement studies to examine the construction in the two languages (for details Featherston, 2005). In the experiments six different types of wh-items were used, always two per sentence. In each experimental condition, one was raised into sentence-initial position and one remained at its base site. We report the results first by in-situ wh-items, see Figure 7.

The essential result is simple: the in-situ subjects in both German (*wer*) and English (*who*) were judged worse than all the others.[7] This may be seen as confirmation that German knows some form of the Superiority effect.

7 Note that the set of wh-items tested varied slightly between German and English. The English experiment has two forms of the subject, both a bare wh-item *who* and also a discourse-linked *which N*;

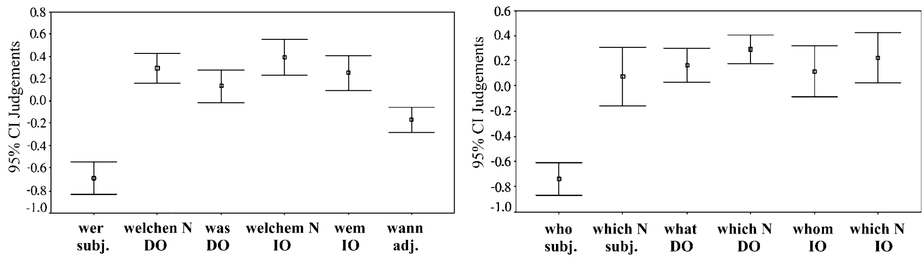


Figure 7: Multiple wh-questions in German and English, by in-situ wh-items.

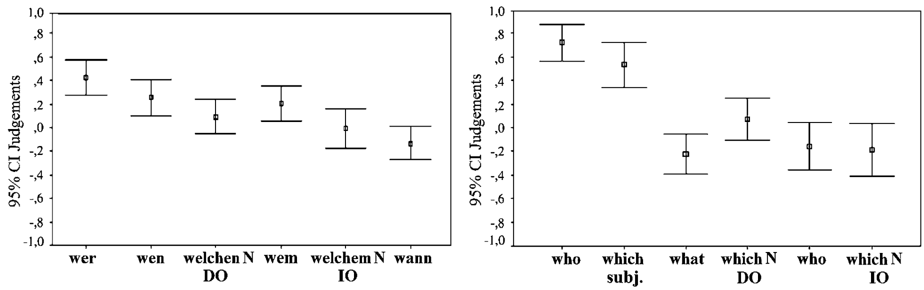


Figure 8: Multiple wh-questions in German and English, by raised wh-items.

A different picture emerges if we look at the results distinguished by the raised wh-item in Figure 8. While raised subjects in English are judged clearly better than all other raised wh-items, this effect is only marginally apparent in German.

Purely descriptively, this result is not surprising: English has a strong preference for starting a sentence with a subject, while German, as a V2 language, is much more tolerant of inversions and object topicalization. But the implication for the nature of the Superiority effect is striking: the effect in English is not one single factor but the cumulative effect of two separate factors. When there are two wh-items, we perceive a positive effect when the wh-subject is in the raised position as in Figure 8 (presumably: because this conforms to SVO). But we perceive a separate negative effect when the wh-subject is in the in-situ position – Figure 7 (because another wh-item has moved over the wh-subject). These are two separate effects, but they occur in so nearly complementary distribution that people have evidently mistaken them for a single effect.

Crucially, German has a similar dispreference for in-situ wh-subjects – Figure 7, but barely knows the positive preference for raised wh-subjects – Figure 8. The apparent relative weakness of the effect in German is thus due to it having only one of the two components which makes up the apparent robust English 'Superiority effect'. This

---

the German experiment has only the first of these. German also has a temporal adjunct *wann* ('when'), which was omitted in the English study.

offers a ready explanation of the doubts that German linguists have expressed about its applicability to German: they are partly right.

Our data makes it clear that the English Superiority effect is not a unitary factor, but rather the combination of two factors which apply with almost complementary distribution. Only examples with direct and indirect objects as wh-items such as (19) provide cases where neither of these apply and thus allow us to capture the well-formedness of examples with neither raised nor in-situ wh-subjects.

(19) a. Who did you send what?
b. What did you send who?

But what is important for us here is the effect of the model of well-formedness adopted. The traditional binary model of well-formedness cannot capture cumulative effects. A structure is either good or bad and undifferentiatedly good or bad. A gradient model of well-formedness allows us to see that an apparent homogeneous effect may in fact be made up of multiple strictly separate effects (for another example see Hofmeister et al., 2015). Such a model therefore permits theoretical progress which a categorical model does not.

5 Evaluation

The Decathlon Model is to be conceived as a psycholinguistic theory. That means that the evaluation criteria that it is designed to fulfil are related to the way that humans store and process language information and how they apply this in language use. The evaluation criteria are thus those listed in Goals above.

6 Sample analysis

In order to allow a direct comparison of the different grammar models in this handbook, each model is invited to provide an analysis of this example – (20)

(20) After Mary had introduced herself to the audience, she turned to a man she had met before.

There are various things that could be said about this example, but we shall concentrate on just one aspect here: the lack of a relative marker in the relative clause. We shall discuss the question why we do not find '...a man that she had met' or '...a man who she had met'. This reduces the example under discussion to (21).

(21) She turned to a man who/that/Ø she had met before.

So the question is: What can the Decathlon Model say about these alternatives, if indeed they are alternatives? Under what circumstances is each of these options possible or preferred? To provide a partial answer to this I should like to present the results of an experiment carried out with Katharina Salkowski (Salkowski, 2014). This offers an illustration of the perspective on such questions within the Decathlon Model.

Our experiment looked at subject and object restrictive relative clauses with the relative markers *who* and *that*, building particularly on the work in Wasow et al. (2011), Race and MacDonald (2003), and Gordon et al. (2004). A third factor varied was the length of the other argument within the relative clause. It appeared either as a pronoun, as a short NP, or as a long NP. It was thus a 2x2x3 design with the following variables and values:

RC Type {ObjRC, SubjRC}
RC Marker {that, who}
NP Length {Pronoun, smallNP, longNP}

The experimental materials were as illustrated in (22). These were produced in twelve lexical variants and then divided into six experiment versions following the latin square procedure, so that each person saw each condition and each lexical variant exactly twice. These were randomly ordered and mixed with another twenty filler items which were designed to span the full range of accessible naturalness, so that each participant saw a total of 44 sentences. The experiment versions were presented randomly to 52 native speakers of English, who were instructed to judge their naturalness on a seven-point scale.

(22) I think you know the teacher …
  a. …that/who I questioned at the school.
  b. …that/who questioned me at the school.
  c. …that/who the parents questioned at the school.
  d. …that/who questioned the parents at the school.
  e. …that/who the overbearing parents questioned at the school.
  f. …that/who questioned the overbearing parents at the school.

The results show several interesting findings. Let us first of all consider Figure 9. This graph offers us assurance that our test is capturing the perception of well-formedness as intended. We can be sure of this as the results reveal two known effects: first, that subject relative clauses are generally perceived to be easier than object relative clauses, and second, that examples which are longer, because their RCs contain a longer NP, receive slightly lower ratings. In the light of this, we can address the further findings about relativizers.

We next consider Figure 10. Here we see that there is an interaction between the relativizers *that* and *who* and the grammatical role of the extracted element. If the head of the RC is the subject, *who* is preferred to *that*. If it is the object, then *that* scores

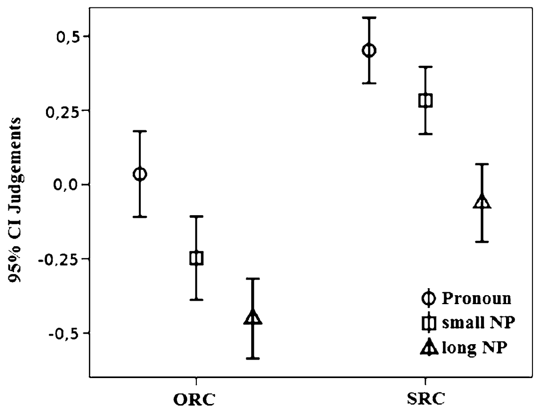


Figure 9: Judgements of subject and object relative clauses, differentiated by the length of NP inside the RC.

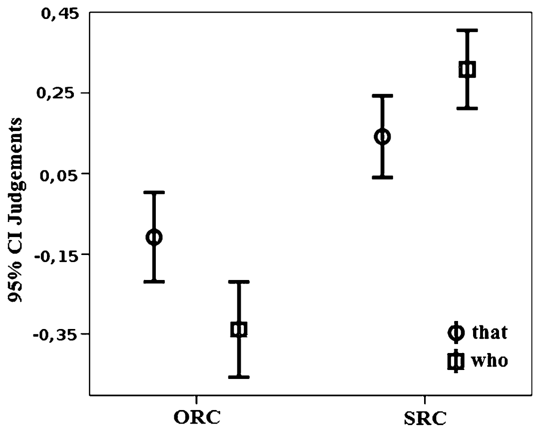


Figure 10: This chart shows the interaction of relative clause type object RC, subject RC and relativizer choice.

better than *who*. One might speculate whether this preference is an avoidance effect because of the requirement to choose between *who* and *whom* in the object case (cf. Quirk et al., 1985, 1251), or whether it has more to do with the prototypical animate nature of subjects and inanimate nature of objects. This might cause *who* to be expected to be a subject and *that* to be expected to be an object. We have found animacy effects in relativizers in other cases, but would note that the effect here can only be related to the prototypical nature of subjects and objects, since both were always animate in these experimental materials.

The final aspect of the results is illustrated in Figure 11. We see that the choice of relativizer is also affected by the length of the NP within the RC. While the relativizer

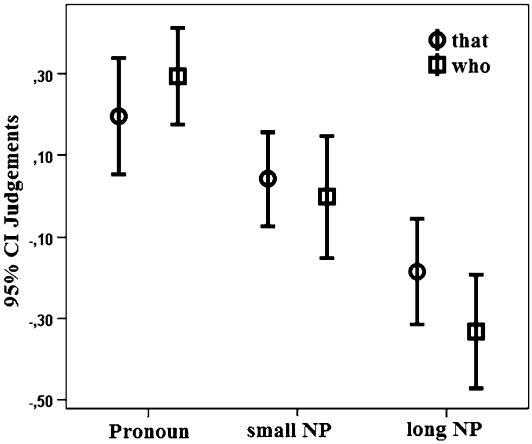


Figure 11: This chart shows the perceived well-formedness of the two relativizers in relation to the length of the NP in the RC.

who is preferred with short NPs such as pronouns, increasing NP length reverses this association, so that *that* is favoured as relativizer when the NP is 6–7 syllables long, as in our materials here. We cannot debate the cause of this in detail here, but it seems likely that it can be attributed to the quantity of information delivered by the different relativizers on the one hand and the NP on the other hand. The relativizer *that* is not marked for case or animacy, and it is superficially identical to both the demonstrative *that* and the complementizer *that*; it is thus extremely unspecific. The relativizer *who*, on the other hand, can only be an animate pronominal form and thus tells us more about the following structure. It is plausible that there is an advantage to having the more informative relativizer when the content of the RC is less informative, and vice versa.

This section is not about relativizer choice but should provide a sample analysis within the Decathlon Model. Our aim here was to illustrate the sort of detailed question that can profitably be examined if the structural assumptions and methodological priorities of the Decathlon Model are applied. A grammar model which only allows for two values, grammatical and ungrammatical, can say nothing about the sort of questions of relativizer choice that we have addressed in this section; it cannot even capture the distinctions in perceived well-formedness that underlie our discussion. Instead it is forced to assume a line across the middle of each of our charts, and claim that those conditions which fall above that line have one theoretical status and those that fall below have another. We can do better.

Nor does the traditional categorical model of grammar allow us to capture such features of the data as cumulativity or interactions. In Figure 9 we saw that the cost in well-formedness due to RC length was independent of the cost of type of RC; the two factors added together cumulatively. In figures 10 and 11 we saw a quite different relationship between two factors, that of interaction. For example in Figure 11 we seem to see that the informational content in the relativizer interacts with informational content in the NP within the RC: the more information we have in the NP, the less we need in the relativizer, and vice versa. The data is complex, we therefore need a well-formedness model which can encompass this complexity. The Decathlon Model is a first step towards this goal.

7 Summary

The Decathlon Model is an attempt develop a grammar model in the generative tradition which is empirically grounded. The collection of high quality data and the close inspection of the patterns found in the data suggest that corpus frequencies and introspective judgements are closely related, but judgements are ontologically precedent to occurrence frequencies, so that judgement data patterns are a better guide to the architecture of the grammar.

Looking then at this data, we see support for many aspects of the Chomskyan family of grammars, but one important feature which contradicts mainstream assumptions. The pattern of well-formedness that we see is consistently gradient, not categorical, which suggests that the categorical model of well-formedness is a simplifying assumption.

Considering what sort of rule system could yield such a pattern, we suggest that grammar 'rules' have the form of conditional constraints, with both a description of the structure causing a violation and a specification of what cost in well-formedness a violation incurs. This requires that our conception of a grammar have an additional component, that of a *parameter of violation costs*.

References

Alexopoulou, Theodora & Frank Keller. 2007. Locality, Cyclicity, and Resumption: At the Interface between the Grammar and the Human Sentence Processor. *Language* 83(1). 110–160.

Bader, Markus & Jana Häussler. 2010. Toward a model of grammaticality judgments. *Journal of Linguistics* 46(2). 273–330.

Bech, Gunnar. 1983. [1955/1957]. *Studien über das deutsche Verbum infinitum, Band 1* (Linguistische Arbeiten 139). Tübingen: Niemeyer.

Behaghel, Otto. 1909/1910. Beziehungen zwischen Umfang und Reihenfolge von Satzgliedern. *Indogermanische Forschungen* 25. 110–142.

Boersma, Paul & Bruce Hayes. 2001. Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry* 32. 45–86.

Bornkessel-Schlesewsky, Ina & Matthias Schlesewsky. 2007. The wolf in sheep's clothing: Against a new judgement-driven imperialism. *Theoretical Linguistics* 33(3). 319–333.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, Massachusetts: MIT Press.

Chomsky, Noam. 1973. Conditions on transformations. In Stephen Anderson & Paul Kiparsky (eds.), *A Festschrift for Morris Halle*, 232–286. New York: Holt, Reinhart & Winston.

Chomsky, Noam. 1986. *Knowledge of Language: Its Nature, Origin, and Use*. New York: Praeger.

Featherston, Sam. 2005. Universals and grammaticality: Wh-constraints in German and English. *Linguistics* 43(4). 667–711.

Featherston, Sam. 2007. Data in generative grammar: The stick and the carrot. *Theoretical Linguistics* 33(3). 269–318.

Featherston, Sam. 2008. Thermometer judgements as linguistic evidence. In Claudia Maria Riehl & Astrid Rothe (eds.), *Was ist linguistische Evidenz?*, 69–90. Aachen: Shaker Verlag.

Featherston, Sam. 2009. Why linguistics needs boiling and freezing points. In Sam Featherston & Susanne Winkler (eds.), *The Fruits of Empirical Linguistics. Volume 1: Process*, Berlin: de Gruyter.

Featherston, Sam. 2011. Reply: Distinguishing lexical and syntactic exceptions. In Horst Simon & Heike Wiese (eds.), *Expecting the Unexpected: Exceptions in Grammar*, 335–337. Berlin: de Gruyter Mouton.

Featherston, Sam & Wolfgang Sternefeld. 2003. The interaction of factors in judgements of reflexive structures: Data from object coreference in German. In Lutz Gunkel, Gereon Müller & Gisela Zifonun (eds.), *Arbeiten zur Reflexivierung* (Linguistische Arbeiten 481), 25–50. Tübingen: Niemeyer.

Gänzle, Miriam. 2017. Resumptive Pronouns and the Empty Category Principle. Dissertation, University of Tübingen.

Gerbrich, Hannah, Vivian Schreier & Sam Featherston. 2016. Standard items for English judgement studies: Syntax and semantics. ms Tübingen University.

Gordon, Peter, Randall Hendrick & Marcus Johnson. 2004. Effects of noun phrase type on sentence complexity. *Journal of Memory and Language* 51. 97–114.

Grewendorf, Günther. 1988. *Aspekte der Deutschen Syntax. Eine Rektions-Bindungs-Analyse*. Tübingen: Narr.

Haider, Hubert. 1993. *Deutsche Syntax – Generativ*. Tübingen: Narr.

Hawkins, John A. 1994. *A Performance Theory of Order and Constituency*. Cambridge: Cambridge University Press.

Hofmeister, Phillip, Peter W. Culicover & Susanne Winkler. 2015. Effects of processing on the acceptability of 'frozen' extraposed constituents. *Syntax* 18(4). 464–483.

Hofmeister, Phillip & Elisabeth Norcliffe. 2013. Does resumption facilitate sentence comprehension? In Phillip Hofmeister & Elisabeth Norcliffe (eds.), *The Core and the Periphery: Data-Driven Perspectives on Syntax Inspired by Ivan A. Sag*, 225–246. Stanford, CA: CSLI Publications.

Jones, Daniel. 1967. *An Outline of English Phonetics*. Cambridge: Heffer.

Keller, Frank. 2000. Gradience in grammar: Experimental and computational aspects of degrees of grammaticality. Ph.D. thesis, Edinburgh University.

Lasnik, Howard & Mamoru Saito. 1984. On the Nature of Proper Government. *Linguistic Inquiry* 15(2). 235-289.

McRoy, Susan & Graeme Hirst. 1990. Race-based parsing and syntactic disambiguation. *Cognitive Science* 14. 313–353.

Pesetsky, David. 1987. *Wh*-in-situ: Movement and unselective binding. In Eric Reuland & Alice ter Meulen (eds.), *The Representation of (In)definiteness*, 98–129. Cambridge, Massachusetts: MIT Press.

Phillips, Colin. 1996. Order and Structure. PhD dissertation, MIT. Distributed by MIT Working Papers in Linguistics.

Prince, Alan & Paul Smolensky. 1993. Optimality Theory. Constraint Interaction in Generative Grammar. Rutgers University Center for Cognitive Science Technical Report 2.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.

Race, David & Maryellen MacDonald. 2003. The use of "that" in the production and comprehension of object relative clauses. *Proceedings of the 26th Annual Meeting of the Cognitive Science Society*. 946–951.

Salkowski, Katherina. 2014. Conflicting constraints: An experimental approach. Dissertation, University of Tübingen.

Schütze, Carson. 1996. *The Empirical Base of Linguistics: Grammaticality Judgements and Linguistic Methodology*. Chicago: University of Chicago Press.

Smolensky, Paul & Géraldine Legendre. 2006. *The Harmonic Mind: From Neural Computation to Optimality-Theoretic Grammar*. Cambridge, MA: MIT Press.

Sternefeld, Wolfgang & Sam Featherston. 2004. Long passives and the third construction: Tests and evidence on coherence in German. Talk given in Linguistics Colloqium, University of Konstanz, Germany.

Wasow, Thomas. 2002. *Postverbal Behavior*. Standford: CSLI Publications.

Wasow, Thomas, Florian Jaeger & David Orr. 2011. Lexical variation in relativizer frequency. In Horst Simon & Heike Wiese (eds.), *Expecting the Unexpected: Exceptions in Grammar*, 175–195. Berlin: de Gruyter Mouton.

Zwicky, Arnold. 2002. I wonder what kind of construction that this example illustrates. In David Beaver, Luis Casillas Martínez, Brady Clark & Stefan Kaufmann (eds.), *The Construction of Meaning*, 219–248. Stanford: CSLI Publications.

Norbert Hornstein

7 The Stupendous Success of the Minimalist Program

Abstract: This chapter is a full-throated argument that the Minimalist Program has been tremendously successful in answering the questions it has posed. Perceptions to the contrary rest on misunderstanding what these questions are. Once clearly identified, the chapter outlines the central minimalist thesis, the Merge Hypothesis, and outlines the reasons for thinking that versions thereof (the Extended Merge Hypothesis and the Fundamental Principle of Grammar that it suggests) are fertile avenues for explaining many of the features of FL/UG that previous GB inquiry empirically established. The chapter also attempts to explain why the Minimalist Program has received a hostile reception. This is due to the fact that its central question makes prominent a long-established division within linguistics which had been obscured until the Minimalist inquiry brought it to prominence

1 Data and goals

1.1 The generative world view

What is linguistics about? What is its subject matter? Here are two views.

One standard answer is "language". Call this the "***la**nguistic* (LANG) perspective". Languists understand the aim of a theory of grammar to describe the properties of different languages and identify their common properties. Languists frequently observe that there are very *few* properties that *all* languages share. Indeed, in my experience, the LANG view is that there are no such universal features (i. e. ones that hold without exception) and that languages can and do vary arbitrarily and limitlessly. In this sense of 'universal,' LANGers assume that almost none exist, or if there are any, then they are of the Greenbergian variety, more often statistical tendencies than categorical absolutes.

There is a second answer to the question, one associated with Chomsky and the tradition in Generative Grammar (GG) his work initiated. Call this the "***li**nguistic* (LING) perspective". Until very recently, linguists have understood grammatical theory to have a pair of related objectives: (i) to describe the mental capacities of a native speaker of a particular language L (e. g. English) and (ii) to describe the meta-capacity that allows any human to acquire the more specific mental capacities underlying a native speaker facility in a particular L (i. e. the meta-capacity required to acquire a particular grammar G). LINGers, in other words, take the object of study to be two kinds of mental states, one that grammars of particular languages (i. e. G_L) describe

https://doi.org/10.1515/9783110540253-007

and one that "Universal Grammar" (UG) describes. UG, then, names not Greenbergian generalizations about languages but design features of human mental capacity that enable them to acquire G_Ls. For linguists, the study of languages and their intricate properties is useful exactly to the degree that it sheds light on both of these mental capacities. As luck would have it, studying the products of these mental capacities (both at the G and UG level) provides a good window on these capacities.

The LANG vs LING perspectives lead to different research programs based on different ontological assumptions.[1] LANGers take language to be primary and grammar secondary. G_Ls are (at best) generalizations over regularities found in a language (often a more or less extensive corpus or lists of "grammaticality" judgments serving as proxy).[2] For LINGers, G_Ls are more real than the specific linguistic objects they generate, the latter being an accidental sampling from an infinite set of *possible* legitimate objects.[3] On this view, the aim of a theory of a G_L is, in the first instance, to describe the *actual* mental state of a native speaker of L and thereby to indirectly circumscribe the *possible* legitimate objects of L. So for LINGers, the mental state comes first (it is more ontologically basic), the linguistic objects are its products and the etiology of those that publically arise (or are elicited in some way) only partially reflect the more stable, real, underlying mental capacity. Put another way, the products are *interaction effects* of various capacities and the visible products of these capacities are the combination of their adventitious complex interaction. So the actual public products are "accidental" in a way that the underlying capacities are not.

LANGers disagree. For them the linguistic objects (be they judgments, corpora, reaction times) come first, G_Ls being inductions or "smoothed" summaries of these more basic data. For LINGers the relation of a G_L to its products is like the relation

1 For reasons I cannot fathom, some have taken the distinction between LANG and LING to be invidious, the LANG moniker being a slight of some sort. I apologize (well sort of) to all those so discomfited. I chose the terms for their mnemonic utility. LANGers take ***lang***uage as cynosure. LINGers take ***i***-language as object of study. Neither type of study is inherently more valuable than the other. However, they are different and recognizing this is, in my view, important to appreciating the current state of play within professional linguistics, as I will argue in what follows.

2 There are few more misleading terms in the field than "grammaticality judgment". The "raw" data are better termed "acceptability" judgments. Native speakers can reliably partially order linguistic objects with regard to relative acceptability (sometimes under an interpretation). These acceptability judgments are, in turn, partial reflections of grammatical competence. This is the official LING view. LANGers need not be as fussy, though they too must distinguish core data from noise. The reason that LANGers differ from LINGers in this regard reflects their different views on what they are studying. I leave it to the reader to run the logic for him/herself.

3 The idea that a language is a *set* of expressions should not be taken too seriously. There is little reason to think that languages are sets with clear in/out conditions or that objects that G_Ls generate are usefully thought of as demarcating the boundaries of a language (see, for example, Chomsky's ruminations on degrees of grammaticality). In fact, LINGers don't assume that the notion of a *language* is clear or well conceived. What LINGers do assume is that native speakers have a sense of what kinds of objects their native capacities extend to and that this is an open ended capacity that is (indirectly) manifest in their linguistic behavior (production and understanding) of linguistic objects.

between a function and its values. For a LANGer it is more like the relation between a scatter plot/histogram and the smoothed distributions that approximate it (e. g. a normal distribution).

LINGers go further: even G_Ls are not *that* real. They are less real than UG, the meta-capacity that allows humans to acquire G_Ls. Why is UG more "real" than G_Ls? Because in a sense that we all understand, native speakers only *accidentally* speak the language they are native in. Basically, it is a truism universally acknowledged that *any* kid could have been native in *any* language. If this is true (and it is, really), then the fact that a particular person is natively proficient in a particular language is a historical accident. Indeed, just like the visible products of a G_L result from a complex interaction of many more basic sub-capacities, a particular individual's G_L is also the product of many interacting mental modules (memory size, attention, the particular data mix a child is exposed to and "ingests", socio-economic status, the number of loving hugs and more). In this sense, every G_L is the product of a combination of accidental factors and adventitious associated capacities combined with the meta-capacity for building G_Ls that humans as a species come equipped with.

If this is right, then there is no *principled* explanation for why it is that Norbert Hornstein (NH) is a linguistically competent speaker of Montreal English. He just happened to grow up on the West Island of that great metropolis. Had NH grown up in the East End of London he would have been natively proficient in another "dialect" of English and had NH been raised in Beijing then he would have been natively proficient in Mandarin. In this very clear sense, then, NH is only *accidentally* a native speaker of the language he actually speaks (i. e. has acquired the particular grammatical sense (i. e. G_L) he actually has) though it is no accident that he speaks *some* native language. At least not a biological accident for NH is the *type* of animal that *would* acquire *some* G_L as a normal matter of course (e. g. absent pathological conditions) if not raised in feral isolation. Thus, NH is a native speaker of *some* language as a matter of biological necessity. NH comes equipped with a meta-capacity to acquire G_Ls in virtue of the fact that he is human and it is biologically endemic to humans to have this meta-capacity. If we call this meta-capacity the *Faculty of Language* (FL), then humans necessarily have an FL and necessarily have UG, as the latter is just a description of FL's properties. Thus, what is most *real* about language is that any human can acquire the G_L of any L as easily as it can acquire any other. A fundamental aim of l***i***nguistic theory is to explain *how* this is possible by describing the fine structure of the meta-capacity (i. e. by outlining a detailed description of FL's UG properties).

Before moving on, it is worth observing that despite their different interests LINGers and LANGers can co-exist (and have co-existed) quite happily and they can fruitfully interact on many different projects. The default assumption among LINGers is that currently the *best* way to study G_Ls is to study its products as they are used/queried. Thus, a very useful way of limning the fine structure of a particular G_L is to study the expressions of these G_Ls. In fact, currently, some of the best evidence concerning G_Ls comes from how native speakers use G_Ls to produce, parse and judge linguistic artifacts (e. g. sentences). Thus, LINGers, like LANGers, will be interested in what native speakers say and what they say about what they say. This will be a common focus of interest and cross talk can be productive.

Similarly, seeing how G_Ls vary can also inform one's views about the fine structure of FL/UG. Thus both LINGers and LANGers will be interested in comparing G_Ls to see what, if any, commonalities they enjoy. There may be important differences in how LINGers and LANGers approach the study of these commonalities, but at least in principle, the subject matter can be shared to the benefit of each. And, as a matter of fact, until the Minimalist Program (MP) arose, carefully distinguishing LINGer interests from LANGer interests was not particularly pressing. The psychologically and philologically inclined could happily live side by side pursuing different but (often enough) closely related projects. What LANGers understood to be facts about language(s), LINGers interpreted as facts about G_Ls and/or FL/UG.

MP adversely affects this pleasant commensalism. The strains that MP exerts on this happy LING/LANG co-existence is one reason, I believe, why so many GGers have taken a dislike to MP. Let me explain what I mean by discussing what the MP research question is. For that I will need a bit of a running start.

1.2 The minimalist turn

Prior to MP, LING addressed two questions based on two evident rationally uncontestable facts (and, from what I can tell, these facts have not been contested). The first fact is that a native speaker's capacities cover an unbounded domain of linguistic objects (phrases, sentences etc.). Following Chomsky (1964) we can dub this fact "Linguistic Creativity" (LC).[4] I've already adverted to the second fact: any child can acquire any G_L as easily as any other. Let's dub this fact "Linguistic Promiscuity" (LP). Part of a LINGers account for LC postulates that native speakers have internalized a G_L. G_Ls consist of generative procedures (recursive rules) that allow for the creation of unboundedly complex linguistic expressions (which partly explains how a native speaker effortlessly deals with the novel linguistic objects s/he regularly produces and encounters).

LINGers account for the second fact, LP, in terms of the UG features of FL. This too is a partial account. UG delineates the *limits* of a possible G_L. Among the possible G_Ls,

4 Here's Chomsky's description of this fact in his (1964, 7):

> [...] a mature native speaker can produce a new sentence of his language on the appropriate occasion, and other speakers can understand it immediately, though it is equally new to them. Most of our linguistic experience, both as speakers and hearers, is with new sentences; once we have mastered a language, the class of sentences with which we can operate fluently is so vast that for all practical purposes (and, obviously, for all theoretical purposes), we may regard it as infinite.

the child builds an actual one in response to the linguistic data it encounters and that it takes in (i. e. the Primary Linguistic Data (PLD)).

So two facts, defining two questions and two kinds of theories, one delimiting the range of *possible* linguistic expressions for a given language (viz. G_Ls) and the other delimiting the range of *possible* G_Ls (viz. FL/UG). As should be evident, as a practical matter, in addressing LP it is useful to have to hand candidate generative procedures of specific G_Ls. Let me emphasize this: though it is morally certain *that* humans come equipped with a FL and build G_Ls it is an *empirical* question what properties these G_Ls have and what the fine structure of FL/UG is. In other words, *that* there is an FL/UG and *that* it yields G_Ls is not really open for rational debate. What is open for a lot of discussion *and is a very hard question* is exactly what features these mental objects have. Over the last 60 years GG has made considerable progress in discovering the properties of particular G_Ls and has reasonable outlines of the overall architecture of FL/UG. At least this is what LINGers believe, I among them. And just as the success in outlining (some) of the core features of particular Gs laid the ground for discovering non-trivial features of FL/UG, so the success in limning (some of) the basic characteristics of FL/UG has prepared the ground for yet one more question: why do humans have the FL/UG they do and not some other? This is the MP question. It is a question about *possible* FL/UGs.

There are several things worth noting about this question. First, the target of explanation is FL/UG and the principles it embodies. Thus, MP only makes sense qua program of inquiry if we assume that we know some things about FL/UG. If nothing is known, then the question is premature. In fact, even if something is known, it might be premature. I return to this anon.

Second, the MP question is specifically about the structure of FL/UG. Thus, unlike earlier work where LANG research can be used to obliquely address LC and LP, the MP question only makes sense from a LING perspective (i. e. there is no LANG analogue) because MP specifically asks about possible FL/UGs and, as FL/UGs are cognitive objects, asking this question requires adopting a mentalist stance. Discussing languages and their various properties had better bottom out in some claim about FL/UG's limits if it is to be of MP relevance. This means that the kind of research MP fosters will often have a different focus from that which has come before. This will lead LANGers and LINGers to a parting of the investigative ways. In fact, given that MP takes as more or less *given* what linguists (and many languists) have heretofore investigated to empirically establish, MP theorizing is not really an alternative to pre-MP theory. More specifically, MP can't be an *alternative* to GB because, at least initially, MP is a consumer of GB results.[5] What does this mean?

5 Personally, I am a big fan of GB and what it has wrought. But MP style investigations need not take GB as the starting point for minimalist investigations. *Any conception of FL/UG will do* (e. g. HPSG, RG, LFG etc.). In my opinion, the purported differences among these "frameworks" (something that this

An analogy might help. Think of the relationship between thermodynamics and statistical mechanics. The laws of thermodynamics are grist for the stats mechanics mill, the aim being to derive the thermodynamic generalizations (i. e. as limit theorems) in a more principled atomic theory of mechanics. The right way to think of the relation between MP and earlier grammatical theory is analogous. Take (e. g.) GB principles and see if they can be derived in a more principled way. In my opinion, that's the right way of understanding MP. It is the position that I elaborate in what follows. Note, if this is right, then just as many thermodynamical accounts of, say, gas behavior will be preserved in a reasonable statistical mechanics, so too many GB accounts will be preserved in a decent MP theory of FL. On this view, the relation between GB and MP is not that between a true theory and a false one, but a descriptive theory (what physicists call an "effective" theory) and a more fundamental one.

If this is right, then GB (or whatever FL/UG is presupposed) accounts will mostly be preserved in MP reconstructions. And this is a very good thing! Indeed, this is precisely what we expect in science; results of past investigations are preserved in later ones with earlier work preparing the ground for deeper questions.[6] *Why* are they preserved? Because they are roughly correct and thus *not* deriving these results (at least approximately) is excellent indication that the subsuming proposal is off on the wrong track. Thus, a sign that the more fundamental proposal is worth taking seriously is that it *recapitulates* earlier results. Thus a reasonable initial goal of inquiry is to explicitly *aim* to redo what has been done before (hopefully, in a more principled fashion).

In light of this, it should be evident why many might dismiss MP inquiry. First, it takes as true (i. e. presupposes) what many will think contentious. Second, it doesn't (initially) aim to do much more than derive "what we already know" and so does not appear to add much to our basic knowledge, except, perhaps, a long labored (formally involved) deduction of a long recognized fact.

Speaking personally, my own work takes GB as a *roughly correct* description of FL/UG. Many who work on refining UGish generalizations will consider this tendentious. So be it. *Let it be stipulated that at any time in any inquiry things are more complicated than they are taken to be*. It is also always possible that we (viz. GB) got things entirely wrong. The question is not whether this is an option. Of course it is. The question is how seriously we should take this truism.

The MP project *starts* from the assumption that we have a fairly accurate picture of some of the central features of FL and considers it fruitful to inquire as to why we have

edited collection highlights) have been overstated. To my eye, they say more or less the same things, identify more or less the same limiting conditions and do so in more or less the same ways. In other words, these differing frameworks are for many practical purposes largely interchangeable. Thus, from where I sit, you can engage in minimalist investigation starting with any of these "frameworks". Given that I grew up as a GB syntactician, I will concentrate on results in this idiom.

6 Indeed, preserving earlier results is one of the hallmarks of science, giving rise to one of its most distinctive features; it's cumulative nature.

found *these* features. In other words, MP assumes that time is ripe to ask more fundamental questions because we have reasonable answers to less fundamental questions. If you don't believe this then MP inquiry is not wrong but footling. Confusing wrong with nugatory is to confuse an empirical question with a value issue. In my view, much of the criticism of MP rests on this fact/value confusion. Let me say this another way.

Many who are disappointed in MP don't actually ask (or care) if MP has failed on its own terms, given its own assumptions. Rather criticism generally challenges MP's assumptions. It takes MP to be not so much false as irrelevant or premature by (tacitly) denying that we know enough about FL/UG to even *ask* the MP question. I believe that these objections are misplaced. In what follows, I will assume that GBish descriptions of FL/UG are adequate *enough* (i. e. are right enough) to start asking the MP question. If you don't buy this, MP research will not be to your taste. If you don't buy this, now is a good time to stop reading in order to avoid a stroke. With this public service warning, let's turn to the questions.

2 Tools and particulars[7]

2.1 The Merge Hypothesis: Explaining some core features of FL/UG

Here is a list of *some* characteristic features of FL/UG and its G_Ls:[8]

(1) a. Hierarchical recursion
b. Displacement (aka, movement)
c. Gs generate natural formats for semantic interpretation
d. Reconstruction effects
e. Movement targets c-commanding positions
f. No lowering rules
g. Strict cyclicity
h. G rules are structure dependent
i. Antecedents c-command their anaphors
j. Anaphors never c-command their antecedents (i. e. Principle C effects and Strong Cross Over Effects)

7 At the behest of the editors I title this long section "tools". It is a misleading section heading as we are not going to discuss tools in any recognizable sense of the term. Rather the focus will be on how minimalist analyses go about trying to explain salient properties of FL/UG, which, recall, is what MP is all about.

8 I provide a fuller list of about 30 principles here (http://facultyoflanguage.blogspot.com/2015/03/a-shortish-whig-history-of-gg-part-3.html)

k. XPs move, X's don't, X^0s might
l. Control targets subjects of "defective" (i. e. Tns or agreement deficient) clauses
m. Control respects the Principle of Minimal Distance
n. Case and agreement are X^0-YP dependencies
o. Reflexivization and Pronominalization apply in complementary domains
p. Selection/subcategorization are very local head-head relations
q. Gs treat arguments and adjuncts differently, with the former less grammatically opaque than the latter

Note, I am not saying that this exhausts the properties of FL/UG, nor am I saying that all LINGers agree with *all* of these accurately describe FL/UG.[9] What I am saying is that (1a–q) identify empirically robust(ish) properties of FL/UG and the generative procedures its G_Ls allow. Put another way, I am claiming (i) that certain facts about human G_Ls (e. g. that they have hierarchical recursion and movement and binding under c-command and display principle C effects and obligatory control effects, etc.) are empirically well-grounded and (ii) that it is appropriate to ask *why* FL/UG allows for G_Ls with these properties and not others. If you buy this, then welcome to the Minimalist Program (MP).

I would go further; not only are the assumptions in (i) reasonable and the question in (ii) appropriate, MP has provided answers to the question in (ii) showing how to derive *some* of the properties in (1) from simpler more principled assumptions.[10] One well-known approach to (1a–h), the *Merge Hypothesis* (MH), unifies all these properties, deriving them from the core generative mechanism Merge. More specifically, MH postulates that FL/UG contains a very simple operation (aka, Merge) that suffices to generate unbounded hierarchical structures (1a) and that these Merge generated hierarchical structures also have the seven additional properties (1b–1h). Let's examine the features of this simple operation and see how it manages to derive these eight properties?

Unbounded hierarchy implies a recursive procedure.[11] MH explains this by postulating a simple operation ("Merge") that generates the requisite unbounded hierarchical structures. Merge consists of a simple recursive specification of *Syntactic Objects* (SO) coupled with the assumption that complex SOs are sets.

9 For example, fans of Dependent Case Theory will reject (1n).
10 Note the 'some'. There are still outstanding unresolved topics. So for example, I will say little about labels here, though it is a very important topic. I have discussed this issue elsewhere. Those interested can consult Hornstein (2009) for some discussion.
11 Recall, that LP requires recursion and linguistics has discovered ample evidence that G_Ls can generate structures of arbitrary depth.

(2) a. If α is a lexical item then α is a SO[12]
b. If α is an SO and β is an SO the Merge(α, β) is an SO

(3) For α, β, SOs, Merge(α,β) $\rightarrow \{\alpha, \beta\}$

The inductive step (2b) allows Merge to apply to its own outputs and thus licenses unboundedly "deep" SOs with sets contained within sets contained within sets . . . . The Merge Hypothesis, MH, is that the "simplest" conception of this combinatoric operation (the minimum required to generate unbounded hierarchically organized objects) suffices to explain why FL/UG has many of the other properties listed in (1).

In what way is Merge the "simplest" specification of unbounded hierarchy? The operation has three key features: (i) it directly and uniquely targets hierarchy (i. e. the basic complex objects are sets (which are unordered), not strings), (ii) it in no way changes the atomic objects combined in combining them (Inclusiveness), and (iii) it in no way changes the complex objects combined in combining them (Extension). Inclusiveness and Extension together constitute the "No Tampering Condition" (NTC). Thus, Merge recursively builds hierarchy (and only hierarchy) without "tampering" with the inputs in any way save combining them in a very simple way (i. e. just hierarchy no linear information).[13] The key theoretical observation is that *if* FL/UG has Merge as (one of) its primary generative mechanism(s),[14] then it delivers G_Ls with properties (1a–1h). And if this is right, it provides a proof of concept that it is not premature to ask *why* FL/UG is structured as it is. In other words, this would be a very nice result given the stated aims of MP. Let's see how Merge so conceived derives (1a–h).

It should be clear that Gs with Merge can generate unbounded hierarchical dependencies. Given a lexicon containing a finite list of atoms $\alpha, \beta, \gamma, \delta, \ldots$ we can, using the definitions in (2) and (3) form structures like (4) (try it!).

(4) a. $\{\alpha, \{\beta, \{\gamma, \delta\}\}\}$
b. $\{\{\alpha, \beta\}, \{\gamma, \delta\}\}$
c. $\{\{\{\alpha, \beta\}, \gamma\}, \delta\}$

And given the recursive nature of the operation, we can keep on going *ad libitum*. So Merge suffices to generate an unbounded number of hierarchically organized syntactic objects.

12 The term *lexical item* denotes the atoms that are not themselves products of Merge. These roughly correspond to the notion *morpheme* or *word*, though these notions are themselves terms of art and it is possible that the naïve notions only roughly corresponds to the technical ones. Every theory of syntax postulates the existence of such atoms. Thus, what is debatable is not their existence but their features.

13 In my opinion, this line of argument does not require that Merge be the "simplest" possible operation. It suffices that it be natural and simple. The conception of Merge in (2) and (3) meets this threshold. Moreover, the assumption that Merge forms sets from inputs is simply the technical reflection of the idea that is the combinatoric operation that targets hierarchy and obeys the NTC. Sets have the three features required and so are good ways of representing what Merge does.

14 In the best of all worlds, the sole generative procedure.

Merge can also generate structures that model displacement (i. e. movement dependencies). Movement rules code the fact that a single expression can enjoy multiple relations within a structure (e. g. it can be both a complement of a predicate and the subject of a sentence).[15] Merge allows for the derivation of structures that have this property. And this is a very good thing given that we know (due to over 60 years of work in Generative Grammar) that displacement is a key feature of human G_Ls.

Here's how Merge does this. Given a structure like (5a) consider how (2) and (3) yield the movement structure (5b). Observe that in (5b), β occurs *twice*. This can be understood as coding a movement dependency, β being both sister of the SO α and sister of the derived SO $\{\gamma, \{\lambda, \{\alpha, \beta\}\}\}$. The derivation is in (6).

(5) a. $\{\gamma, \{\lambda, \{\alpha, \beta\}\}\}$
b. $\{\beta, \{\gamma, \{\lambda, \{\alpha, \beta\}\}\}\}$

(6) The SO $\{\gamma, \{\lambda, \{\alpha, \beta\}\}\}$ and the SO β (within $\{\gamma, \{\lambda, \{\alpha, \beta\}\}\}$) merge to from $\{\beta, \{\gamma, \{\lambda, \{\alpha, \beta\}\}\}\}$

Note that this derivation assumes that once an SO always an SO. Thus, Merging an SO α to form part of a complex SO β that contains α does not change (tamper with) α's status as an SO. Because complex SOs are composed of SOs Merge can target a subpart of an SO for further Merging. Thus, NTC allows Merge to generate structures with the properties of movement; structures where a SO is a member of two different "sets".

Let me emphasize an important point: the key feature that allows Merge to generate movement dependencies (viz. the "once an SO always an SO" assumption) follows from the assumption that Merge does *nothing more* than take SOs and form them into a unit. It otherwise leaves the combined objects alone. Thus, if some expression is an SO before being merged with another SO then it will retain this property after being Merged given that Merge in no way changes the expressions but for combining them. NTC (specifically the Inclusiveness and Extension Conditions) leaves all properties of the combining expressions intact. So, if α has some property before being combined with β (e. g. being an SO), it will have this property after it is combined with β. As being an SO is a property of an expression, Merging it will not change this and so Merge can legitimately combine a subpart of an SO to its container.

Before pressing on, a comment: unifying movement and phrase building is an MP innovation. Earlier theories of grammar (and early minimalist theories) treated phrasal dependencies and movement dependencies as the products of entirely *different kinds of rules* (e. g. phrase structure rules vs transformations/Merge vs Copy + Merge). Merge *unifies* these two kinds of dependencies and treats them as different

15 A phrase marker is just a list of relations that the combined atoms enjoy. Derivations that map phrase markers into phrase markers allow an expression to enjoy different relations coded in the various relations it enters into in the varying phrase markers.

outputs of a single operation. As such, the fact that FL yields Gs that contain both unbounded hierarchy *and* displacement operations is unsurprising. Hierarchy and displacement are flip sides of the same combinatoric coin. Thus, if Merge is the core combinatoric operation FL makes available, then MH explains why FL/UG constructs G_Ls that have both (1a) and (1b) as characteristic features.

Let's continue. As should be clear, Merge generated structures like those in (5) and (6) also provides all we need to code the two basic types of semantic dependencies: predicate-argument structures (i. e. thematic dependency) and scope structure. Let me be a bit clearer. The two basic applications of Merge are those that take two separate SOs and combine them and those that take two SOs with one contained in the other and combine them. The former, E-Merge, is fit for the representation of predicate-argument (aka, thematic structure). The latter, I-Merge, provides an adequate grammatical format for representing operator/variable (i. e. scope) dependencies. There is ample evidence that Gs code for these two kinds of semantic information in simple constructions like *Wh*-questions. Thus, it is an argument in its favor that Merge as defined in (2) and (3) provides a syntactic format for both. An argument saturates the predicate it E-merges with and scopes over the SO it I-merges with. If this is correct, then Merge provides structure appropriate to explain (1c).

And also (1d). A standard account of Reconstruction Effects (RE) involves allowing a moved expression to function *as if* it still occupied the position from which it moved. This *as-if* is redeemed theoretically if the movement site contains a copy of the moved expression. Why does a displaced expression semantically comport itself as if it is in its base position? Because a copy of the moved expression *is* in the base position. Or, to put this another way, a copy theory of movement would go a long way towards providing the technical wherewithal to account for the possibility of REs. But Merge based accounts of movement like the one above embody a copy theory of movement. Look at (5b): β is in two positions in virtue of being I-merged with its container. Thus, β is a member of the lowest set and the highest. Reconstruction amounts to choosing which "copy" to interpret semantically and phonetically.[16] Reducing movement to I-merge explains why movement permits REs.[17]

Furthermore, MH *entails* the copy theory of movement. How so? Recall that MH eschews tampering. Thus, if movement is a species of Merge then NTC requires coding movement with "copies". To see this, contrast how movement is treated in GB.

Within GB, if α moves from its base position to some higher position a trace is left in the launch site. Thus, a GB version of (5b) would look something like (7):

(7) $\{\beta_1, \{\gamma, \{\lambda, \{\alpha, t_1\}\}\}\}$

16 Copy is simply a descriptive term here. A more technically accurate variant is "occurrence". β occurs twice in (5b). The logic, however, does not change.

17 A full theory of REs would articulate the principles behind choosing which copies to interpret. See Sportiche (forthcoming) for an interesting substantive proposal.

Two features are noteworthy; (i) in place of a copy in the launch site we find a trace and (ii) that trace is co-indexed with the moved expression β.[18] These features are built into the GB understanding of a movement rule. Understood from a Merge perspective, this GB conception is doubly suspect for it violates the Inclusiveness Condition clause of the NTC twice over. It replaces a copy with a trace and it adds indices to the derived structure. Empirically, it also mystifies REs. Traces have no contents. That's what makes them traces (see note 18). Why are they then able to act as if they did have them? To accommodate such effects GB adds a layer of theory *specific* to REs (e. g. it invokes reconstruction rules to undo the effects of movement understood in trace theoretic terms). Having copies in place of traces simplifies matters and explains how REs are possible. Furthermore, if movement is a species of Merge (i. e. I-merge) then SOs like (7) are not generable at all as they violate NTC. More specifically, *the only kosher way to code movement and obey the NTC is via copies*. So, the only way to code movement given a simple conception of syntactic combination like Merge (i. e. one that embodies NTC) results in a copy theory of movement that serves to rationalize REs without a theoretically bespoke theory of reconstruction. Not bad![19]

So Merge delivers properties (1a–1d), and the gifts just keep on coming. It also serves up (1e,f,g) as consequences. This time let's look at the Extension Condition (EC) codicil to the NTC. EC requires that inputs to Merge be preserved in the outputs to Merge (any other result would "change" one of the inputs). Thus, if an SO is input to the operation it will be a unit/set in the output as well because Merge does no more than create linguistic units from the inputs. Thus, whatever is a constituent in the input appears as a constituent with the same properties in the output. This implies (i) that all I-merge is to a c-commanding position, (ii) that lowering rules cannot exist, and (iii) that derivations are strictly cyclic.[20] The conditions that movements be always upwards to c-commanding positions and strictly cyclic thus follows trivially from this simple specification of Merge (i. e. Merge with NTC understood as embodying the EC).

An illustration will help clarify this. NTC prohibits deriving structure (8b) from (8a). Here we Merge γ with α. The output of this instance of Merge obliterates the fact that $\{\alpha, \beta\}$ had been a unit/constituent in (8a), the input to Merge. EC prohibits this.

18 Traces within GB are indexed contentless categories: $[_1\ \text{ec}]$.

19 We could go further: Merge based theory cannot have objects like traces. Traces live on the distinction between Phrase Structure Rules and lexical insertion operations. They are effectively the phrase structure scaffolding without the lexical insertion. But, Merge makes no distinction between structure building and lexical insertion (i. e. between slots and contents). As such, *if* traces exist, they must be lexical primitives rather than syntactically derived formatives. This would be a very weird conception of traces, inconsistent with the GB rendering in note 18. The same, incidentally, goes for PRO, which we will talk about later on. The upshot: not only would traces violate NTC, they are indefinable given the "bare phrase structure" nature of movement understood as I-merge.

20 The first conjunct only holds if there is no inter-arboreal/sideward movement. I believe that this is incorrect (and have argued as much) but in this paper I will assume it to be correct.

It effectively restricts I-Merge to the root. So restricted, (8b) is not a licit instance of I-Merge (note that $\{\alpha, \beta\}$ is not a unit in the output). Nor is (8c) (note that $\{\{\alpha, \beta\}, \{\gamma, \delta\}\}$ is not a unit in the output). Nor is a derivation that violates the strict cycle (as in (8d)). Only (8e) is a grammatically licit Merge derivation for here all the inputs to the derivation (i. e. γ and $\{\{\alpha, \beta\}, \{\gamma, \delta\}\}$) are also units in the output of the derivation (i. e. thus the inputs have been preserved (remain unchanged) in the output). Yes a new relation has been added, but no previous ones have been destroyed (i. e. the derivation is info-preserving (viz. monotonic)). Repeat the mantra: once an SO always an SO. In deriving (8b–c) one of the inputs (viz. $\{\{\alpha, \beta\}, \{\gamma, \delta\}\}$) is no longer a unit in the output and so NTC/EC has been violated.

(8) a. $\{\{\alpha, \beta\}, \{\gamma, \delta\}\}$
b. $\{\{\{\gamma, \alpha\}, \beta\}, \{\gamma, \delta\}\}$
c. $\{\{\gamma, \{\alpha, \beta\}\}, \{\gamma, \delta\}\}$
d. $\{\{\alpha, \beta\}, \{\delta, \{\gamma, \delta\}\}$
e. $\{\gamma, \{\{\alpha, \beta\}, \{\gamma, \delta\}\}\}$

In sum, if movement is I-merge subject to NTC then all movement will necessarily be to c-commanding positions, upwards, and strictly cyclic.

It is worth noting that the features here derived via the Merge Hypothesis are not particularly recondite properties of FL/UG. Indeed, they find a place in most GG accounts (HPSG, LFG, RG etc.) of movement. This makes their seamless derivation within a Merge based account particularly interesting.

Last, we can derive the fact that the rules of grammar are structure dependent ((1h) above), an oft-noted feature of syntactic operations.[21] Why should this be so? Well, if Merge is the sole syntactic operation and then non-structure dependent operations are very hard (impossible?) to state. Why? Because the products of Merge are sets and sets impose no linear requirements on their elements.[22] If we understand a derivation to be a mapping of phrase markers into phrase markers *and* we understand phrase markers to effectively be sets (i. e. to only specify constituency and hierarchical relations) then it is no surprise that rules that leverage linear left-right properties of a string cannot be exploited. They don't exist for phrase markers eschew this sort of information and thus operations that exploit left/right (i. e. string based) information cannot be defined. So,

21 For a recent review and defense of the claim see Berwick et al. (2011).

22 Actually, MH does not require that PMs *be* sets. It only requires that Merge do *nothing more* than combine SOs. Sets are adequate vehicles for representing Merge's slightness as simple sets say nothing more of their contents than that they form a unit. In particular, they impose no order on their elements and elements are in no way changed in virtue of becoming members of a set (save being members). So, if Merge is simple in the sense noted above, sets will serve as useful proxies for the objects that they construct. As Stabler (2010) notes however, other formats with the same properties are available and would yield analogous results.

why are rules of G structure dependent? Because this is the only structural information that Merge based Gs represent. So, if the basic combinatoric operation that FL/UG allows is Merge, then FL/UGs restriction to structure dependent operations quickly follows.

This is a good place to pause for a temporary summary: Research in GG over the last 60 years has uncovered several plausible design features of FL/UG. (1a–h) summarizes some uncontroversial examples. *All* of these properties of FL/UG can be unified if we assume that Merge as outlined in (2) and (3) is the basic combination operation that FL/UG affords. Put simply, the Merge Hypothesis has (1a–h) as consequences.

Let me say the same thing more tendentiously. All agree that a basic feature of FL/UG is that it allows for Gs with unbounded hierarchy. A very simple inductive procedure sufficient for specifying this property (1a), also entails many other features of FL/UG (1b–h). What makes this specification simple is that it directly targets hierarchy and requires that the computation be strongly monotonic (embody the NTC). Thus we can explain the fact that FL/UG has these properties by assuming that it embodies a very simple (arguably, the simplest) version of a procedure that *any* empirically adequate theory of FL/UG would have to embody. Or, *given* that FL/UG allows for unbounded hierarchical recursion (a non-controversial fact given the fact of Linguistic Productivity), the *simplest* (or at least, very simple) version of the requisite procedure brings in its train displacement, an adequate format for semantic interpretation, Reconstruction Effects, movement rules that target c-commanding positions, eschew lowering and are strictly cyclic, and G operations that are structure dependent. Thus, if the Merge Hypothesis is true (i. e. if FL/UG has Merge as the basic syntactic operation), it *explains* why FL/UG has this bushel of properties. In other words, the Merge Hypothesis provides a plausible first step in answering the basic MP question: why does FL/UG have the properties it has?

Moreover, it is morally certain that *something* like Merge will be part of any theory of FL/UG precisely because it is so very simple. It is always possible to add bells and whistles to the G rules FL/UG makes available. But any theory hoping to be empirically adequate will contain *at least* this much structure. After all, what do (2) and (3) specify? They specify a recursive procedure for building hierarchical structures that does nothing but build such structures. Given the fact of Linguistic Productivity and Linguistic Promiscuity any theory of FL/UG will contain *at least* this much. If it does not contain much more than this much, then (1a–h) results. Not bad.

2.2 The Extended Merge Hypothesis: Explaining more features of FL/UG

There are proposals in the MP literature that push this MP line of argument harder still. Here's what I mean. The Merge Hypothesis unifies structure building and movement and in the process explains many central properties of FL/UG. MH accomplishes this by reducing phrase building and movement to instances of Merge (i. e. E and I-Merge respectively). We can push this reductive/unificational approach more aggressively by reducing other kinds of dependencies to instances of E or I-Merge. More specifically, if we take seriously the MP proposal that Merge is the *unique fundamental* combinatoric operation that FL/UG affords, then the strongest minimalist hypothesis is that *every* grammatical dependency must be mediated by some instance of Merge. Call this the "Extended Merge Hypothesis" (EMH).[23] In what follows, I would like to review some of the MP literature to see what properties EMH might enjoy. My aim is to suggest that this further unification has properties that nicely track some fundamental features of FL/UG. If this is correct, then it suggests that relentlessly expanding the reach of Merge beyond phrase structure and movement to include construal and case/agreement dependencies has interesting potential payoffs for those interested in MP questions. Once again, it will pay to begin with GB as a jumping off point.[24]

GB is a strongly modular theory in the sense that it describes FL/UG as containing many different kinds of operations and principles. Thus, GB distinguishes construal rules like Binding and Control, from movement rules like Wh-movement and Raising. We classify case relations as different from movement dependencies and both from theta forming dependencies. The primitives are different and, more importantly, the operations and principles that condition them are different. The internal modularity of GB style FL/UGs complicates them. This is, theoretically speaking, unfortunate, especially in light of the fact that the different dependencies that the modules specify share many properties in common. That they do so is something an MP account (indeed, any account) would like to explain. EMH proposes that all the different dependencies in GB's various modules structure are actually only *apparently* distinct. In reality, they are all different instances of chains formed by I-merge.[25] To put this slightly differently, all of the non-local dependencies GB specifies "live on" chains formed by I-Merge.[26] Let's consider some examples.

23 In the conclusion below I dub the assumption embodied in the Extended Merge Hypothesis the Fundamental Principle of Grammar: the only way to establish a grammatical dependency is by merging the dependents.

24 But first a warning: many MPers would agree with the gist of what I outlined in Section 2.1. What follows is considerably more (indeed, much more) controversial. I will not have the space (or, to be honest, the interest) in defending the line of argument that follows. I have written about this elsewhere and tried to argue that, for example, large parts of the rules of construal can be usefully reduced to I-merge. Many have disagreed. For my point here, this may not be that important as my aim here is to see how far this line of argument can go. Showing that it is also the best way to go is less important in the current context than showing that it is a plausible way to proceed.

25 This MP project clearly gains inspiration from the unification of islands under Subjacency, still, in my opinion, one of the great leaps forward in syntactic understanding.

26 GB assumes that chains come in two basic flavors: A vs A'. This makes providing a principled explanation for the A vs A' distinction an important theoretical MP project. Despite its importance, I will not discuss it here because of space limitations.

2.2.1 Case theory

Chomsky (1993) re-analyzes case dependencies as movement mediated. The argument is in two steps.

The first is a critical observation: the GB theory of case is contrived in that it relies on a very convoluted and unnatural notion of government. Furthermore, the contrived nature of the government account reflects a core assumption: accusative case on the internal argument of a transitive verb (sisterhood to a case assigning head) reflects the core configuration for case licensing. Extending sisterhood so that it covers what we see in nominative case and in ECM requires "generalizing" sisterhood to government, with the resulting government configuration itself covering three very distinct looking configurations (see (9)). (9a) is the configuration for accusative case, (9b) for nominative and (9c) for ECM. It is possible to define a technical notion that treats these three configurations as instances of a common relation (viz. government), but the resulting definition is very baroque.[27] The obtuseness of the resulting definition argues against treating (9a) as the core case precisely because the resulting unified theory rests on a gerrymandered (and hence theoretically unsatisfying) conception of government.

(9) a. [V nominal]
b. [Nominal [T^0-finite …
c. [V [Nominal [T^0-non-finite …

The second step in the argument is positive: case theory can be considerably streamlined and unified if we take the core instance of case assignment to be exemplified by what we find with nominatives. If this is the core case configuration then case is effectively a spec-head relation between a case marked nominal and a head that licenses it. Generalizing nominative case configurations to cover simple accusative objects and ECM subjects requires treating case as a product of movement (as with nominatives).[28] Thus, simplifying the case module rests on analyzing case dependencies as products of I-merge (i. e. as non-local relations between a case assigning head and a nominal that has (A-)moved to the specifier of this head).[29] The canonical case configuration is (10), with h^0 being a case assigning head and the nominal being the head of a I-merge generated A-chain.[30]

27 Defining government so that it could do all required of it in GB was a lively activity in the 80s and 90s.

28 At least if we adopt the Predicate Internal Subject Hypothesis which assumes that subjects of finite clauses move to Spec T from some lower predicate internal base position in which the nominals theta role is determined. For discussion see Hornstein et al. (2005).

29 This abstracts away from the issue of assignment versus checking, a distinction I will ignore in what follows.

30 If we assume that structures are labeled and that labels are heads then (10) has the structure in (10') and we can say that the nominal merges with h^0 in virtue of merging with a labeled projection

(10) [Nominal [h^0 …

There is some interesting empirical evidence for this view. First, it predicts that we should find a correlation between case and movability. More specifically, if some position resists movement, then this should have consequences for case. And this seems to be correct. Consider the paradigm in (11):

(11) a. John believes him to be tall
b. * John believes him is tall
c. John was believed t to be tall
d. * John was believed t is tall

Just as A-movement/raising from the subject position of a finite clause is prohibited, so too is accusative case. Why? Because accusative case requires A-movement of *him* in (11b) to the case head that sits above *believe* and this kind of movement is prohibited, as (11d) illustrates.[31]

There is a second prediction. Case should condition binding. On the current proposal, movement feeds case. As movement broadens an expression's scope, it thereby increases its binding potential. This second prediction is particularly interesting for it ties together two features of the grammar that GB approaches to accusative case keep separate. Here's what I mean.

With regard to nominative case assignment, it is well-known that movement "for" case (e. g. raising to subject) can expand an expression's binding potential. *John* can

---

of h. I personally believe that this is the right view. However, as I ignore labeling in this essay, this is not the place to go into these matters.

(10') [$_h$ Nominal [$_h$ h^0 …

31 That case and movement should correlate is implicit in GB accounts as well. Movement in raising and passive constructions is "for" case. If movement is impossible, the case filter will be violated. However, the logic of the GB account based on government is that movement "for" case was special. The core case licensing configuration did not require it. Chomsky's (1993) insight is that if one takes the movement fed licensing examples as indicative of the underlying configuration a more unified theory of case licensing was possible. Later MP approaches to case returned to the earlier GB conception, but, in my view, at a significant cost. Later theory added to Merge an additional G operation, AGREE. AGREE is a long distance operation between a probe and a c-commanded goal. It is possible to unify case licensing configurations using AGREE (and hence rejecting the EMH). However, one looses the correlation between movement and scope (discussed immediately below) unless further assumptions are pressed into service.
Why the shift from the earlier account? I am not sure. So far as I can tell, the first reason was Chomsky's unhappiness with Spec-X^0 relations (Chomsky took these to be suspect in a way that head-complement relations are not (I have no idea why)) and became more suspicious in a label free syntax. If labels are not syntactically active, then there isn't a local relation between a moved nominal and a case licensing head in a Spec-head configuration. So, if you don't like labels, you won't like unifying case under Spec-head. Or, to put this more positively (I am after all a pretty positive fellow), if you are ok with labels (I love them) then you will find obvious attractions in the Spec-head theory.

bind *himself* after raising to subject in (12b) but not without moving to the matrix Spec T (12a). Thus some instances of movement for case reasons can feed binding.

(12) a. * It seems to himself$_1$ [(that) John$_1$ is happy]
b. John$_1$ seems to himself$_1$ [t$_1$ to be happy]

However, the standard GB analysis of case for accusatives has the V assign case to the nominal object in its base position.[32] Thus, whereas case to nominative subjects is a Spec-head relation, case to canonical objects is under sisterhood. *If*, however, we unify nominative and accusative case and assimilate the latter to what we find with nominatives, then movement will mediate accusative case too. If this involves movement to some position above the external argument's base position (recall, we are assuming predicate internal subjects) then accusative case is being assigned in a configuration something like (13). Different epochs of MP have treated this VP external Spec position differently, but the technical details don't really matter. What does matter is that accusative case is *not* assigned in the object nominal's base position, but rather in a higher position at the edge of the VP complex. So conceived, accusative case, like nominative, is expected to expand a nominal's binding domain.

(13) [Nominal$_1$ [$_V$ external argument [$_V$ V . . .

There is evidence supporting this correlation between case value and scope domain.[33] Here is some comparative data that illustrates the point. (14) shows that it is possible for an embedded ECM subject to bind an anaphor in a matrix adjunct, whereas a nominative embedded subject cannot.

(14) a. The lawyer proved [the men$_1$ to be guilty] during each other$_1$'s trials
b. * The lawyer proved [the men$_1$ were guilty] during each other$_1$'s trials

(14a) has a sensible reading in which the *during* phrase modifies the matrix predicate *proved*. This reading is unavailable in (14b), the only reading being the silly one in which the *during* phrase modifies *were guilty*. This is expected if licensing accusative case on the ECM subject requires moving it to the edge of the higher matrix VP (as in (13)). In contrast, licensing nominative leaves *the men* in the embedded spec T and hence leaves the matrix *during* phrase outside its c-command domain prohibiting binding of the reciprocal. Thus we see that case value and scope domain co-vary as the MP story leads us to expect.[34]

32 As is also the case for AGREE based conceptions, see previous note.
33 This reprises the analysis in Lasnik and Saito (1991), which is in turn based on data from Postal. For a more elaborate discussion with further binding data see Hornstein et al. (2005, 133 ff.).
34 Analogous data for the internal argument obtain as well:
c. John criticized the men during each other's trials
I leave unpacking the derivations as an exercise.

In sum, unifying case under I-merge rather than government leads to a nicer looking theory and makes *novel* (apparently accurate) predications concerning the interaction of case and binding.

2.2.2 Obligatory complement control

Consider next obligatory complement control, as exemplified in (15):

(15) a. John_1 hopes [PRO_1 to go to grad school]
b. John persuaded Mary_1 [PRO_1 to go to grad school]

Here are two salient properties of these constructions: (i) PRO is restricted to non-finite subject positions and (ii) PRO requires a local c-commanding antecedent. There are GB proposals to account for the first property in terms of binding theory (the so-called "PRO theorem") but by the early 1990s, its theoretical inadequacies had become apparent and PRO's distributional restrictions were hereafter restricted to the subject of non-finite clauses by stipulation.[35] As regards selecting the appropriate antecedent, this has the remained the province of a bespoke control module with antecedent selection traced to stipulated properties of the embedding predicate (i. e. the controller is a lexical property of *hope* and *persuade*). I believe that it is fair to say that both parts of GB control theory contain a fair bit of ad hocery.[36]

Here's where MP comes to the rescue. A unified more principled account is available by treating construal relations as "living" on chains (in the case of control, A-chains) generated by I-merge.[37] On this view, the actual structure of the sentences in (15) is provided in (16) with the controller being the head of an A-chain with links/copies in multiple theta positions (annotated below).

(16) a. [John [T [John_Θ [hopes [John [to [John_Θ [go to grad school]]]]]]]]
b. [John T [John [persuade [Mary_Θ [Mary to [Mary_Θ [go to grad school]]]]]]]

The unification provides a straightforward account for both facts above: where PRO is found and what its antecedent must be. PROs distribute like links in A-chains. Antecedents for PRO are heads of the chains that contain them. Thus, PRO can appear in positions from which A-movement is licit. Antecedents will be the heads of such licit

35 Usually via a dedicated diacritic feature (e. g. null case) but sometimes even less elegantly (if that is conceivable).

36 In fact, the control relation, though understood to be a species of selection, does not conform to the locality restrictions characteristic of selection and so is a theoretically idiosyncratic and *ad hoc*. I leave showing this as an exercise.

37 The idea that control is a chain relation is not original to MP. It was mooted within GB in Koster (1984) and Manzini (1983). The idea that control is a movement dependency was first proposed by Bowers (1973).

chains. Observe that this implies that PRO has all the properties of a GB A-trace. Thus it will be part of a chain with proximate links, these links will c-command one another and will be local in the way that links in A-chains are local. In other words, a movement theory of control derives the features of control constructions noted above.

We can go further: if we assume that Merge is the *only* way to establish grammatical dependencies, then control configurations *must* have such properties. If PRO is a "trace" then of course it requires an antecedent. If it is a trace, then of course the antecedent must c-command it. If it is an A-trace, then of course the antecedent must be local. And if it is an A-trace then we can reduce the fact that it (typically)[38] appears in the subject position of non-finite clauses to the fact that A-movement is also so restricted:

(17) a. John seems t to like Mary
b. * John seems t will like Mary
c. John expects PRO to like Mary
d. * John expects PRO will like Mary

In sum, if we reduce control dependencies to A-chain dependencies and treat control structures as generated via I-merge it is possible to derive some of its core distributional and interpretive properties.[39] Indeed, I would go further, much further.

First, at this moment, *only* this approach to control offers a possible *explanation* for properties of control constructions. All other approaches *code* the relevant data, they do not, *and cannot* explain them. And there is a principled reason for this. All other theories on the market treat PRO as a primitive lexical element, rather than the residue of grammatical operations, and hand pack the properties of control constructions into the feature specifications of this primitive lexical element. The analysis amounts to showing that checking these features correlates with tracking the relevant properties. The source of the features, however, is grammatically exogenous and arbitrary. The features posited are exactly those that the facts require, thereby allowing for other features were the facts different. And this robs these accounts of any explanatory potential. From a minimalist perspective, one from which the question of interest is *why* FL/UG has the properties it appears to have and not others, this treatment of control is nugatory.

Second, the movement approach to control has a very interesting empirical consequence in the context of standard MP theories. Recall that the copy theory is a consequence of Merge based accounts of movement (see Section 2.1). If control is the product of I-merge then control chains, like other A-chains, have copies as links. If so, part of

38 I say "typically" for A-movement is not always so restricted and it appears that in these Gs neither is control. See Boeckx et al. (2010, chapter 4) for discussion.

39 Again, space prohibits developing the argument in full detail. The interested reader should consult the Boeckx et al. (2010) presentation.

any G will be procedures for phonetically "deleting" all but one of the copies/occurrences. So the reason that PRO is phonetically null is that copies in A-chains are generally phonetically null.[40] Importantly, whatever the process that "deletes" copies/occurrences will apply uniformly to "A-traces" and to PRO as these are the same kinds of things.

There is well-known evidence that this is correct. Consider contraction effects like those in (18). *Wanna* contraction is licensed in (18b) across an A-trace and in (18a) across a PRO, but not in (18c) across a A'-trace. This supports the claim that PRO is the residue of A-movement (rather than being a base generated pronominal primitive or the residue of A'-movement).

(18) a. They want PRO to kiss Mary → They wanna kiss Mary
b. They used t to live in the attic → They usta live in the attic
c. Who do they want t to vanish from the party → *Who do they wanna vanish from the party.

The I-Merge analysis of control also predicts a possibility that PRO based accounts cannot tolerate. Consider, an I-Merge based account of displacement needs a theory of copy/occurrence pronunciation to account for the fact that most copies/occurrences in many languages are phonetically null. So part of any I-Merge theory of displacement we need a theory of copy deletion. A particularly simple one allows higher copies as well as lower ones to delete, at least in principle.[41] This opens up the following possibility: there are control configurations in which "PRO" c-commands its antecedent.[42] Thus, the movement theory of control in conjunction with standard assumptions concerning deletion in the copy theory of movement allow for the possibility of control constructions which apparently violate Principle C. And these appear to exist. It is possible to find languages in which the controllee c-commands its controller.[43] In other words, configurations like (19b) with the standard control interpretation are perfectly fine and have the interpretation of control sentences like (19a). Both kinds of sentences are derivable given assumptions about I-merge and copy deletion. They derive from the common underlying (19c) with either the bottom copy removed (19a) or the top (19b). On this view, the classical control configuration is simply a limiting case of a

40 My own view is that this is probably a reflex of case theory. See Haddad and Potsdam (2013) for a proposal along these lines.

41 We will soon see that in some languages many copies can be retained, but let's put this aside for the moment.

42 As Haddad and Potsdam (2013) note there actually four possibilities: The higher copy is retained, the lower, either the higher or lower or both. Haddad and Potsdam (2013) provides evidence that all four possibilities are in fact realized, a fact that provides further support for treating Control as living in I-Merged generated A-chains plus some deletion process for copies.

43 For discussion, see Boeckx et al. (2010) and the review in Haddad and Potsdam (2013).

more general set of possibilities, that but for phonetic expression, have the same underlying properties.[44]

(19) a. DP_1 V [PRO_1 VP]
b. PRO_1 V [DP_1 VP]
c. DP_1 V [DP_2 VP]

This kind of data argues *against* classical PRO based accounts (decisively so, in my opinion), while being straightforwardly compatible with movement approaches to control based on I-merge.

One last point: Given standard MP assumptions, something like the movement theory of control is inevitable once PRO is discarded. MP theories have dispensed with D-structure as a level of representation (see Chomsky, 1993), and with this a principled prohibition against moving a DP into multiple theta positions. Thus, there is nothing to prevent DPs from forming control chains via I-merge given the barest MP assumptions. In this sense, control as movement is an MP inevitability. It is possible to block this implication, but only by invoking additional ad hoc assumptions. Not only is control as movement *compatible* with MP, it is what we will find unless we try to specifically avoid it. That we find it, argues for the reduction of control to I-merge.[45]

2.2.3 Principle A effects

The same logic reduces principle A-effects to I-merge. It's been a staple of grammatical theory since LGB that A-traces have many of the signature properties of reflexives, as illustrated by the following paradigm:

(20) a. * John seems [t is intelligent]
b. * John believes [himself is intelligent]
c. John seems [to be intelligent]
d. John believes [himself to be intelligent]
e. * John seems it was told t that Sue is intelligent
f. * John wants Mary to tell himself that Sue is intelligent

LGB accounts for this common pattern by categorizing A-traces as anaphors subject to principle A. Thus, for example, in LGB-land the reason that A-movement is always upwards, local and to a c-commanding position is that otherwise the traces left by

44 Observe, for example, that control is still a chain relation linking two theta positions, the embedded one being the subject of a non-finite clause.
45 There are many other properties of control constructions that an I-Merge account explains (e. g. the Principle of Minimal Distance). For the curious, this is reviewed in Boeckx et al. (2010).

movement are unbound and violate principle A. What's important for current purposes is to observe that LGB unifies A-anaphoric binding and movement. The current proposal that *all* grammatical dependencies are mediated by Merge has the LGB unification as a special case if we assume that A-anaphors are simply the surface reflex of an underlying A-chain. In other words, the data in (20) follow directly if reflexives "live on" A-chains. Given standard assumptions concerning I-merge this could be theoretically accommodated if "copies" can convert to reflexive in certain configurations (as in (21)).[46]

(21) [John believes [John (→ himself) to be intelligent]]

Like cases of control, reflexives are simply the morphological residue of I-merge generated occurrences/copies. Put another way, reflexives are the inessential morphological detritus of an underlying process of reflexivization, the latter simply being the formation of an A-chain involving multiple theta links under I-Merge.

If correct, this makes a prediction: reflexives *per se* are inessential for reflexivization. There is evidence in favor of this assumption. There are languages in which copies can stand in place of reflexive morphemes in reflexive constructions. Thus, structures like (22a) have reflexive interpretations, as witnessed by the fact that they license sloppy identity under ellipsis (22b).[47]

(22) a. John saw John (= John saw himself)
b. John saw John and Mary too (= and Mary say Mary)

Note that given standard assumptions regarding GB binding theory examples like (22) violate principle C.

We also have apparent violations of principle B where pronouns locally c-command and antecede another pronoun (structure in (23)):

(23) Pronoun likes pronoun and Mike too (= and Mike likes Mike)

These puzzles disappear when these are seen as the surface manifestations of reflexivization chains under I-merge. The names and pronouns in object positions in (22) and (23) are just pronounced occurrence/copies. There is a strict identity condition on the copies in copy reflexive constructions, again something that an I-merge view of these constructions would lead one to expect. Interestingly, we find similar copies possible in "control" structures:

(24) a. Mike wants Mike to eat
b. The priest persuaded Mike Mike to go to school

46 This partly resurrects the old Lees-Klima theory of reflexivization, but without many of the problems. For discussion see Lidz and Idsardi (1998), Hornstein (2001) and Zwart (2002).
47 See Boeckx et al. (2010) and references therein for discussion.

This is to be expected if indeed both Reflexive and Control constructions are mediated by I-merge as proposed here.[48]

2.3 Summary

Let me sum up. Section 2.1 showed that we can gain explanatory leverage on several interesting features of FL/UG if we assume that Merge is the fundamental operation for combining lexical atoms into larger hierarchical structures. In this section I argued that one can get leverage on other fundamental properties if we assume that *all* grammatical dependencies are mediated by Merge. This implies that non-local dependencies are products of I-merge. This section presented evidence that case dependencies, control and reflexivization "live on" A-chains formed by I-merge. I have shown that this proposal accounts for much of the conventional data in a straightforward way *and* that it is compatible with data that goes against the conventional grain (e. g. backwards control, apparent violations of principle B and C binding effects). Moreover, all of this follows from two very simple assumptions: (i) that Merge is a basic combinatoric operation FL/UG makes available and (ii) that all grammatical dependencies are mediate via Merge (i. e. Merge is the unique way of establishing grammatical dependencies). We have seen that the second assumption underwrites I-merge analyses of case, control and reflexivization, which in turn explain some of the key features of case, control and reflexivization (e. g. case impacts scope, PRO occupies the subject position of non-finite clauses and requires a local c-commanding antecedent and that languages that appear to violate conditions B and C are not really doing so). Thus, the Merge hypothesis so extended resolves some apparent paradoxes, accounts for some novel data, covers the standard well-trodden empirical ground *and* (and this is the key part) *explains* why it is that the FL/UG properties GB identified hold in these cases. The Extended Merge Hypothesis explains why these constructions have the general properties they do by reducing them to reflexes of the A-chains they live on generated by I-merge. If this is on the right track, then the EMH goes some way towards answering the question that MP has set for itself.

3 Conclusion

The above was intended to be a full-throated advertisement for the success of MP. The original paper for this volume included further extensions and also identified some problems ripe for minimalist research. Sadly, space limitations forced their removal

48 This proposal also predicts that backwards reflexive constructions should be possible, and indeed, Polinsky and Potsdam (2002) argues that these exist in Tsez.

(though I am working on a book where I will lay this all out).[49] Suffice it to say, much remains to be done.

That conceded, contrary to a widespread perception, it is apparent to me that MP has been wildly successful. It has posed an interesting question (why does FL/UG have the properties it has) and has advanced interesting (albeit partial) answers (Merge generates certain kinds of dependencies in terms of which we can explain many of the fundamental properties of FL/UG). The Extended Merge Hypothesis offers a way of explaining many of the features of FL/UG noted in (1). Thus, the above has offered accounts of (1a–j and l–o). Though there here are still many properties left unexplained, that's just the nature of inquiry. So far as I can discern there are no fundamental reasons for thinking that the MP enterprise should not continue to advance our understanding of FL/UG along the lines indicated.

Moreover, MP has led to novel empirical insights: linking case and scope, allowing for backwards control and copy reflexivization/control, suggesting the existence of a novel kinds of movement (e. g. movement into theta positions) and linking hyperraising and control to the same parameter. So, the EMH has unified previously disparate theoretical domains and has generated novel predictions and proposed novel kinds of dependencies. On its own terms then, MP has been quite successful.[50]

So why is there a widespread belief that MP has been a step backward? Let me end by returning to what I said at the outset.

MP only makes sense in the context of a biological/cognitive inquiry into the structure of human linguistic capacity. Because FL/UG is the cynosure of MP inquiry, it really makes no sense from a philological perspective. Thus the minimalist program disrupts the heretofore peaceful coexistence between l***i***nguistics and l***a***nguistics. In addition, it valorizes purely theoretical work more greatly than has been the norm within GG. Let me pound this last point into the ground.

For MP to succeed requires presenting a simple vision of FL/UG. If GB is taken to be a decent first approximation of what FL/UG looks like, MP requires a radical unification of the modules. And figuring out how to do this is first and foremost a theoretical challenge. Thus, MP prizes the kind of theoretical work that languists see as at best of secondary importance. So, not only does MP serve to clarify the aims

49 For example, discussion of binding principles B and C has been excised. See Kayne (2002) and Hornstein (2007) for some tentative discussion.

50 Moreover, as noted, this tentative success suggests a strong version of the Merge Hypothesis, call it the Fundamental Principle of Grammar (FPG).

FPG: For α and β to be grammatically dependent, α and β must merge

FPG requires all grammatical dependency formation to take place between sistered (labeled) expressions (e. g. selection, antecedence, binding, control, theta marking etc.). In effect, the same operation that explains the fundamental property of Gs (their unbounded hierarchical complexity) would also explain everything else. This conclusion is almost certainly too good to be true. But, it is not too good to explore. The Merge Hypothesis and the Extended Merge Hypothesis are steps in establishing the FPG.

of generative inquiry thereby conceptually separating it from the kind of language centered descriptive philological inquiry practiced by most of the field, it favors a form of investigation that is far more abstract and removed from language specific details than has been the case before. MP, in other words, sharpens divisions that have been latent in the discipline for 60 years and prizes the kind of work that is less knee deep in descriptive detail than is most work in GG. And (surprise, surprise) some don't like this. And this malignity has prevented many from evaluating MP on its own terms. That is too bad, for when so evaluated, the achievements have been reasonably impressive. There are reasonable proposals (e. g. the MH, EMH) that if correct would explain why FL/UG has many of the properties GB has claimed for it. These proposals make some novel empirical predictions and resolve apparent paradoxes. Some require re-thinking of earlier assumptions. Some do not. That's what a fecund research program does. On its own terms, MP has been very productive. Unfortunately, many have missed this. The fault does not lie with MP.

4 Sample analysis

The editors asked all contributors to provide an analysis of (25):

(25) After Mary introduced herself to the audience, she(1) turned to the man that she(2) met before

Before offering a partial analysis, consider the standard GB analysis of (25). The *after* phrase is an adjunct adjoined to TP (or maybe higher). *Herself* is a reflexive with *Mary* as antecedent. *Mary* might also be antecedent for *she* sitting in Spec T and this subject might be antecedent for the embedded *she*, though neither pronoun need have a sentence internal antecedent at all. *Mary* and *she* receive nominative case from the respective finite T^0s whose specs they occupy. *Herself* receives accusative case from the transitive verb *introduced*. Moreover, *Mary* receives the external theta role and *herself* the internal theta role from *introduced*, while *she*(1) gets the external role of *turn*, *she*(2) the external role of *meet* and the head of the relative clause *the man* gets the internal role of *meet* before moving to the edge of the relative clause. A MP derivation would discharge these dependencies via a series of combinations of Merge (both E and I). The final product of the syntactic derivation before Spell Out would look something like (26):[51]

51 I leave some details aside, most importantly this paper has completely ignored the labeling conventions and a discussion of Spell Out.
I here use conventional GB labeling (including phrasal marking) despite the fact that all MP accounts include some version of Bare Phrase Structure and so deny that labels include bar level information. Though this feature of labels is not contested, much else is (in particular whether or not labels are

(26) [$_{T/T}$ [after [$_{TP}$ Mary [$_{T'}$ T(past) [$_{vP}$ Mary(=Herself) [$_{v'}$ Mary [$_{v'}$ v [$_{VP}$ introduce Mary to the audience]]]]]]]] [$_{TP}$ She [$_{T'}$ T (past) [$_{vP}$ she(1) [$_{v'}$ v [$_{VP}$ turn [$_{PP}$ to [$_{DP}$ the [$_{NP}$ man [$_{CP}$ wh-man [$_{C'}$ that [$_{TP}$ she(2) [$_{T'}$ T (past) [$_{vP}$ [wh-man] [$_{v'}$ she(2) [$_{VP}$ meet [wh-man] before]]]]]]]]]]]]]]

(27) [$_{T/T}$ [after [$_{TP}$ Mary [$_{T'}$ T(past) [$_{vP}$ ~~Mary(=Herself)~~ [$_{v'}$ ~~Mary~~ [$_{v'}$ v [$_{VP}$ introduce Mary(=herself) to the audience]]]]]]]] [$_{TP}$ She [$_{T'}$ T (past) [$_{vP}$ ~~she(1)~~ [$_{v'}$ v [$_{VP}$ turn [$_{PP}$ to [$_{DP}$ the [$_{NP}$ man [$_{CP}$ ~~wh-man~~ [$_{C'}$ that [$_{TP}$ she(2) [$_{T'}$ T (past) [$_{vP}$ ~~[wh-man]~~ [$_{v'}$ ~~she(2)~~ [$_{VP}$ meet ~~[wh-man]~~ before]]]]]]]]]]]]]]

Some remarks on (26)/(27): As noted above, the EMH treats the reflexivization as a species of A-chain with the lower copy spelled out as a reflexive. Thus *Mary* moves through multiple theta positions, one copy ending up in Spec T where it gets nominative case and the lower copy realized as a reflexive. We avoided saying much about pronominalization, but if we adopt the (unlikely but interesting) proposal first mooted by Kayne, then *she* in the matrix can be a binder for the embedded *she*. There is no possible binding relation between *Mary* and *she* so this cataphoric relation falls outside the purview of the grammar (as in standard GB treatments). This EMH treatment of (26) would look like (28).[52]

(28) [$_{TP}$ [After Mary$_1$ introduced Mary/herself$_1$ . . .] [$_{CP}$ [$_{TP}$ She$_1$ Tns turned to the man [She$_1$ that [$_{TP}$ She$_1$ Tns met before]]]]]

On this analysis, the three copies of *she* form an improper A' chain spanning two theta positions, while two occurrences of *Mary* form an A-chain with two theta roles. The relation between *Mary* and *she* is not grammatically coded (note that *Mary* does not c-command the pronoun and so could not have been moved there).

Accusative case is assigned under movement to spec of vP in accordance with the EMH. More conventional MP accounts would likely leave it in situ in the syntax and discharge case under AGREE (an operation I have not discussed here).

---

available in the course of the syntactic derivation). The standard view is that in addition to Merge, Gs contain another operation Agree which allows for G interaction among non adjacent expressions and this is how (for example) case is discharged. The other view is that it is discharged under Merging a DP to the outer spec of its case "assigner". This latter conception requires something like labels. The former does not. I have not discussed in details how labels are assigned, what they are or where they come from (just as I have not discussed Agree). For the nonce, assume that labeling assigns the structures one finds in GB. Thus, *{meet, wh-man}* is a VP (the head being *meet*) and {*she*, {*v*, {*meet*, *wh–man*}}} is a vP (the head being *v*).

As for Spell Out, space limitations preclude detailed discussion. Suffice it to say that in English only one occurrence of an expression survives to be pronounced. (27) provides a plausible representation of the post Spell Out copies available for linearization.

52 An earlier version of this paper discussed this proposal and its virtues/vices, but space limitations forced it to drop out of the final version. Sorry.

In sum, the EMH would produce phrase markers with a larger number of copies and allow for movement into theta positions coupled with rules that convert copies into morphologically distinct pronominal forms (e. g. reflexives).

Before ending, let me note that many card carrying minimalists would disagree with this analysis.

References

Berwick, Robert C., Paul Pietroski, Beracah Yankama & Noam Chomsky. 2011. Poverty of the stimulus Revisited. *Cognitive Science* 35. 1207–1242.

Boeckx, Cedric, Norbert Hornstein & Jairo Nunes. 2010. *Control as Movement*. Cambridge University Press.

Bowers, John. 1973. Grammatical Relations. PhD thesis. Cambridge: MIT.

Chomsky, Noam. 1964. *Current Issues in Linguistic Theory*. Mouton: The Hague.

Chomsky, Noam. 1993. *The minimalist program*. Reprinted as Chapter 3, in Chomsky 1995.

Chomsky, Noam. 1995. *The minimalist program*. (Current studies in linguistics 28). Cambridge, Massachusetts: The MIT Press.

Haddad, Youssef & Eric Potsdam. 2013. Linearizing control relations: a typology. In Theresa Biberauer & Ian Roberts (eds.), *Principles of Linearization*, Berlin: Mouton de Gruyter.

Hornstein, Norbert. 2001. *Move!*. Oxford: Blackwell.

Hornstein, Norbert. 2007. Pronouns in a minimalist setting. In Jairo Nunes & Norbert Corver (eds.), *The Copy Theory of Movement. John Benjamins*, Amsterdam: Boston.

Hornstein, Norbert. 2009. *A Theory of Syntax*. Cambridge: Cambridge University Press.

Hornstein, Norbert, Jairo Nunes & Kleanthes K. Grohmann. 2005. *Understanding Minimalism*. Cambridge University Press: Cambridge.

Kayne, Richard S. 2002. Pronouns and their antecedents. In Sam D. Epstein & T. Daniel Seely (eds.), *Derivation and Explanation in the Minimalist Program*, Oxford: Blackwell.

Koster, Jan. 1984. On binding and control. *Linguistic Inquiry* 15. 417–459.

Lasnik, Howard & Mamoru Saito. 1991. On the subject of infinitives. In L. K. Dobrin, L. Nichols & R. M. Rodroguez (eds.), *Paper from the 27th regional meeting of the Chicago Linguistics Society 1991*, 324–343. Il: Chicago Linguistics Society, Cambridge, MIT. Chapter 3.

Lidz, Jeff & William Idsardi. 1998. Chains and Phono-Logical Form. In *Proceedings of 22nd Penn Linguistics Colloquium*. (Penn Working Papers in Linguistics 5.1).

Manzini, Maria R. 1983. On control and control theory. *Linguistic Inquiry* 14. 421–446.

Polinsky, Maria & Eric Potsdam. 2002. Backward control. *Linguistic Inquiry* 33. 245–282.

Sportiche, Dominique. forthcoming. Reconstruction Theory. Oxford: OUP.

Stabler, Edward. 2010. Computational perspectives on minimalism. In Cedric Boeckx (ed.), *Oxford Handbook of Linguistic Minimalism*, 616–641. Oxford: Oxford University Press.

Zwart, C. Jan-Wouter. 2002. Issues relating to a derivational theory of binding. In S. D. Epstein & T. D. Seely (eds.), *Derivation and Explanation in the Minimalist Program*, Oxford: Blackwell.

Ray Jackendoff and Jenny Audring

8 The Parallel Architecture

Abstract: The Parallel Architecture is a constructional approach to the knowledge of language based on two fundamental ideas. First, semantics, syntax, and phonology are independent combinatorial systems that interact through interfaces, often with mismatch. Second, the lexicon includes not only words but also idioms, collocations, and meaningful constructions. Crucially, it also includes all rules of grammar, which are stated in the form of declarative schemas with the form of pieces of linguistic structure containing variables; sentences are constructed by the operation of Unification over these schemas. This configuration allows syntax to be integrated gracefully with phonology, morphology, semantics, and other cognitive capacities, as well as making it possible to embed the theory of linguistic competence in the theory of performance.

1 Goals

The Parallel Architecture (PA; Jackendoff, 1997, 2002) is a constructional theory, close in many respects to HPSG (Pollard and Sag, 1994; Müller and Machicao y Priemer, this volume), LFG (Bresnan, 1982, 2001; Dalrymple and Findlay, this volume), and Construction Grammar (Goldberg, 1995, 2006; Croft, 2001; Hoffman and Trousdale, 2013; Chaves, this volume). It is intended to address the organization of language as a whole and its place in the mind/brain. Three major subcomponents have been developed in detail: Conceptual Semantics (Jackendoff, 1983, 1990, 2007a), Simpler Syntax (Culicover, 1999; Culicover and Jackendoff, 2005, this volume), and Relational Morphology (Jackendoff and Audring, 2016, forthcoming). However, PA also has implications for phonology, language processing, and language acquisition. The present chapter describes the framework and its bearing on syntactic theory. Culicover and Jackendoff (this volume) demonstrate how Simpler Syntax deals with a number of major syntactic phenomena.

The PA is rooted in a mentalist stance: knowing a language implies something stored in the mind of its speakers. Hence the PA reformulates the traditional concerns of generative linguistic theory as (1).

(1) What is the repertoire of linguistic structures available to a speaker of a language? In particular,
   a. What is stored in memory – both in the lexicon and in the grammar?
   b. In what form is this material stored?
   c. How are items in memory combined to create an unlimited number of novel utterances?

https://doi.org/10.1515/9783110540253-008

The theory should also engage with psychological concerns:

(2) How are linguistic structures deployed in the mind when producing and comprehending language?

(1) and (2) raise the question of how the speaker's linguistic repertoire comes about:

(3) How does the mind acquire linguistic knowledge? What prior resources does the mind require, and what does the learner require from the environment? How do these resources constrain possible languages?

Pushing still further, we can ask:

(4) Which aspects of the resources for language acquisition are specifically linguistic, and which belong to general mental phenomena such as the ability to categorize, imitate, and engage in social interaction?[1]

Goals (1)–(4) are articulated in Chomsky, 1965. Even if one questions Chomsky's own implementations of these goals (as the PA does), they still remain central.

Ideally, we would add questions such as:

(5) How can (1)–(4) be instantiated in neural storage and neural computation?

(6) How does the brain develop, such that the genetic code can lead to (5), and thence to (1)–(4)?

(7) How did the genetic code that creates a "language-ready brain" develop over the course of the biological evolution of our species?

These goals are largely beyond reach today. Although brain imaging and computational modeling have made some inroads,[2] it is unknown how something like even simple speech sounds are neurally instantiated, much less how biological development builds a brain. Still, linguistic theory should keep these goals in mind, in the hope of eventually closing the formidable gaps between linguistic theory, neuroscience, and developmental biology. (Fitch, 2010 is a good survey of current progress.)

A more immediate aspiration for the Parallel Architecture is *internal* integration. Here we find much contemporary theory wanting. The most influential contemporary approaches to syntax, semantics, and phonology certainly capture important insights about their own domains, but they each have their own formal machinery, only marginally compatible with the others. In contrast, the PA aspires to a theory of lan-

1 Many approaches – e. g. Cognitive Grammar (Langacker, 1987; Broccias, this volume), connectionism (McClelland and Rumelhart, 1986), some versions of Construction Grammar (Goldberg, 1995; Chaves, this volume; Bybee, 2010; Tomasello, 2003) – posit that *every* aspect of language arises from domain-general factors alone: there is no dedicated language faculty. The PA, however, takes the existence of language-specific aspects of mind to be an empirical issue.

2 Smolensky and Legendre (2006) offer an ambitious account of the connection between the "symbolic" digital character of linguistic representations and the more analog processes of neural computation.

guage in which phonetics, phonology, morphology, syntax, semantics, and pragmatics fit together gracefully.

We would also like a theory that affords a natural approach to signed as well as spoken language; to conversation, narrative, and song; to reading; to bilingualism and code-switching; to gesture; and to social aspects of language use. And we would like the theory to engage with accounts of other faculties of mind, affording explanations of, for instance, how we talk about what we see and how we use language to support reasoning (see Section 3.1).

2 Data

Given its broad goals, research within the Parallel Architecture framework is open to any sort of evidence: from introspection, corpus analysis, computational modeling, and genetics, as well as from experimental studies of processing, acquisition, and language impairment, using both behavioral techniques and brain imaging. The framework is also open to evidence from any language, from language variation and language change, and from other cognitive domains and from neuroscience.

In practice, the data motivating the Parallel Architecture have come predominantly from English, but with frequent reference to other languages, primarily Indo-European, but also including for instance Riau Indonesian (Gil, 2009), Pirahã (Everett, 1986, 2005), Al-Sayyid Bedouin Sign Language (Sandler et al., 2005) and Central Taurus Sign Language (Ergin, 2017) (see also Jackendoff and Wittenberg, 2014). The data has largely come from introspective judgments, but evidence has also come from corpus searches, psycholinguistic experimentation, and computational modeling (Culicover and Nowak, 2003).

An important focus of the data invoked by the Parallel Architecture has been unusual and understudied constructions. These “syntactic nuts” (Culicover, 1999) shed important light on the character of the grammar as a whole, as we will see in Section 3.2 below.

Finally, goal (4) above is to determine what aspects of the language faculty are domain-specific and what aspects are domain-general. This requires evidence concerning the structure of other cognitive domains. Accordingly, the PA has examined music (Lerdahl and Jackendoff, 1983; Jackendoff and Lerdahl, 2006), visual/spatial cognition (Jackendoff, 1987; Landau and Jackendoff, 1993), action planning (Jackendoff, 2007a), social cognition (Jackendoff, 2007a), and sequential visual images such as comic strips (Cohn, 2013).

3 Tools

The Parallel Architecture grows out of four fundamental issues: the arrangement and interaction of levels of linguistic structure, the relation between grammar and lexicon, unification-based combinatoriality, and the generative vs. relational functions of linguistic patterns. This section takes these up in turn.

3.1 Parallel combinatorial components and interfaces

The basic premise of the Parallel Architecture is that linguistic structure is determined by three independent generative systems – phonology, syntax, and semantics – plus, crucially, the linkages between them. Similar ideas appear in Stratificational Grammar (Lamb, 1966), Lexical-Functional Grammar (Bresnan, 1982, 2001; Dalrymple and Findlay, this volume), Autolexical Grammar (Sadock, 1991), Role and Reference Grammar (Van Valin and LaPolla, 1997), and others. This contrasts with the traditional "syntactocentrism" of generative grammar (Chomsky, 1965, 1981, 1995), which assumes without argument that the only "generative engine" in the grammar is syntax, and that phonology and semantics are derived from syntactic structure.[3]

The independence of phonology from syntax is motivated by the fact that phonological structure includes syllable and foot structure, a metrical grid, an intonation contour, and, where appropriate, a tone tier, none of which can be derived from syntactic structure (Goldsmith, 1979; Liberman and Prince, 1977; Selkirk, 1984). Similarly, in any substantive theory of meaning,[4] semantic structure is built out of entities such as (conceptualized) objects, events, times, and places, rather than NPs and VPs.

Syntactocentrism presumes that phonological and semantic constituency are determined from syntax. However, constituency often fails to match across components, as seen in (8), for example.

(8) a. *Phonology-syntax mismatch*:
Syntax: [[Sesame Street] [is [a production [of [the Children's Television Workshop]]]]]
Phonology: [Sesame Street is a production of] [the Children's Television Workshop]

b. *Syntax-semantics mismatch*
Syntax: that travesty of a theory [*travesty* is head, *theory* is modifier]
Semantics: 'that theory, which is a travesty' [*theory* is head, *travesty* is modifier]

---

3 Chomsky makes this assumption explicit several times in *Aspects* (Chomsky, 1965, 16, 17, 75, 198). To our knowledge he has never defended it or even questioned it since, at least on empirical grounds.

4 In addition to Conceptual Semantics (the semantic component of PA), such approaches include formal semantics (Heim and Kratzer, 1998), Cognitive Grammar (Lakoff, 1987; Langacker, 1987; Talmy, 1978; Broccias, this volume), and approaches from artificial intelligence (e. g. Schank, 1973; Minsky, 1975; Rumelhart, 1980).

Syntactocentric approaches account for these mismatches by positing covert syntactic structure that matches the semantics, related to the surface by movement and deletion. In contrast, the PA accounts for these misalignments through *interface principles* that link two kinds of structure. Between phonology and syntax, the default linkage matches constituency and linear order; and between syntax and semantics, the default linkage matches head-argument relations. But non-default linkings such as those in (8) are endemic (see also (22)–(24) below).

The upshot is a conception of grammar like Figure 1. The formation rules define well-formed structures in their respective components. The interfaces define permissible linkings between structures in two domains. Hence a well-formed sentence has well-formed structures in each of the three domains, plus well-formed links among the structures. (The direct link between phonology and meaning in Fig. 1 cannot exist in a syntactocentric grammar, but it is altogether possible in PA, for example to link intonation contour in phonology directly to information structure in semantics.)

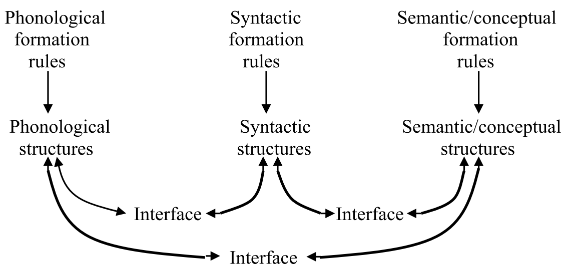


Figure 1: The Parallel Architecture.

The individual components in Fig. 1 have similar fine-scale architecture. Phonology is itself composed of independent generative tiers for segmental structure, tone, metrical structure, and prosodic contour, all linked by interfaces. Similarly, semantics contains quasi-independent but linked subcomponents of argument/event structure and information structure (topic and focus). In syntax, f-structure in LFG (Dalrymple and Findlay, this volume) and the grammatical function tier in Simpler Syntax (Culicover and Jackendoff, this volume) are likewise independent tiers from phrase structure. Exactly what components are necessary, what each of them accounts for, and how they interface with the others are empirical issues (see Section 5.1).

Because its components are related by interfaces rather than derivations, the PA can situate the language faculty comfortably in an overall view of the mind. In particular, it claims that the internal components of language are connected to each other in the same way as language is connected with nonlinguistic mental structure, and in the same way as other faculties of mind are connected with each other, as in Figure 2. For instance, in order to talk about what we see, linguistic representations must

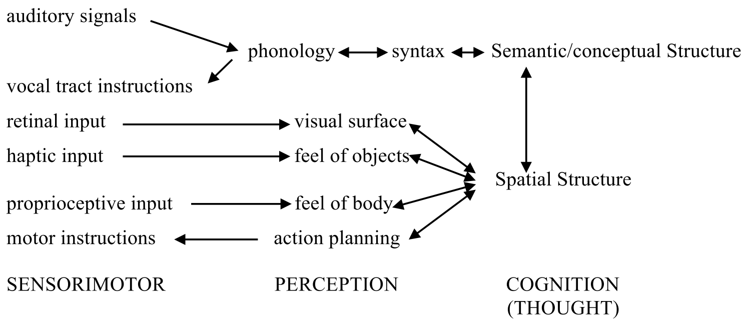


Figure 2: Overall layout of cognitive capacities.

be connected to visually-based representations of how objects look and how they are situated in the environment. Such representations cannot be derived from language or vice versa: vision and language have to connect through a system of linking principles (Jackendoff, 1987, 1996; Landau and Jackendoff, 1993). Similarly, phonological structures must be linked to auditory input and to vocal tract motor instructions by interface principles that have nothing to do with syntax (Jackendoff, 1997, Section 2.1.).

Moving to the rest of the mind, visually based representations of shape and spatial layout must be correlated with representations derived by hapsis (the sense of touch) and proprioception (the body senses). None of these can be derived from the others; rather they must be related by principles that establish equivalences between them – interfaces in our sense. Navigation and manipulating objects can be visually guided, through principled correspondences between visual perception and the formulation of action. This overall uniformity is impossible to achieve in a classical framework. (See Jackendoff, 2011 for more discussion.)

The PA also offers connections to additional domains of linguistic structure. For instance, sign language has phonology built from structured gesture, with interfaces to vision and the motor system. Formal versus casual speech register can be treated as an interface between perceptions of social situation and the use of particular words or expressions. Orthography has internal principles that define the alphabet, plus an interface to the visual system, so that the reader can see writing, plus an interface to phonology and/or morphology that stipulates the correspondence between spelling and speech.[5] (See Jackendoff and Audring, forthcoming.)

5 One near-universal principle of spelling is that orthographic linear order corresponds to phonological temporal order. However, even this basic principle is violated in English orthography (a) by silent *e*, which affects vowels two segments to the left (*can*/*cane*, *pin*/*pine*), and (b) by combinations like <$5>, read in the opposite order, 'five dollars'. Similarly for German numerals, e. g. <29> = 'neunundzwanzig' ('nine-and-twenty').

3.2 The structure of the lexicon

We return to the goals in (1), restated here:

(9) a. What is stored in memory that constitutes one's knowledge of language?
b. In what form is this material stored?
c. How are items in memory combined so as to make it possible to create an unlimited number of novel utterances?

Traditional generative grammar answers that words (or morphemes) are stored in the *lexicon*, while rules constitute the *grammar* and are responsible for all combinatorial processes. Lexical items are taken to be arbitrary associations of phonological, syntactic, and semantic structure. Anything that can be predicted by a rule is absent: the lexicon is "simply an unordered list of all lexical forms" (Chomsky, 1965, 84), "really an appendix of the grammar, a list of basic irregularities" (Bloomfield, 1933, 274). This encourages the view that the grammar and the lexicon are altogether different kinds of knowledge and are stored differently in the brain (as has in fact been claimed by Ullman, 2015).

The PA, along with Cognitive Grammar and especially Construction Grammar, rejects this strict distinction between grammar and lexicon. Rather, rules of grammar are stated in the same format as words: they are pieces of stored linguistic structure, and hence can be considered lexical items. There is a continuum between words and rules: a lexical item is more word-like if it consists of fully specified material; it is more rule-like to the extent that it contains variables (underspecified material). Construction Grammar calls the latter items *constructions*; following the terminology of Construction Morphology (Booij, 2010), we will call them ***schemas***.[6]

Let us illustrate the continuum between words and schemas. Every theory views a word as linking a piece of phonology, a piece of meaning, and a collection of syntactic features such as category, number, grammatical gender, and so on. These components are customarily enclosed in square brackets. Our notation instead coindexes the three components, as in (10).

(10) Semantics: $[\text{CAT}]_1$
Syntax: $[\text{N}, \textsc{sg}]_1$
Phonology: $/\text{kæt}/_1$

We take the subscripts to mark the ends of association lines: *this* piece of semantics is linked to *these* syntactic features and *this* pronunciation. Upon hearing /kæt/, one

6 Radical Construction Grammar (Croft, 2001) extends the term *construction* even to words like *dog*. We prefer the term *lexical item*. We also avoid the term *constraint*, used in "constraint-based" theories such as LFG, HPSG, and Optimality Theory (Legendre, this volume). A "constraint" *constrains* what you can say, or stipulates what you *can't* say or want to *avoid* saying. We prefer to think of schemas as encoding affordances for expression – generalizations about what you *can* say.

can posit a noun in syntax and CAT in semantics; and one can express the meaning CAT by connecting it to a noun in syntax and to /kæt/ in phonology. In other words, a word is a small piece of the interface components in Fig. 1, and the coindex serves as an ***interface link***. There is no separate "lexical interface"; words participate in the more general interfaces among the three structures.

(10) is a stereotypical word, with structure in semantics, syntax, and phonology. But some words lack one or more of the components. For instance, the words in (11) can occur alone as full utterances, and they do not combine with other words, except by parataxis (12a) and in quotative and metalinguistic contexts (12b,c), into which anything can be inserted, even a phrase of another language.

(11) *Phonology and semantics, no syntax*:
hello, ouch, yes, oops, gosh, dammit, upsy-daisy, allakazam

(12) a. Hello, Bill.
b. "Hello," she said.
c. the word *hello*

The PA characterizes these words as linkages of phonological and semantic structure. They offer no evidence for syntactic category or for a more canonical "underlying" syntactic structure from which they can be derived. Lacking syntactic features, they appear only in contexts where syntactic features play no role.[7]

A few words of English, underlined in (13), have syntax and phonology but no meaning, functioning only as "grammatical glue."

(13) *Phonology and syntax, no semantics*:
a. <u>It</u>'s hot in here.
b. <u>Do</u> you want a drink?
c. a picture <u>of</u> Bill
d. I forgot <u>that</u> it's going to rain.
e. He wants <u>to</u> get out <u>of</u> here.

There are even stored pieces of phonology with neither syntax nor meaning. Their function is just to fill up metrical structure in nursery rhymes and songs. Someone who knows the lyrics in question knows these items.

(14) *Phonology, no syntax or semantics*:
fiddle-de-dee, hey-diddle-diddle, e-i-e-i-o, hickory-dickory-dock, eenie-meenie-minie-moe

7 These items are examples of the dashed phonology-semantics interface in Fig. 1. Hence they are problematic for a syntactocentric theory, in which semantics and phonology must be derived from syntax.

The simplest treatment of these items is to put them in the lexicon alongside ordinary words. They (mostly) obey English phonotactics, so they have phonological structure. But lacking syntax and semantics, they cannot appear in ordinary sentences except as phonological intrusions. In other words, their intrinsic properties account for their distribution, and the theory need not distinguish them further, say as elements of a "babblecon."[8]

Words with argument structure have variables in their lexical representation, so they are already somewhat schema-like. (15) shows the verb *devour*. The underlined parts are variables that must be instantiated.

(15) Semantics: $[_{\text{Event}}\ \text{DEVOUR}_2\ (\text{Agent: }\underline{\text{X}}, \text{ Patient: }\underline{\text{Y}}_{\text{y}})]$
Syntax: $[_{\text{VP}}\ \text{V}_2\ \underline{\text{NP}}_{\text{y}}]$
Phonology: $/\text{dəvawr}/_2$

Coindex 2 links the semantics with a verb in syntax and with the pronunciation *devour*. The semantic structure is an Event with an Agent and a Patient (plus further omitted details). The Agent (variable *X*) is linked to subject position in syntax by principles we won't describe here (see Jackendoff, 1990); the Patient (variable *Y*) is linked to the direct object position in syntax. The syntactic structure contains a variable NP, which marks this verb as obligatorily transitive, contrasting with, say, *eat*, whose Patient need not be expressed.[9] The coindex that encodes the linking of Patient to direct object is a variable *y* rather than a number. This reflects the fact the Patient variable in semantics is linked to whatever instantiates the direct object variable in syntax.

The lexicon also contains units larger than words: thousands of idioms, clichés, and other "prefabs." For example, *chew the fat* has phonological structure, the syntactic structure of a VP, and a semantic structure approximately equivalent to *converse idly* (again, the variable *X* is the Agent argument).

(16) Semantics: $[_{\text{Event}}\ \text{CONVERSE\^{}IDLY}\ (\text{Agent: }\underline{\text{X}})]_3$
Syntax: $[_{\text{VP}}\ \text{V}_4\ [_{\text{NP}}\ \text{Det}_5\ \text{N}_6]]_3$
Phonology: $/\text{čuw}/_4\ /\text{ðə}/_5\ /\text{fæt}/_6$

What makes (16) an idiom is that the entire VP is co-indexed with the meaning (coindex 3), but its words are not (coindices 4, 5, 6). Hence the whole means something different from its parts. We know that (16) has internal syntactic structure, because

8 The items in (13) and (14) are problematic for those versions of HPSG, Cognitive Grammar, and Construction Grammar that insist that every lexical item is a sign that pairs a form (phonology and syntax) with a function (semantics). These items have form, but they have no function in the intended sense.
9 This notation expresses subcategorization without introducing subcategorization features such as [+ ___ NP] or abstract case-marking features.
A complication: In Simpler Syntax, as in LFG, the semantic variable Y is actually linked to the second position in the Grammatical Function tier, which is realized in syntax as either the direct object of an active or the subject of a passive. See Culicover and Jackendoff (this volume, 2005, chapter 6).

the verb conjugates like a normal verb: its past tense is *chewed the fat*, not **chew the fatted*. Again, idioms need not be housed in another "place" in the grammar, separate from words; they are idioms simply by virtue of their internal syntactic structure and their noncompositional meaning.[10]

Stored items need not be semantically idiosyncratic. Clichés and collocations like (17) mean pretty much what they should mean compositionally, but they are still identifiable as stored pieces of English that are part of native command of the language. (Examples from the Wheel of Fortune corpus, Jackendoff, 1997; see also Corrigan et al., 2009.)

(17) baby-blue eyes, open twenty-four hours, quiet as a mouse, reckless driving, rosy cheeks, see you later

The idioms in (18) have argument structure: they take a direct object, encoded as a variable in the idiom's syntactic structure. This is linked to a variable in semantics, parallel to the variables in *devour* (15).

(18) take NP for granted
put NP on ice
give NP the once-over

Most idioms have canonical syntactic structure; for instance (17)–(18) are standard VPs. However, the syntax of idioms like (19) is unusual.[11]

(19) day in day out
by and large
for the time being
all of a sudden
over (and over) (again)

Some idioms with non-canonical structure, such as those in (20), serve as full utterances, and they do not embed except in quotative contexts and possibly in indirect speech.

(20) How about XP? (cf. **I'm wondering* (*whether*) *how about lunch*.)
Prt with NP! (*off with his head*, *down with the government*, etc.)
(cf. **The crowd demanded* (*that*) *off with his head*.)
far be it from NP to VP (cf. **Fred said that far be it from him to give up*.)
suffice it to say that S (cf. **Henk said that suffice it to say that* . . .)]

10 The existence of (16) does not preclude a literal interpretation as well. On the partial mobility of only some idioms (*the cat was let out the bag*; **the bucket was kicked*), see Jackendoff (1997, 166–171).
11 Distributed Morphology (Marantz, 1997; Embick and Noyer, 2007) constructs the syntax of idioms by normal syntax, and the Encyclopedia provides the idiomatic meaning. However, this offers no way to generate idioms with idiosyncratic syntax such as (19) and (20).

Another class of “constructional idioms” have canonical syntax but unusual semantics. These play a major role in both the Parallel Architecture and Construction Grammar. Consider (21) (Jackendoff, 1990, Goldberg, 1995).

(21) *Way*-construction:
$[_{\mathrm{VP}}$ V pro’s way PP]
Jerry joked his way out of the meeting. (= ‘Jerry went out of the meeting joking’)

Syntactically, (21) is a typical verb phrase. However:

– The verb *joke* doesn’t usually take a direct object, but (21) has the phrase *his way* apparently in object position. The verb cannot have a direct object of its own (**Jerry told jokes his way out of the meeting*).
– *Joke* doesn’t normally select a path expression (**Jerry joked out of the meeting*).
– The sentence means that Jerry *went* out of the meeting, even though there is no verb of motion.
– *Joke* describes the manner in which he went out of the meeting (alternatively, the means by which he got out of the meeting).

Hence (21) involves an absurdly non-canonical mapping between syntax and semantics. Any verb of the appropriate syntactic and semantic type is possible: you can drink your way across the country or knit your way through a conference.

The construction can be formalized approximately as the schema in (22), which links semantic, syntactic, and phonological structures. It is noncanonical because the verb is linked to a manner (or means) modifier in semantics (coindex *z*), while the semantic head is the function GO, which does not link to syntax at all. Meanwhile, the direct object in syntax links to the phonology *way* but not to semantics (coindex 8), and the PP links to the path of motion.[12] (We omit notation for binding the pronoun to the subject.)

(22) *Way-construction*
Semantics: $[\mathrm{GO\ (Agent{:}\ }\underline{\mathrm{X}},\ \underline{\mathrm{Path}}_{y});\ \mathrm{WHILE\ (}\underline{\mathrm{F}}_{z}\mathrm{\ (Agent{:}\ }\underline{\mathrm{X}})]_{7}$
Syntax: $[_{\mathrm{VP}}\ \underline{\mathrm{V}}_{z}\ [_{\mathrm{NP}}\ \underline{\mathrm{Pro}}\mathrm{+Poss\ N}_{8}]\ \mathrm{PP}_{y}]_{7}$
Phonology: $/\mathrm{wej}/_{8}$

(23) illustrates further constructional idioms, with different semantics, marked by *away* and *head* (or other body part) *off*.

(23) a. *Time-away construction*:
$[_{\mathrm{VP}}\ \underline{\mathrm{V}}\ [_{\mathrm{NP}}\ (\underline{\textit{time}})]\ \mathrm{away}]$
Fred drank the afternoon away. (= ‘Fred spent/wasted the afternoon drinking’)

12 Goldberg, 1995 has a slightly different account of the role of *way*.

b. *Head off construction*:
[$_{VP}$ V Pro's head/tush/butt off]
Suzie sang her head off. (= 'Suzie sang a lot/intensely')

Knowing these constructions is part of knowing English; there are not precise cognates even in its close relatives German and Dutch. For each construction, a speaker has to learn and store its syntactic structure, how its constituents correspond to semantics in other than the normal way, and the phonology of the designated elements *way*, *away*, and *head off* that signal that something unusual is going on.

Other constructions of this sort have no distinguishing phonological content, so they depart still farther from canonical word- and idiomhood.

(24) a. *Sound+motion construction*:
[$_{VP}$ V PP]
The bus rumbled around the corner. (= 'The bus went around the corner, rumbling')

b. *Inverted NP construction*:
[$_{NP}$ a/this/that N of an N]
that travesty of a theory (= 'that theory, which is a travesty')

c. *Light verb construction*:
[$_{VP}$ V NP NP]
Pat gave Kim a hug. (= 'Pat hugged Kim')

d. Casual paratactic conditional:
[S, S]
You break my chair, I break your arm. (Ed Merrin, 12 June 2014)

In (24a), the bus *went* around the corner, making rumbling sounds, but without a verb of motion. In (24b) (= (8b)), the syntactic head is *travesty*, and *theory* is a modifier. But semantically, *theory* is the referent of the expression and *travesty* is an evaluative modifier. In (24c), the direct object *a hug* provides the semantic content normally expressed by a verb, and the verb is a dummy that contributes aspectuality and provides argument positions. Finally, (24d) has no *if-then*, but the conditional meaning is perfectly clear.

A speaker has to learn and store each of these rule-like constructional idioms as an association between a syntactic complex and a semantic complex. But the basic formalism is the same as for words – except that the syntax is composite, as with idioms, and in these cases there is no associated phonology.

Again, it is important to recognize that constructional idioms are not rare oddities: languages are full of them. Within the PA, they emerge as stored pieces of noncanonical syntax-semantics correlations containing variables, sometimes with accompanying phonology, sometimes not.

Now: The very same formalism can be used to state phrase structure rules as syntactic schemas, without any association to phonology or semantics. For instance, the traditional rule (25a) for the English transitive VP can be restated as the declarative schema (25b), a syntactic "treelet" in the sense of Fodor (1998).

(25) a. VP → V – (NP) – . . .
b. Syntax: $[_{\mathrm{VP}}\ \underline{\mathrm{V}} - (\underline{\mathrm{NP}}) - \ldots]$

Such pieces of pure syntactic structure are also basic building blocks for Tree-Adjoining Grammar (Joshi, 1987; Frank and Kroch, 1995).

How should we think of schema (25b)? Extending the analysis of the phenomena in (11)–(24) just one step further, we conclude that (25b) is just another sort of lexical item. The lexicon has items with just phonology like *fiddle-de-dee*, items with just phonology and syntax like do-support *do*, items with just phonology and semantics like *hello*, and items such as idioms and meaningful constructions with phrasal syntactic structure and syntactic variables. (25b) is an item with only syntactic structure – just one more logical possibility. Its variables license novel, freely combined VPs in syntax.

The upshot of all this is that the lexicon – the storehouse of linguistic knowledge – looks very different from the traditional lexicon. Stereotypical words are fully specified lexical items; stereotypical rules (now stated as schemas) are lexical items that consist entirely of variables. In between are many intermediate cases: words that subcategorize complements (and therefore contain variables), idioms, and meaningful constructions with or without phonology. These items are all encoded as pieces of linguistic structure in one or more domains; if in more than one domain, the lexical item specifies the interface links between them.

This continuity between words and rules is an important feature of the Parallel Architecture, shared with Construction Grammar and Cognitive Grammar. It contrasts with LFG and HPSG, which maintain a strict lexicon/grammar distinction. It contrasts more drastically with mainstream generative grammar, whose lexicon is an unstructured "list of exceptions" (Chomsky, 1965) or a collection of "syntactic atoms" (Berwick and Chomsky, 2016), and whose grammar consists of procedural rules.

3.3 Unification-based combinatoriality

We next focus on the role of schemas in the grammar, comparing them to traditional rules. Traditional rules are procedural: combine *this* item with *that* item, move *this* item to *here*, delete *this* item in *this* context. In contrast, schemas are declarative: they stipulate pieces of structure and interface links among them. In order for PA to build novel utterances, a procedural component is necessary. Here PA, like other constraint-based theories, adopts Unification (Shieber, 1986) as the appropriate procedural operation.

Unification is a sort of Boolean union over structures: it *superimposes* one structure on another, preserving the unique parts of both, without doubling the shared parts. For instance, unifying the string ABC with the string BCD yields ABCD, not, say, ABCBCD.

To illustrate how Unification works, suppose we wish to generate the phrase *that cat*. This requires three pieces of stored structure: the two words and the phrase structure schema for (this part of) the NP.[13]

(26)	Semantics:	a.	$[\text{DISTAL}]_{11}$	b.	$[\text{CAT}]_{1}$	c.	(*no intrinsic semantics*)
	Syntax:		$[\text{Det, SG}]_{11}$		$[\text{N, SG}]_{1}$		$[_{\text{NP}}\ [\underline{\text{Det}}, \underline{\alpha}\text{NUM}], [\underline{\text{N}}, \underline{\alpha}\text{NUM}]]$
	Phonology:		$/\text{ðæt}/_{11}$		$/\text{kæt}/_{1}$		(*no intrinsic phonology*)

(26a) unifies with the part of schema (26c) whose variables it satisfies; (26b) does the same, yielding (27).

(27) Semantics: $[\text{CAT}_1; \text{DISTAL}_{11}]_{12}$
Syntax: $[_{\text{NP}}\ [\text{Det, SG}]_{11}\ [\text{N, SG}]_{1}\ ]_{12}$
Phonology: $/\text{ðæt}_{11}\ \text{kæt}_{1}\ /_{12}$

Four remarks: First, phrase structure schema (26c) stipulates the word order in syntax, which maps canonically into linear order in phonology. Second, the alphas in schema (26c) stipulate that the noun and determiner must agree in number. In (27) they indeed do agree. But in **those cat* and **that cats* they do not, so these are syntactically ill-formed. Third, DISTAL and CAT supplement each other semantically, so they are in effect conjoined into [CAT; DISTAL]. Fourth, the entire unified expression receives an interface link (coindex 12), to associate the meaning of the whole with the NP and the entire phonological string. We assume this is an automatic consequence of Unification.

For a slightly more complicated case, consider again the lexical representation of *devour*.

(28) Semantics: $[\text{DEVOUR}_2\ (\text{Agent: } \underline{\text{X}}, \text{Patient: } \underline{\text{Y}}_{\text{y}})]$
Syntax: $[_{\text{VP}}\ \text{V}_2\ \underline{\text{NP}}_{\text{y}}]$
Phonology: $/\text{dəvawr}/_{2}$

The selectional restrictions of *devour* are encoded as properties of its semantic variables. *X* in (28) is not just an Agent, it is an animate Agent; and *Y* is not just a Patient, it is an object or substance with appropriate properties or affordances. When *devour* is unified with its arguments, these features of the variables are unified with the features of the corresponding arguments. If the argument is semantically underspecified,

13 DISTAL has structure that denotes the spatial relation of a denoted object to the speaker or hearer. We set these details aside.

as in <u>*It*</u> *devoured the pie*, the verb supplies the requisite semantic detail through Unification: *it* denotes an animate. If the argument's meaning conflicts with the variable's, the sentence may be coerced into a metaphorical reading, as in *Evil thoughts devoured me*, or, failing that, it will be judged anomalous, as in **My pencil sharpener devoured The Hague*.[14]

3.4 Generative and relational functions of schemas

The PA thus separates the traditional function of a phrase structure rule into two components: a declarative schema that encodes the desired structure, plus the process of combining the parts through Unification. Unification is hence the only *procedural* rule in the grammar.[15] Using Unification to combine words and schemas enables the creation of an unlimited number of novel utterances, like traditional procedural rules. We call this the ***generative*** function of schemas.

However, schemas also have a second function. Consider again a VP idiom such as *chew the fat*. Because of its idiosyncratic meaning, it must be stored in the lexicon. Yet because its syntactic structure is a canonical VP, traditional generative grammar must, paradoxically, construct it *outside* the lexicon. The PA offers a resolution: a schema can be used not only to generate new utterances, it can also capture generalizations among existing lexical items, thereby partially motivating them, decreasing their arbitrariness (Culicover et al., 2017). Thus the VP schema (25b) and the NP schema (26c) together motivate the whole class of VP idioms such as *chew the fat*, *kick the bucket*, and *bite the bullet*, making them less arbitrary than the syntactically exotic *by and large*, *for the time being*, and *day in day out*. We call this use of schemas their ***relational*** function. Traditional generative grammar sometimes attributes this function to "lexical redundancy rules," distinct from the generative rules. In the PA, both functions can be performed by the very same schema.[16]

Crucially, not all schemas have both functions. Some have only the relational function: their instances are all listed, and it is impossible to create new ones. One example is the little determiner construction in (29).

(29) a. [$_{\text{Det}}$ such/what/quite/many/hardly a] tree
  b. [$_{\text{Det}}$ [$_{\text{AP}}$ this/that/so/too/how/as tall] a] tree

14 The sentence may still be acceptable if embedded under a predicate that selects for anomaly, as in *It makes no sense to speak of pencil sharpeners devouring The Hague*! For more detail on Unification within the PA, see Jackendoff (1990) (called there Argument Fusion), Jackendoff (2011).

15 This position excludes the possibility of movement rules, the procedural rules par excellence. See Section 5.2.

16 This dual function of schemas is suggested by Jackendoff (1975), Aronoff (1976), and Barlow and Kemmer (1994). However, it makes more sense in the PA, whose rules are all in the lexicon, than in more traditional frameworks where some rules are "in the grammar" and some "in the lexicon."

The overall pattern is a determiner (if one can call it that), of the form *X a*. The pattern has six instances: the five in (29a), plus the subpattern in (29b) of a degree word, an adjective, and *a*. Novelty is possible only in the choice of adjective and noun: e. g. *that beautiful a song*, *so deep a lake*. Otherwise, the pattern is closed; no general, productive rule captures this peculiar range of possibilities. Yet it is certainly a pattern of English, worthy of a schema, even if it only captures a generalization among listed lexical items. Such nonproductive schemas are rather rare in syntax, but they are commonplace in morphology, especially derivational morphology.

Could schemas exist that are used only productively, to create novel structures? No, because any novel form can be memorized; and once memorized, its structure has to be accounted for relationally, like idioms and stored collocations. Hence *all* schemas can be used relationally, but only a subset can be used generatively. Expanding the knowledge of language to include nonproductive patterns enlarges the scope of linguistic theory in what we consider an important fashion. (Jackendoff and Audring, forthcoming show how productive schemas are distinguished from nonproductive.)

The notion of schemas with variables extends readily to poetic forms such as sonnets and limericks. A text counts as a limerick if it conforms to a particular metrical pattern and a particular rhyme scheme; moreover, a large class of limericks begin with the text *There once was a X from Y, who. . . .* These characteristics can be encoded as a schema for the limerick form, which captures the structure of existing limericks and can also be used to construct novel exemplars. Knock-knock jokes even have a two-person schema; both participants have to know it to perform the joke properly.

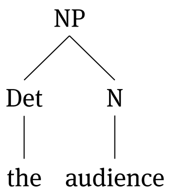
Beyond language, abstract musical forms such as 12-bar blues can be specified as abstract schemas of meter and harmony, to be fleshed out by the creator's choice of conforming notes. And it is plausible to think of knowledge of rituals (formal and informal, including games) as schemas that stipulate what the participants are expected to do. The actual participants on any particular occasion then satisfy the variables in the schema. This is the idea behind Schank's (1973) "scripts," Minsky's (1975) "frames," Fillmore's (2003) frame semantics, Rumelhart's (1980) "schemas," and Goffman's (1974) elaborate "frames." In short, the formal tools of the PA are to some degree domain-general; they did not have to be invented just for syntax.

4 Sample analysis

For a fuller picture of linguistic structure in the Parallel Architecture, we offer an analysis of the clause *After Mary introduced herself to the audience*,.... (Culicover and Jackendoff, this volume, treat the second assigned clause, *she turned to a man that she had met before*.)

From a traditional perspective, what stands out here is that the words are not terminal elements in the syntactic tree, as in the usual notation (30).

(30)



"*The*" and "*audience*" in (30) stand for complexes of semantic, syntactic, and phonological features. The PA insists that semantic and phonological features belong not to syntax, but to their own proprietary structures, *linked* to the syntax. Fig. 3 reflects this perspective: each word is distributed among the three structures. For instance, *audience* consists of the three pieces of structure coindexed 11 in Fig. 3.

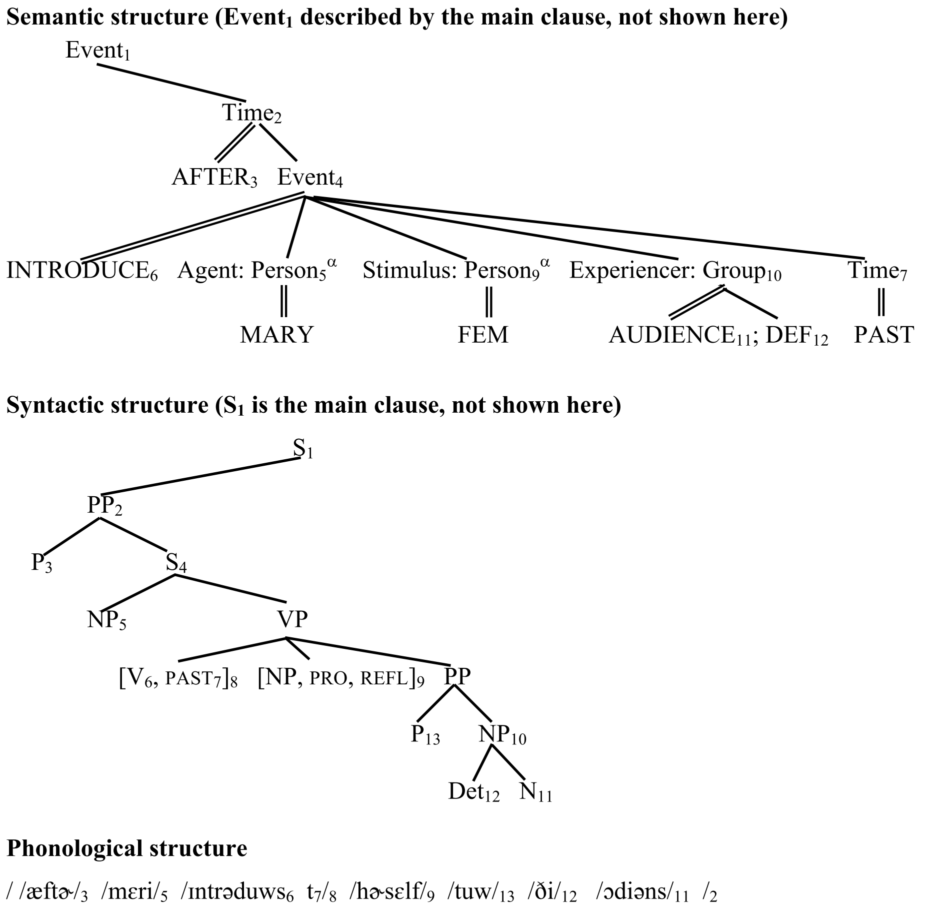


Figure 3: Structure of *After Mary introduced herself to the audience*.

The phonological structure in Fig. 3 has been simplified, omitting syllable structure, stress, and intonation. Double slashes at the beginning and end surround the entire string, which is linked with the entire clause in syntax and semantics (coindex 2).

Semantic structure is notated as a tree, somewhat more perspicuous than the labeled bracketing used so far; the two are intended as notational variants. A function such as INTRODUCE is connected to its mother by a double line; its arguments, for instance MARY, are connected to the same mother by a single line. Modifiers are connected by a dashed line.

The semantic structure in Fig. 3 says that an Event (coindex 1), described by the clause omitted here (*she turned to a man*...), occurs at a Time (coindex 2) after another Event (coindex 4). In this latter Event, set in the past, a Person (Mary) introduces a female Person to a Group which is an audience and definite. The two Persons are coreferential, as indicated by the co-superscript αs; that is, coreference is encoded in semantic structure rather than (or in addition to) syntactic structure. The function INTRODUCE itself is a sort of causative psych predicate: roughly, Mary (an Agent) brings herself (a Stimulus) to the attention of the audience (an Experiencer).

Turning finally to the syntactic structure: Its terminal elements link to words in the phonology, and most of its constituents link to semantic structure. The structure itself is straightforward; we mention only four points. First, the name *Mary* (coindex 5) and the reflexive pronoun *herself* (coindex 8) are full NPs; other treatments are possible. Second, *after* (coindex 3) is semantically a function of an Event or Time. Syntactically, it is a preposition whose complement can be either NP or, as in this case, S. It shares this semantic structure and subcategorization with e. g. *before* and *until* (Klima, 1965; Jackendoff, 1977).

Third, the preposition *to* (coindex 12) has structure in syntax and phonology but not in semantics. No other preposition is possible in this position. Rather, *to* functions like a quirky case-marker or idiom chunk: it is determined by the verb. This stipulation can be encoded in the lexical entry for *introduce* as in (31) (reverting for convenience to labeled bracketing).

(31) Semantics: $[_{\text{Event}}\ \text{INTRODUCE}_6\ (\text{Agent: } \underline{\text{X}}, \text{Stimulus: } \underline{\text{Y}}_{\text{y}}, \text{Experiencer: } \underline{\text{Z}}_{\text{z}})]$
Syntax: $[_{\text{VP}}\ \text{V}_6\ \underline{\text{NP}}_{\text{y}}\ [_{\text{PP}}\ \text{P}_{12}\ \underline{\text{NP}}_{\text{z}}]\ ]$
Phonology: /ɪntrəduws/$_6$ /$\underline{...}$/$_{\text{y}}$ /tuw/$_{12}$ /$\underline{...}$/$_{\text{z}}$

(31) has a specified verb, like *devour* in (28), plus a specified preposition, with variable direct and oblique objects, corresponding to Stimulus and Experiencer respectively.

The last point of interest is the treatment of tense. Syntactic category (here, V) and inflection (here, PAST) are treated as features of the node in the syntactic tree (Jackendoff and Audring, forthcoming). Thus the regular past tense is a Time modifier in semantics, a morphosyntactic feature in syntax, and a suffix in phonology, linked by coindex 7 – another mismatch among the components of the grammar. This mismatch is captured directly in the interface links, rather than by any sort of movement. Accordingly, a simplified lexical entry for the regular past tense (ignoring the *d/t/ed* allomorphy) can be formulated as (32).

(32) Semantics: $[_{\text{Situation}}\ \underline{\text{X}};\ [_{\text{Time}}\ \text{PAST}]_7\ ]_y$
Syntax: $[_{\text{S}}\ \underline{\ldots}\ [\underline{\text{V}}_x,\ \textsc{past}_7]\ \underline{\ldots}\ ]_y$
Phonology: $/\ \underline{\ldots}\ /\ /\underline{\ldots}_x\ \text{t}_7/\ /\underline{\ldots}/_y$

Verbs with irregular past tenses have different links to phonology (see Jackendoff & Audring, forthcoming). But whether the verb is regular or irregular, the syntax contains morphosyntactic PAST, which can link not only to semantic PAST, but also to hypothetical conditional (e. g. *What if I* <u>*came*</u> *tomorrow?*). So (32) is only part of the structure associated with the syntactic past tense.

A more traditional semantic account treats the Tense as an operator over the rest of the sentence, as in (33).

(33) $[_{\text{Situation}}\ \text{PAST}_7\ (\ [_{\text{Event}} \ldots]_4\ )]$.

The same problems of mismatch arise, just in a different configuration.

Summing up, linguistic structure in the PA encompasses three linked structures. In simple cases, the constituents of the three components align; but, as also seen in Section 3.2, numerous mismatches are possible, including missing components (no *to* in semantics) and mismatched hierarchical structures (past tense).

5 Evaluation

Evaluation of the Parallel Architecture framework encompasses two different endeavors. "Internal" evaluation compares alternative treatments of a phenomenon *within* the framework; "external" evaluation compares the theory to other frameworks. We take these up in turn.

5.1 Choosing the correct account within the framework

Within the PA framework, a frequent question is which component is responsible for the observed facts: Is this a matter of syntax, of semantics, of their interface, or of some combination? Here are three cases.

First consider the principle that a German determiner agrees in gender and number with its head noun, regardless of whether the gender of the noun is motivated by semantics. We conclude that gender is a morphosyntactic feature, and that gender agreement is firmly lodged in the well-formedness principles for syntax.

For a second case, consider aspectual coercion (Verkuyl, 1972; Talmy, 1978; Dowty, 1979; Platzack, 1979; Hinrichs, 1985; Jackendoff, 1991, 1997). (34a) describes multiple jumps; but (34b), with a different preposition, describes a single jump, and (34c), with a different verb, describes a single act of sleeping.

(34) a. Sam jumped until the bell rang.
b. Sam jumped before the bell rang.
c. Sam slept until the bell rang.

The multiple event interpretation emerges when (a) the verb denotes a completed event and (b) the time expression sets a boundary on an otherwise unbounded event. Multiplicity could potentially be encoded syntactically, using an invisible formative whose presence depends on the right combination of features in the verb and preposition. However, the PA framework offers a simpler solution: the sentences in (34) have identical syntax, with no hidden formatives or features. The aspectual distinctions among them have to be present in semantic structure in any event, because they are part of the interpretation. Thus the semantics can account directly for the sense of multiplicity: when a temporal boundary on a process (denoted by the preposition *until* but not *before*) is imposed on a point-action (denoted by the verb *jump* but not *sleep*), it thereby coerces the action into a sequence of actions. Hence in this case, the phenomenon is localized in semantics.

Finally, consider the construction illustrated in *that travesty of a theory* (8b). In principle, a syntactic movement could derive it from something like *that theory, which is a travesty*. However, such a movement would be unprecedented, raising a noun out of an appositive while demoting the underlying syntactic head. Moreover, the construction has characteristic semantics: the head noun is understood as evaluating the subordinate noun. (35a) is unacceptable, even though its putative source (35b) is fine, because *sailor* cannot be understood as an evaluative term. On the other hand, (35c) is understood as criticizing the violinist's performance – coercing *butcher* into an evaluation.

(35) a. * that sailor of a butcher
b. that butcher, who is a sailor
c. that butcher of a violinist

This sensitivity to meaning requires involvement of semantic structure. The construction therefore must be an interface principle: To express an evaluation of an individual, one may use the canonical structure N_1 *of a* N_2, in which the evaluation is linked to N_1 and the individual being evaluated is linked to N_2, reversing the usual linkage between syntax and semantics.

These three cases illustrate the sorts of decisions faced within the Parallel Architecture, and how they come to be settled in favor of syntax, semantics, and/or the interface.

5.2 External evaluation and summary: Comparing the Parallel Architecture to alternative approaches

For external evaluation, we first mention two common objections to the PA. First, if syntactic rules are stated declaratively, as in the PA (and other "constraint-based" theories), then the procedural notion of movement has to be expunged from syntax. So it is crucial to figure out how a declarative theory deals with passive, subject-auxiliary inversion, wh-question formation, and the like, which have always seemed very natural under a movement hypothesis. HPSG, LFG, Simpler Syntax, and other frameworks work this out, with considerable success. (Some of the arguments are addressed in Culicover and Jackendoff, this volume.)

Second, it has been argued that the PA is not sufficiently constrained (Marantz, 2005; Phillips and Lau, 2004), because it involves three "generative engines" instead of syntax alone. We find this criticism misguided. Every theory of language must account not only for syntactic well-formedness but also phonological and semantic well-formedness – counterparts of the PA's phonological and semantic "generative engines." And every theory must account for the interfaces among these levels, in order to connect sound to meaning.

In fact, PA syntax is *highly* constrained: it lacks movement and cycles/phases, minimizes phonologically null syntactic heads, and makes no distinction of components between lexicon and rules. The PA's interfaces are more flexible than in other approaches, partially compensating for the loss of syntactic power. However, the PA thereby gains coverage of a plethora of noncanonical phenomena such as those in Section 3.2. Many of these have never (to our knowledge) been addressed in traditional generative grammar and its direct descendants, being regarded as irrelevant exceptions or "periphery" (Chomsky, 1981). Still, if anything, the hard question is whether the PA's relatively lean set of formal devices is powerful *enough* to account for the fullest range of linguistic phenomena.

Perhaps the most important innovation of the PA (shared with Cognitive Grammar and Construction Grammar) is to eliminate the distinction between lexicon and grammar, couching both as pieces of linguistic structure with some combination of phonology, syntax, and/or semantics, and with some combination of constants and variables. In the place of this distinction is a greatly enriched lexicon, in many respects like that of Construction Grammar.

The PA also eliminates the distinction between "grammatical" rules and "lexical" rules, which many approaches treat as distinct grammatical modules. In the PA, both are schemas. The PA replaces the distinction between them with a distinction between two uses of schemas. *All* schemas function relationally, supporting or motivating more highly specified lexical items; this use corresponds to "lexical rules." *Some* schemas also function generatively, unifying with other lexical items to create novel structures; this corresponds to productive "rules of grammar." This duality of function is to our knowledge not central to other frameworks.

The PA takes significant steps toward integrating semantics, syntax, morphology, and phonology into a coherent whole, linking these independent sources of combinatorial structure by interface principles. This overall organization arguably obtains in other faculties of mind as well, and permit the language faculty to interact with other faculties such as vision. This feature of the PA is to our knowledge unique.

An important feature of the PA is that it can be integrated smoothly into accounts of language processing and language acquisition, bridging the competence-performance divide (Jackendoff, 2002, 2007b; Jackendoff and Audring, 2016, forthcoming). This compatibility arises because schemas are lexical items, right alongside words. In the course of parsing a heard sentence, all relevant lexical items are activated – both words and schemas; and candidate parses are constructed by Unification. It is not necessary to go to another "place" to apply rules. Similarly, acquiring a rule amounts to adding a new schema to the lexicon. This schema shares the structure of the items that motivate it, but it contains variables where these items differ. There is no need to construct something in an entirely different "rule" format that goes in a different "place." More generally, PA-style analyses have provided a basis for experimental work on sentence comprehension, e. g. Piñango et al. (1999), Paczynski et al. (2014), Wittenberg et al. (2014), Wittenberg (2016).

The PA also responds to goal (4) in Section 1: determining which aspects of the language faculty are domain-specific and which are domain-general. We speculate that Unification is a domain-general cognitive process for creating combinatorial structures (for discussion, see Jackendoff, 2011), and we suggest that the notion of stored schemas with variables appears in a wide range of cognitive domains (Jackendoff and Audring, forthcoming). What is likely to be specific to language is the repertoire of phonological and syntactic elements out of which words and schemas are built. (In a sense this concurs with Bod's (2006) proposal of "Universal Representation" instead of "Universal Grammar.")

To sum up: While we recognize the danger of viewing the world in terms of one's own theory, we believe that the goals set out in Section 1 are appropriate ambitions for linguistic theory. We further believe that, more than other frameworks we are acquainted with, the Parallel Architecture offers an avenue towards reaching those goals.

References

Aronoff, Mark. 1976. *Word Formation in Generative Grammar*. Cambridge, MA: MIT Press.

Barlow, Michael & Suzanne Kemmer. 1994. A schema-based approach to grammatical description. In Susan D. Lima, Roberta Corrigan & Gregory K. Iverson (eds.), *The Reality of Linguistic Rules*, 19–42. Amsterdam/Philadelphia: Benjamins.

Berwick, Robert C. & Noam Chomsky. 2016. *Why only us?* Cambridge, MA: MIT Press.

Bloomfield, Leonard. 1933. *Language*. New York: Holt.

Bod, Rens. 2006. Exemplar-based syntax: How to get productivity from examples. *The Linguistic Review* 23. 291–320.
Booij, Geert. 2010. *Construction Morphology*. Oxford: Oxford University Press.
Bresnan, Joan (ed.). 1982. *The Mental Representation of Grammatical Relations*. Cambridge, MA: MIT Press.
Bresnan, Joan. 2001. *Lexical Functional Syntax*. Oxford: Blackwell.
Broccias, Cristiano. *This volume*. Cognitive Grammar.
Bybee, Joan L. 2010. *Language, Usage, and Cognition*. Cambridge: Cambridge University Press.
Chaves, Rui. *This volume*. Construction Grammar.
Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
Chomsky, Noam. 1981. *Lectures on Government and Binding*. Dordrecht: Foris.
Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.
Cohn, Neil. 2013. *The Visual Language of Comics*. New York: Bloomsbury.
Corrigan, Roberta, Edith A. Moravcsik, Hamid Ouali & Kathleen M. Wheatley (eds.). 2009. *Formulaic Language, vols. 1 and 2*. Amsterdam/Philadelphia: John Benjamins.
Croft, William. 2001. *Radical Construction Grammar*. Oxford: Oxford University Press.
Culicover, Peter W. 1999. *Syntactic Nuts: Hard Cases, Syntactic Theory, and Language Acquisition*. Oxford: Oxford University Press.
Culicover, Peter W. & Ray Jackendoff. 2005. *Simpler Syntax*. Oxford: Oxford University Press.
Culicover, Peter W., Ray Jackendoff & Jenny Audring. 2017. Multiword constructions in the grammar. *Topics in Cognitive Science* 9. 552–568. doi:10.1111/tops.12255.
Culicover, Peter W. & Andrzej Nowak. 2003. *Dynamical Grammar*. Oxford: Oxford University Press.
Dalrymple, Mary & Jamie Findlay. *This volume*. Lexical Functional Grammar.
Dowty, David R. 1979. *Word Meaning and Montague Grammar*. Dordrecht: Reidel.
Embick, David & Rolf Noyer. 2007. Distributed Morphology and the Syntax/Morphology interface. In Gillian Ramchand & Charles Reiss (eds.), *The Oxford Handbook of Linguistic Interfaces*, 289–324. Oxford: Oxford University Press.
Ergin, Rabia. 2017. Central Taurus Sign Language: A Unique Window on Language Creation. Ph.D. dissertation, Tufts University.
Everett, Daniel L. 1986. Pirahã. In Desmond D. Derbyshire & Geoffrey K. Pullum (eds.), *Handbook of Amazonian Languages, Vol. 1*, 200–325. Berlin: Mouton de Gruyter.
Everett, Daniel L. 2005. Cultural constraints on gammar and cognition in Pirahã: Another look at the design features of human language. *Current Anthropology* 46. 621–646.
Fillmore, Charles J., Christopher R. Johnson & Miriam R. L. Petruck. 2003. Background to Framenet. *Journal of Lexicography* 16. 235–250.
Fitch, W. Tecumseh. 2010. *The Evolution of Language*. Cambridge: Cambridge University Press.
Fodor, Janet D. 1998. Learning to parse. In David Swinney (ed.), *Anniversary Issue of Journal of Psycholinguistic Research 27*, 285–318.
Frank, Robert & Anthony Kroch. 1995. Generalized transformations and the theory of grammar. *Studia Linguistica* 49. 103–151.
Gil, David. 2009. How much grammar does it take to sail a boat? In Geoffrey Sampson, David Gil & Peter Trudgill (eds.), *Language Complexity as an Evolving Variable*, 19–33. Oxford: Oxford University Press.
Goffman, Erving. 1974. *Frame Analysis: An essay on the organization of experience*. Cambridge, MA: Harvard University Press.
Goldberg, Adele E. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.
Goldberg, Adele E. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.

Goldsmith, John. 1979. *Autosegmental Phonology*. New York: Garland.
Heim, Irena & Angelika Kratzer. 1998. *Semantics and Generative Grammar*. Oxford: Blackwell.
Hinrichs, Erhard W. 1985. A Compositional Semantics for Aktionsarten and NP Reference in English. Ph.D. dissertation, Department of Linguistics, Ohio State University.
Hoffman, Thomas & Graeme Trousdale (eds.). 2013. *The Oxford Handbook of Construction Grammar*. Oxford: Oxford University Press.
Jackendoff, Ray. 1975. Morphological and semantic regularities in the lexicon. *Language* 51. 639–671.
Jackendoff, Ray. 1977. *X-Bar Syntax*. Cambridge, MA: MIT Press.
Jackendoff, Ray. 1983. *Semantics and Cognition*. Cambridge, MA: MIT Press.
Jackendoff, Ray. 1987. On Beyond Zebra: The Relation of Linguistic and Visual Information. *Cognition* 26. 89–114.
Jackendoff, Ray. 1990. *Semantic Structures*. Cambridge, MA: MIT Press.
Jackendoff, Ray. 1991. Parts and Boundaries. *Cognition* 4. 9–45.
Jackendoff, Ray. 1996. The architecture of the linguistic-spatial interface. In Paul Bloom, Mary A. Peterson, Lynn Nadel & Merrill F. Garrett (eds.), *Language and Space*, 1–30. Cambridge, MA: MIT Press.
Jackendoff, Ray. 1997. *The Architecture of the Language Faculty*. Cambridge, MA: MIT Press.
Jackendoff, Ray. 2002. *Foundations of Language*. Oxford: Oxford University Press.
Jackendoff, Ray. 2007a. *Language, Consciousness, Culture*. Cambridge, MA: MIT Press.
Jackendoff, Ray. 2011. What is the human language faculty? Two views. *Language* 87. 586–624.
Jackendoff, Ray & Jenny Audring. 2016. Morphological schemas: Theoretical and psycholinguistic issues. *The Mental Lexicon* 11. 467–493.
Jackendoff, Ray & Jenny Audring. Forthcoming. *The Texture of the Lexicon: Relational Morphology and the Parallel Architecture*. Oxford: Oxford University Press.
Jackendoff, Ray & Fred Lerdahl. 2006. The capacity for music: What's special about it? (with Fred Lerdahl). *Cognition* 100. 33–72.
Jackendoff, Ray & Eva Wittenberg. 2014. What you can say without syntax: A hierarchy of grammatical complexity. In F. J. Newmeyer & L. Preston (eds.), *Measuring Grammatical Complexity*, 65–82. Oxford: Oxford University Press.
Joshi, Aravind, K. 1987. An introduction to Tree-Adjoining Grammars. In Alexis Manaster-Ramer (ed.), *Mathematics of Language*, 87–114. Amsterdam: Benjamins.
Klima, Edward S. 1965. Studies in Diachronic Syntax. Ph.D. dissertation, Harvard University.
Lakoff, George. 1987. *Women, Fire, and Dangerous Things*. Chicago: University of Chicago Press.
Lamb, Sydney. 1966. *Outline of Stratificational Grammar*. Washington: Georgetown University Press.
Landau, Barbara & Ray Jackendoff. 1993. What" and "where" in spatial language and spatial cognition. *Behavioral and Brain Sciences* 16. 217–238.
Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar, vol. 1*, Stanford: Stanford University Press.
Legendre, Géraldine. *This volume*. Optimality-theoretic Grammar.
Lerdahl, Fred & Ray Jackendoff. 1983. *A Generative Theory of Tonal Music*. Cambridge, MA: MIT Press.
Liberman, Mark & Alan Prince. 1977. On stress and linguistic rhythm. *Linguistic Inquiry* 8. 249–336.
Marantz, Alec. 1997. No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. *Philadelphia: University of Pennsylvania Working Papers in Linguistics* 4(2). 201–225.
Marantz, Alec. 2005. Generative linguistics within the cognitive neuroscience of language. *The Linguistic Review* 22. 429–446.

Minsky, Marvin. 1975. A framework for representing knowledge. In Patrick Winston (ed.), *The Psychology of Computer Vision*, 211–277. New York: McGraw-Hill.

Müller, Stefan & Antonio Machicao y Priemer. *This volume*. Head-Driven Phrase Structure Grammar.

Paczynski, Martin, Ray Jackendoff & Gina Kuperberg. 2014. When events change their nature: The neurocognitive mechanisms underlying aspectual coercion. *Journal of Cognitive Neuroscience* 26. 1905–1917.

Phillips, Colin & Ellen Lau. 2004. Foundational issues (Review article on Jackendoff 2002). *Journal of Linguistics* 40. 1–21.

Piñango, Maria Mercedes, Edgar Zurif & Ray Jackendoff. 1999. Real-Time Processing Implications of Enriched Composition at the Syntax-Semantics Interface. *Journal of Psycholinguistic Research* 28. 395–414.

Platzack, Christer. 1979. *The Semantic Interpretation of Aspect and Aktionsarten*. Dordrecht: Foris.

Pollard, Carl & Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. Chicago: University of Chicago Press.

Rumelhart, David. 1980. Schemata: The building blocks of cognition. In R. Spiro, B. Bruce & W. Brewer (eds.), *Theoretical Issues in Reading Comprehension*, 33–58. Hillsdale, NJ: Erlbaum.

Rumelhart, David & James McClelland. 1986. On learning the past tense of English verbs. In James McClelland, David Rumelhart & the PDP Research Groups, *Parallel Distributed Processing, Vol. ii*, 216–271. Cambridge, MA: MIT Press.

Sadock, Jerrold. 1991. *Autolexical Syntax*. Chicago: University of Chicago Press.

Sandler, Wendy, Irit Meir, Carol Padden & Mark Aronoff. 2005. The emergence of grammar: Systematic structure in a new language. *Proceedings of the National Academy of Sciences* 102(7). 2661–2665.

Schank, Roger. 1973. Identification of Conceptualizations Underlying Natural Language. In Roger Schank & Kenneth Colby (eds.), *Computer Models of Thought and Language*, 187–248. San Francisco, CA: Freeman.

Selkirk, Elizabeth O. 1984. *Phonology and Syntax: The Relation between Sound and Structure*. Cambridge, MA: MIT Press.

Shieber, Stuart. 1986. *An Introduction to Unification-based Approaches to Grammar*. Stanford, CA: CSLI Publications.

Smolensky, Paul & Géraldine Legendre. 2006. *The Harmonic Mind*. Cambridge, MA: MIT Press.

Talmy, Leonard. 1978. The relation of grammar to cognition: A synopsis. In David Waltz (ed.), *Theoretical Issues in Natural Language Processing 2*, 14–24. New York: Association for Computing Machinery. Revised version in Leonard Talmy. 2000. *Toward a Cognitive Semantics, Volume 1: Concept Structuring Systems. & Volume 2: Typology and Process in Concept Structuring*. Cambridge: MIT Press.

Tomasello, Michael. 2003. *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge: Harvard University Press.

Ullman, Michael T. 2015. The declarative/procedural model: A neurobiologically motivated theory of first and second language. In B. VanPatten & J. Williams (eds.), *Theories in second language acquisition: An introduction*, 135–158. 2d edn. New York: Routledge.

Van Valin, Robert D. & Randy LaPolla. 1997. *Syntax: Structure, Meaning and Function*. Cambridge: Cambridge University Press.

Verkuyl, Henk. 1972. *On the Compositional Nature of the Aspects*. Dordrecht: Reidel.

Wittenberg, Eva. 2016. *With light verb constructions from syntax to concepts*. (Potsdam Cognitive Science Series 7). Potsdam: Universitätsverlag Potsdam.

Wittenberg, Eva, Martin Paczynski, Heike Wiese, Ray Jackendoff & Gina Kuperberg. 2014. The difference between ”giving a rose” and ”giving a kiss”: Sustained neural activity to the light verb construction. *Journal of Memory and Language* 73C. 31–42.

Ritva Laury and Tsuyoshi Ono

9 Usage-based Grammar

Abstract: This article introduces two major approaches to usage-based study of syntax, Emergent Grammar and Interactional Linguistics. Grammarians studying human languages from these two approaches insist on basing their analyses on data from actual language use, especially everyday conversation in a range of languages. Grammar is viewed as emerging from language use in context, and thus grammatical structure is seen as provisional, negotiable, and ever changing. Linguistic units and categories need to be based on what is found in actual use. Traditional notions are not accepted a priori, unless it is shown that speakers actually orient to them in their everyday use. Findings are stated in terms of both form and function. Since the availability of videotaped data, there has been increasing attention paid to embodied behavior as a component of linguistic communication. Frequently used structures are seen as more basic than rarely used ones, and for this reason, quantitative approaches are common. Building carefully designed, balanced corpora of everyday speech in a range of contexts for multiple languages will be a next major step which would make this usage-based endeavor to human language a viable option.

1 Introduction

This chapter represents Usage-based grammar as it is conceived of in discourse-functional and functional-typological approaches to human languages, such as what came to be called West Coast Functionalism, and more recently in approaches concerned with the use of grammar in interaction. In this article, we focus in particular on two approaches, namely Emergent Grammar (Hopper, 1987, 2011) and Interactional Linguistics (Selting and Couper-Kuhlen, 2001; Couper-Kuhlen and Selting, 2018). These approaches had their beginnings in the 1970s and 80s, when linguists such as Li (1976), Givón (1979), Chafe (1980), Du Bois (1980), Hopper and Thompson (1980, 1984), and Bybee (1985) turned to functional aspects of language, first using constructed examples, and then discourse data, starting with written data and elicited narratives (e. g. Chafe, 1980, Hopper and Thompson, 1980 and 1984, Du Bois, 1987), in search of how language was used in and shaped by actual use. Emergent Grammar originated in Hopper's (1987) critique of the then and perhaps still dominant view of

Note: We would like to thank the editors of the current volume for their timely response and critical yet very constructive comments which have turned our paper into a much more readable piece. In some cases, this has led to clarification of our position on critical issues that we are dealing with. In writing this paper, we have also benefited immensely from our regular discussion with Marja-Liisa Helasvuo, Toshi Nakayama, Ryoko Suzuki and Sandy Thompson.

https://doi.org/10.1515/9783110540253-009

grammar as a stable, autonomous system which exists prior to use, and the promotion of an approach which would see grammar as emerging from language use in context and grammatical structure as "always provisional, always negotiable, and in fact epiphenomenal". These ideas on the nature of grammar, radical at the time, came to have a strong influence on linguistic approaches to everyday talk-in-interaction. Even earlier, sociologists had become interested in the structure of conversation (Sacks et al., 1974, Sacks, 1992), which later resulted in the development of Conversation Analysis (Schegloff, 2007; Sidnell, 2009; Sidnell and Stivers, 2012). By the late 1980s and early 1990s, functionally oriented linguists also began studying grammar in conversation (e. g. Fox, 1987; Auer, 1992; Ford, 1993; Ono and Thompson, 1995; Ochs et al., 1996). Out of these beginnings developed the approach now called Interactional Linguistics (Selting and Couper-Kuhlen, 2000, 2001; Couper-Kuhlen and Selting, 2018), which combines insights from the above-mentioned Conversation Analysis and functionally oriented linguistics.

2 Data

As the term usage-based grammar indicates, linguists working in this paradigm generally base their analyses on naturally occurring, corpus-based data, although experimental, elicited and even introspective data are also used.[1] The focus on corpus data, especially in Emergent Grammar, arises from the basic tenet shared by practitioners of this approach, namely that grammar is not only a repository of knowledge but rather a dynamic product of language use in context. The form that utterances take in interaction is strongly influenced by what has happened just prior ('positionally sensitive grammar', Schegloff, 1996), and, more broadly, what is going on in the interaction, as utterances are shaped online in response to factors constantly emerging as talk goes on ('online syntax', Auer, 2009). That is, grammar does not exist *a priori*, as a stable repository of items and rules in place in the minds of its users before being put to use, but rather it emerges, is created and constantly modified by the use of language in context (Hopper, 1987, 2011).

In many usage-based approaches, and especially in Interactional Linguistics, there is an emphasis on the role of ordinary everyday spoken language as the primary

1 One approach where introspective data are used is Langacker's Cognitive Grammar (Langacker, 1987, 1991); the creator of the approach, Ronald Langacker, uses constructed data in his English-based research. However, many cognitive grammarians also use both spoken and written corpus data (see Langacker, 1999; Barlow and Kemmer, 2000; Broccias, this volume). These types of data are also examined in other usage-based approaches to grammar, such as experimental and quantitative corpus linguistics, not focused on in this article (on a useful introduction, see Gries, 2009). On the use of experimental data in Conversation Analysis, see e. g. Stevanovic, 2016.

and most basic form of language (see, e. g. Schegloff, 1996). This is because it is acquired first, exists in a language community before the written form of language is developed, and is still the only form used by speakers of many, probably most, human languages. While written corpora are also studied, especially by grammarians working on diachronic issues from a usage-based perspective, it has been observed that our understanding of human language in general has been skewed by a traditional focus on written language, what Linell (2005) calls 'written language bias'. In this article we will focus on those approaches to usage-based grammar which take spoken language as their primary focus, as that seems to be a common trend among many usage-based grammarians examining discourse data in recent years.[2]

In the study of language in interaction, there is an increasing emphasis on videotaped data, as it has become clear that much of what goes on in face-to-face language use involves embodied action. For example, knowing who is targeted as the recipient of a particular utterance and turn is difficult or even impossible to determine without visual access (e. g. Lerner, 1996), and it is known that speakers design their turns for their particular addressee(s) ("recipient design", Schegloff, 2006, 89). However, while significant insights into language use have been obtained from the study of embodiment (see especially the studies by Charles and Marjorie Harness Goodwin (e. g. Goodwin, 1986, 2003; Goodwin and Goodwin, 1986) and Lorenza Mondada (e. g. Mondada, 2007); see also Haddington (e. g. Haddington, 2012)), our knowledge of what all is involved in it, and how it is integrated with verbal aspects of interaction is still limited.

A complicating factor in the study of the role of embodiment in language is that since videotaped data has been more widely used by interactional linguists only in the last ten years or so, the availability of videotaped data is still also very limited even for heavily studied languages. Furthermore, the study of embodied action is extremely time-consuming; for this reason, it may not be compatible with quantitative approaches which are often required for significant generalization. Issues that arise regarding the protection of the identity of the speakers and other sensitive personal data in the collection, processing, storage and publication of corpus data are becoming increasingly prominent; they are even more demanding with video data than audio data.

When choosing among alternative hypotheses, corpus data count more than other kinds of data. Ultimately, researchers are interested in knowing what occurs,

2 It seems at the very least premature to assume that there is one grammar which would account for both written and spoken language (see the Multiple Grammars Hypothesis, Iwasaki, 2015). Crucially, our understanding of how spoken language functions is still incomplete. Since speech is the most fundamental form of language, that should be the primary target of research if one aims at reaching an understanding of how human language in general works. There is also the question of forms of language which seem to utilize features of both speech and writing, such as many online varieties, which may require a separate theory.

what does not occur, and what occurs most often in particular contexts. This shows what participants orient to in carrying out actions with others (on participant orientation, see Thompson et al., 2015; Couper-Kuhlen and Selting, 2018; Laury et al., to appear). Furthermore, usage-based linguistics has long emphasized the role of frequency in grammar (see, e. g., Bybee and Hopper, 2001). Frequently occurring structures are more basic than rarely used ones and have a role to play in language change, contributing to the spreading of systematicity in language, what we think of as grammar. As Du Bois has put it, "grammars code best what speakers do most" (1985, 363). Frequently used patterns may also create economically motivated asymmetries in grammar (Haiman, 1983; Haspelmath, 2008). Frequently used expressions are known to receive less coding than expressions that are rarely used. Similarly, predictable entities receive less coding than unpredictable entities. Consider, for example, reflexives. Haiman (1983, 803) argues that in *Max washed* the reflexive is not mentioned because the object is predictable, whereas in *Max kicked himself* the reflexive is mentioned, since kicking oneself is an unexpected activity. Besides economy, such patterns have also been shown to be motivated by interactional and sequential factors. Frequently occurring actions tend to be associated with certain types of grammatical constructions, which in turn occur in certain sequential positions. Telephone call opening routines are a case in point; the order of what is said and how it is responded to has been found to be highly routinized (on landline phone calls Schegloff, 1986 on English; Hakulinen, 1993; on mobile phone calls, see Arminen and Leinonen, 2006).

3 Goals

Emergent Grammar and Interactional Linguistics are presented as alternatives to traditional approaches whose focus is constructed data, and with the exclusive focus on naturally occurring everyday conversation, they may be understood to extend other usage-based approaches such as Construction Grammar (see Chaves, this volume) and Cognitive Grammar (see Broccias, this volume; Etelämäki and Visapää, 2014). Problems that Emergent Grammar and Interactional Linguistics deal with are rather open-ended at this point in the enterprise. Researchers are particularly interested in exploring fundamental questions such as what language is, what it consists of, and how it works. More generally, we aim to provide a realistic picture of how speakers of human languages actually use language. This requires understanding the nature of language and grammar and determining what categories and units languages consist of, what those categories and units are like, and how languages develop and change over time. In practice, we typically give structural and functional descriptions, show how they are related to each other, and propose hypotheses about why grammars of languages are the way they are.

Minimally, but crucially, we have to be able to account for the form and function of everyday spoken language, the only type of language shared by the majority of world languages where we know for sure that the linguistic ability is being exploited in its production.[3] In fact, initial investigations of syntax in everyday speech, though observational and a little crude they may have been, led to a whole series of discoveries which were completely new in the context of traditional linguistics, which was built on the study of constructed examples. These discoveries include the lack of the 'sentence' in spoken English (Miller, 1995; Miller and Weinert, 1998), the use of 'neverending sentences' in German (Auer, 1992), "subordinate" clauses functioning like main clauses (Nakayama and Ichihashi-Nakayama, 1997); now called 'insubordination' (Evans, 2007). Another major finding is that cross-linguistically, syntactic subjects overwhelmingly tend to encode "given" or "old" information in the form of affixes and pronouns (Chafe, 1987; Du Bois, 1987). As an example of early findings based on everyday conversation, we give below a simplified transcript and translation of a German "neverending sentence" from Auer (1992, 43–44).[4]

01M: des auf der. einen Seite is also Aussen sonne Hülse.
that is on the one side is kind of a sheath on the outside,

02F: =j[a
yeah

03M: [rund,
round

04 [1.0; gulps]

05 und in der Mitte is bei dem ein n Docht,
and in the middle this one has a wick,

06 n massiver Do[cht, n d'ünner,
a solid wick, a thin one,

07F: [m

08M: un auf der andern Seite vom selben Kabel
and on the other side of the same wire

3 It seems fair to say that other types of data such as constructed examples and experimental data are produced using some type of linguistic ability, but that ability cannot be assumed to be the same as the ability which produces language in naturally occurring everyday speech. This is because these types of data either lack or manipulate discourse context which is exactly what is not found in naturally occurring speech. For this reason, practitioners of Emergent Grammar and Interactional Linguistics tend to consider those kinds of data less central to their work or even unworthy of consideration.
4 Note that in transcripts of spoken language, punctuation marks such as the period, the comma, the semicolon, the question mark and the exclamation mark are used for indicating prosodic features, not syntactic boundaries or sentence function, as in writing.

09 [1.0, gulps]

10 is n Docht der hohl is.
is a wick which is hollow.

11 (1.0)

12 der m bissl dicker is.
which is a little bit thicker.

13 des sin/ die des sin die Kabel.
these are the wires.

14 an besten suchs ma nach som mittelbrauen Kabel wo
the best thing to do is to look for a brownish wire with

15 vorne und hinten so(n) runder Stecker dran [is. das is
in front and behind is like a round plug. that is

16F: [ja. Also
yeah. so

17M denn genau s richtige
exactly the right one

18F: nich verschiedene kleine – äh Pinne komm da raus
it not different little – uh pins sticking out there

19 sondern *ei*[n – dicker Docht,
but one – thick wick,

20M: [genau
exactly

((*etc*.))

Auer shows (43–46) that parsing such an indeterminate stretch of talk into units such as clauses and sentences runs into a number of problems. For example, unfinished clausal constructions such as the first part of line 1 are problematic in this respect. The construction *das ist auf der einen Seite* lacks a predicate, but the speaker then uses the local adverbial *auf der einen Seite* to build another clause in an apokoinu construction.[5] Similar problems arise from the continuation of what seem to be already completed sentences. In line 10, the clausal unit is brought to a completion syntactically and prosodically, but in line 12, after a pause, the speaker begins a relative clause.

5 An apokoinu construction involves the use of a single linguistic element in two constructions; in this case, *auf der einen Seite* is first used as an adverbial in a clause beginning with *das ist*, but the speaker abandons that plan and instead builds another clause beginning with the adverbial. The adverbial phrase thus has membership in two clausal units.

The question then becomes, whether there is a sentence boundary at the end of line 10 or not, and just where, if anywhere, there is a sentence boundary in line 1. Auer argues that a more reasonable direction to go would be prosodic segmentation, and even more, analyzing spoken conversation in terms of turntaking. Stretches of language, he suggests, constitute turns at talk rather than grammatical units such as sentences. Auer allows that syntactic, as well as prosodic and semantic analysis of such units is also a worthwhile and important pursuit, but he questions the usefulness of the concept of 'sentence' in such analysis (47). As we will show, much subsequent work has followed along these same lines.

In general, practitioners of usage-based approaches do not provide coherent models as the general sentiment is that it is premature to adopt that type of approach. This is because we started examining everyday speech only recently, as noted above. However, that said, there are certainly different attitudes about modeling even among usage-based linguists who examine discourse data, as researchers have different training and orientations. It should be pointed out, however, that having to explain why one does not provide an explicit model shows a bias in our field where many researchers share the view that a scientific approach to language study requires the construction of formal models. We might note that there are linguists who build models based on data they have themselves constructed. The results obtained from such an exercise may have little to do with what actual speakers do in actual contexts of use (see Laury and Ono, 2005).

The present authors' view is that we should first observe what interactants do in order to avoid as much bias as possible which can be created by what we think happens, or what the theory leads us to assume. Obviously this is easier said than done because one's description is always influenced or informed by assumptions and theories one may have. Still the preference toward avoiding preconceptions is an attitude commonly shared by practitioners of Emergent Grammar and Interactional Linguistics. Even given the paucity of appropriate data in nearly all languages of the world, it is becoming abundantly clear that non-verbal behaviors are as critical as speech in making interaction possible. Various functions/actions depend on the involvement of both of these two modes and for that reason, separating the two and/or focusing just on speech might not be wise.

Usage-based accounts ultimately aim at both description and explanation as seen even in early work such as Geluykens (1992) where, using English corpus data, he demonstrates that what has been analyzed as a syntactic phenomenon, 'left dislocation' is best characterized as a grammaticized construction from a frequent way of introducing 'topic' into conversation (but see Pekarek Doehler et al. (2015) for a more recent analysis, where left dislocation is seen as a resource for turn-taking and sequence organization). Similarly, several recent studies have shown that prosody and embodied features of language are important in creating meaning in language use. For example, Thompson, Fox and Couper-Kuhlen show that intonational packaging plays a decisive role in responses to informings. (2015, 56–138). With rising intonation, “the recipient is soliciting more work on the informing, making it public that s/he is not quite ready to accept it as fact, and requesting verification or confirmation of it (135)”. That is, even though the linguistic form of a response remains constant, the prosodic form it takes carries additional meaning which other participants respond to. Consider the following excerpt, taken from Thompson et al. (2015, 109), where Michael asks what kind of solution Vivian uses for her contact lenses (lines 1–2). Vivian then gives the requested information (line 3). In response, Michael uses an interrogatively formatted utterance with rising intonation, marked in the transcript with a question mark (line 5), which indicates that Vivian has provided new information, but also requests verification.

```
1   MIC:   what kind of solution
2          yo[u – you: uh: u:se. you use –
3   VIV:     [Bausch and Lomb
4   (0.3)
5   MIC:   oh do you? Is tha:t what you us[e?
6   VIV:                                  [°yeh°
```

Here, Michael’s interrogative with rising intonation *oh do you?* is followed by another request for verification, which Vivian then provides. Thompson et al. show that otherwise similarly formatted requests with falling intonation do not ask for, and do not get, additional verification (111).

Further, already in early work on videotaped conversation, it became apparent that speakers closely monitor the embodied behavior of their recipients, such as gaze direction, and tailor their language accordingly, even modifying utterances already in progress in order to direct them at particular recipients (Ford and Fox, 1996).

Most of these descriptions are not formalized, with notable exceptions (see, e. g., Lindström, 2014, Auer, 2014). A general feeling is that it is premature to go into formalization due to the lack of basic description, as we only recently began to examine naturally occurring everyday conversation as our primary data. Yet, preliminary studies clearly show that we need to develop radically new ways of conceptualizing and examining language. The traditional formalization tools, however, may not be adequate or useful to capture the multi-modality, variation, ongoing change, and other dimensions of these new data. Explicit descriptions appear to serve us better at this juncture than premature formalization.[6]

6 One of the editors remarked that it is not the maturity of the paradigm but rather the complexity of the phenomena requiring formalization which makes the current tools available for formalization inadequate, and moreover, that formalization is not inherently a goal usage-based approaches should strive for. This is precisely how we also view the issue. By ‘premature’, we simply mean, ‘too early’.

4 Tools

4.1 Descriptive tools

Most commonly, in usage-based approaches to grammar, descriptive statements are given both in terms of form and function, often along with frequency and percentage information, sometimes with statistical figures (e. g., Helasvuo and Kyröläinen, 2016). Structural description can include not only syntax but other dimensions such as prosody and nonverbal movements, especially in recent work (e. g. Barth-Weingarten, 2016; Rauniomaa and Keisanen, 2012). Functional description includes semantic, discourse, social and interactional dimensions. Form and function are discussed together because many practitioners of Emergent Grammar and Interactional Linguistics share a basic attitude against compartmentalized views of language (i. e., modularity), in contrast to the traditional assumption that language consists of discrete components which can be studied one at a time. It seems rather obvious and only natural that form and function are found together in actual usage, and in fact recent studies, especially those taking multimodal approaches, have been uncovering intricate mechanisms which highlight intimate connections between form and function (e. g., Goodwin, 2007; Mondada, 2009; Kärkkäinen and Keisanen, 2012; Haddington et al., 2013). For this reason, a viable alternative view might be that these two areas are represented together in the grammar and thus they are studied and described together in usage-based approaches to grammar.

Equally importantly, as language is part of human behavior, and thus crucially associated with constant change, one naturally comes across multiple facets of ongoing change in actual speech data. This is why statements about variation and change are often part of the statements about the form and function in Emergent Grammar and Interactional Linguistics.

Since we are interested in what speakers regularly do, we often seek quantitative support, mainly in terms of frequency of occurrence and percentage. That is, our investigation begins with what speakers produce regularly, not with some forms and functions which might be theoretically important but rarely found in actual talk; forms and functions, unless regularly used, are not important to the line of research we are highlighting here.

Quantitative information such as frequency and percentage is used mostly for descriptive purposes in our research. That is, such information is typically employed as a descriptive measure of how frequent some forms and/or functions are. It is also often employed to show how commonly some forms and functions occur together, and sometimes to show which among competing form-function pairings have better fits which gives an indication as to which pairings might be more recognized and more actively employed by speakers (Thompson et al., 2015). It is also employed to show the degree and nature of on-going language change (Ono et al., 2012b). For example, it seems that while verbs of cognition, like other verbs, have full paradigms in a number of languages, only certain forms are frequently used and consequently become fixed expressions or 'epistemic fragments' (Thompson, 2002) with distinct functions such as stance taking (see, for example, Kärkkäinen, 2003 on the English *I think*; Keevallik, 2003, 2016 on Estonian *mai tea* 'I don't know'; Endo, 2013 on the Mandarin *wo juede* 'I feel/think'; Deppermann and Reineke, 2017 on the German *ich dachte* 'I thought').

Use of quantitative information naturally makes one think of the employment of statistical methods as a next logical step. Unfortunately, these methods are of limited value to us since we typically lack carefully designed balanced corpora involving naturally occurring everyday speech, as mentioned above. As noted, for most languages, we only have a small collection of random sets of recordings and their transcripts at best, and for most world languages we have no data consisting of naturally occurring talk. Statistical methods can of course be applied to existing data to obtain statistical figures, but in the absence of adequate, relevant and sufficiently similar data, in the end the results may not tell us much about what actual speakers do in real life. In fact, as more and more researchers are interested in non-verbal aspects of interaction which are, as we have noted, directly tied to or perhaps even inseparable from verbal aspects, the limitations of the existing corpora become even more problematic even for qualitative single case studies, as most of them do not include video. For these reasons we believe that our immediate efforts should be to directed to video recording of naturally occurring interaction and producing high quality transcripts, neither of which is unfortunately a simple task to accomplish. However, these are the primary data to study human language as it is used by real people.[7]

4.2 Components, levels, categories and relations

Our basic attitude concerning notions such as components, levels, and categories of language and the relations between them is empirically oriented. That is, they need to be informed and supported by what speakers do in interaction. Otherwise, they remain assumptions or hypotheses at best. As starting points and/or out of convenience (i. e., just to name components and levels which we question), we refer to traditional components such as phonetics, phonology, morphology, syntax, and semantics, but that does not mean at all that we believe in their reality as traditionally conceived (e. g., as discrete entities) or their reality themselves. In fact, as noted, many (if not all) usage-based grammarians are wary of the modular view of language.

7 In situations where primary data is not easily accessible, linguists with traditional training typically try to find another way of getting to the research topic or question, most typically using constructed and/or experimental data. Unfortunately, being devoid of speech context, these types of data are not useful for our goal of understanding human language. Without empirical evidence, we cannot simply suppose that constructed and elicited language has anything to do with what actual speakers do in real life.

Even a casual glance at actual speech data easily supports an earlier claim by Langacker (1987) that there is no clear division between grammar and lexicon (see Laury and Ono, to appear), which have been portrayed as a prime example of two separate components in language. Further, in closely examining language use, one becomes aware, for instance, that lexicon, semantics and phonology, as well as embodied/nonverbal features of conversational interaction might be better characterized to form an inseparable whole with syntax. In fact, most recent studies in Interactional Linguistics (e. g., Deppermann, 2013; Selting, 2013; Li, 2014; Li and Ono, in press) reveal a tight link between these verbal/nonverbal materials and social actions engaged in by interactants, suggesting that the nature and reality of these components and levels require new description and theorization. For example, Rossi (2015) shows that the use of different linguistic formats of requesting in Italian depends on factors such as the projectability of the requested action, the requestee's visual attention, how the requested action is related to what the requestee is currently doing, and the availability of objects. Similar factors have been shown to be relevant to how requests made by shoppers to employees in Finnish convenience stores are formatted (Sorjonen and Raevaara, 2014). That is, the motivation for the choice of particular verbal formats or constructions is entirely due to nonverbal factors in the environment.

With regard to what might be characterized as lower levels and parts of language, it is safe to say that most of the researchers in Emergent Grammar and Interactional Linguistics operate with the assumption that human language has categories, but some would even challenge that. That is, the type and nature of categories, or even the idea of categorization can be challenged (see Ford et al., 2013 for relevant comments).

Again, traditional labels such as part of speech categories and phrases/clauses are employed as starting points and/or for convenience, but our descriptive attempts and theoretical attitudes have been to see if those categories and units as traditionally conceived are identifiable in actual speech and if they are oriented to by speakers themselves. That is, trying to identify, describe, and establish categories and units in everyday speech (rather than assuming them) is a major component of research.

Prominent examples of such work are the large body of research on fixed expressions that develop from fully fledged main clauses, becoming morphologically and phonetically reduced and no longer functioning as main clauses, but rather as projector constructions that serve to create a slot for further talk of a particular nature (e. g. Hopper, 2001; Hopper and Thompson, 2008; Günthner, 2008; Pekarek Doehler, 2011; Imo, 2011). One such construction is 'the thing is', a phrase found in both English and German. Günthner (2011) shows that the German *die Sache ist* serves to anticipate, and to focus the recipient's attention on, its speaker's core message, which can take various forms. The phrase can be followed by a complement clause (overwhelmingly without the complementizer *dass*), but also by a main clause, or a larger, complex stretch of discourse similar to Auer's (1992) 'neverending' sentences. In such uses, the expression no longer serves as a matrix clause, and in fact, in such uses, what follows overrides the former 'matrix' (see also Thompson, 2002). Such expressions further call into question the category of 'sentence' in spoken language, and the studies add to the body of research concerning the suitability of traditional linguistic categories for the description of ordinary spoken language in a variety of languages.

It should be highlighted that the series of findings concerning the reality of grammatical units and their nature in conversation were made possible with the empirical orientation in usage-based research where research was conducted without assuming traditional categories and units. Similarly, as a first step, usage-based researchers use traditional relations such as sequential ordering, inclusion, modification but they are fully aware that these relations need to be empirically established by actual speech data.

Overall, we tend to be agnostic toward traditional linguistic notions, which makes us more empirically oriented in dealing with various standard syntactic notions such as components, levels, categories, and their relations. These notions were established based on constructed examples of dominant languages such as English and other Indo-European languages, and a large number of recent studies have highlighted and continue to highlight their inadequacy in describing the grammar of not only other (new) languages (a fraction of the thousands of languages we have little information on) but also everyday conversation of more extensively studied languages like English; but it can be questioned whether the traditional linguistic categories are even relevant for ordinary conversation in English (see Fox et al., 2013).

It is becoming increasingly clear that as our primary data, recordings and transcripts of everyday speech especially with video, become more available, most of the traditional syntactic notions require radical rethinking and reformulation by taking into consideration a number of interrelated factors, including the nature of categories of human behavior, the interactivity and temporality of ongoing talk, issues of ongoing constant change, and cross-linguistic differences.

4.3 Crosslinguistic variation, language change, and cognition

Although Emergent Grammar and Interactional Linguistics have traditionally paid particular attention to crosslinguistic similarities and differences found in actual interaction data (e. g., to name only a few, see Fox et al., 1996; Couper-Kuhlen and Ono, 2007; Haakana et al., 2009; Laury and Ono, 2014; Ono and Thompson, 2017; Ono et al., to appear; Ono et al., 2012a; Ehmer and Barth-Weingarten, 2016; Zinken, 2016; Couper-Kuhlen and Selting, 2018; Lindström et al., 2016), claims about language universals are approached very cautiously. Data requirements discussed throughout this article naturally translate to our views on cross-linguistic research in general and universal claims in particular. That is, cross-linguistic comparisons and claims about universals obviously have to be made based on forms and functions actually employed in everyday talk of the languages examined. Cross-linguistically equivalent forms and functions based on constructed data, the kind most commonly used in typology, for instance, do not meet this minimal level of empiricism for comparison, let alone claims for universals. We stress this simply because comparison is not meaningful unless relevant forms and functions actually occur in the languages which are compared. There has been recent discussion within typological linguistics regarding the crosslinguistic applicability of traditional categories (see especially Haspelmath, 2010). There is also much interest within usage-based linguistics in comparing the inventories of linguistic resources available in different languages for accomplishing similar actions, and how the differences in resources might influence how speakers orient to features of the interactional context. Zinken (2016), for example, shows that due to grammatical structures that exist in Polish but not in English, requests for carrying out small tasks in a family are sensitive to features of context in language-specific ways. In this case, the observable commitment of the other person to a shared task systematically enters into the way a Polish request is built.

The current reality is, however, that only a very small number of (mostly European) languages has been studied well enough to be considered for comparison purposes; most of the world's languages are not even known yet, and we lack everyday interaction data from most languages. Much of our knowledge of less well known (or even well known) languages is based on constructed examples provided by the researchers and/or found in reference grammars.

Constructed data and examples taken from reference grammars are known to be particularly problematic because of the 'written language bias' (Linell, 2005) discussed earlier. They often do not represent the type of language found in speech. These examples are also problematic because they are typically modeled after grammatical categories and units of English and other Indo-European languages. That is, most past and present approaches to grammar are overwhelmingly based on these dominant and colonial languages (for approaches critical of traditional categories, see Haspelmath, 2010; Szczepek Reed and Raymond, 2013), and for that very reason, standard grammatical categories and units presented in the literature are very much like the ones found in those languages. Obviously, these categories and units may not be relevant in less dominant and/or non-Indo-European languages, yet they are still used as a model to construct examples in describing them resulting in a type of examples which suspiciously look like English and other Indo-European languages.[8]

Usage-based studies on languages outside the Indo-European family of languages in fact report that speakers of different languages may not equally orient to well established grammatical categories and units, highlighting the significance of this problem which has been hidden for the past several decades. Although complementation has

8 This is a general tendency observed especially in studies of non-European languages. For an example of this tendency, see representative Japanese studies such as Kuno (1973), Shibatani (1990), Iwasaki (2013), Tsujimura (2013), and Hasegawa (2015).

been proposed as a universal, Englebretson (2003) shows that Indonesian lacks it in conversational language; Nakayama (2002) suggests that the concept of the sentence is not useful for the analysis of Nuuchahnulth, a polysynthetic language; Laury et al. (to appear) show that the concept of clause does not work well for Japanese conversational data. However, this does not mean that studies of related languages would not reveal significant differences in links between form and function, as shown by the Zinken (2016) study quoted just above. Zinken shows that even genetically related European languages such as Polish and English may differ in significant ways, which in turn may result in linguistically and culturally distinct ways of performing 'the same' action. Such effects may even be discernible within the same language spoken in different areas (e. g. Swedish spoken in Finland and Sweden, Nilsson et al., 2018).

Language change and historical facts are considered to be very closely related to synchronic accounts. Purely synchronic accounts are deemed not ideal or even possible as language is a living entity undergoing change at every moment (Hopper, 1987 and, 2011). That is, we simply cannot afford to ignore diachronic factors. Only by considering diachrony can we begin to have an understanding of synchronic patterns and variations and of why facets of language are the way they are. Diachronic changes are reflected in synchronic data, as changing forms and functions may retain their old ones while new forms/functions keep evolving (a basic tenet of grammaticalization theory). This often results in less than perfect paradigms and patterns (i. e., variations), which only diachrony can account for.

Finally, human language has been thought to be part of general cognition (Langacker, 1987, 1991 and, 2003; Tomasello, 2003 and, 2008; Bybee, 2007 and 2010), and more recently it has been suggested that interaction is at least partly responsible for the formation of human cognition (Levinson, 2006). As we stated in an earlier section, many usage-based grammarians do not assume a separate language component in human cognition, an assumption made in autonomous linguistics.[9]

5 Evaluation

When choosing among alternative hypotheses, ultimately researchers are most interested in knowing what occurs, what does not occur, and what occurs most often in particular contexts. Specifically, we seek empirical support from what speakers actually do by examining recordings of talk and its transcripts, a critical component of our work. That is, minimally the form or structure in question and its function have to be what speakers actually, and more critically regularly, employ in naturally occurring

9 For a good current summary of this issue from the perspective in Interactional Linguistics see chapter 9 in Couper-Kuhlen and Selting (2018).

talk.[10] Obviously, neither intuition about what people say nor anecdotal observations about what they said meet this basic level requirement. In fact, researchers who are more interactionally oriented go even further by focusing on what they call 'participant orientation' (see Couper-Kuhlen and Selting, 2018, 25–26); they seek evidence for participants of the conversation orienting to the target form or function. For example, to those researchers, the identification of an utterance as a question minimally requires some indication of the orientation to it by the interactants, such as an answer to the question by the addressee.

6 Sample analysis

As we have noted previously, and as the term usage-based grammar implies, linguists working in this paradigm primarily, and even exclusively, analyze naturally occurring, corpus-based data. Practitioners of the two approaches we focus on here, Emergent Grammar and Interactional Linguistics, are interested in the use of language in context, especially ordinary everyday conversation. For these reasons, sentences such as the one below are not easily analyzable using the tools of usage-based grammar. Here is the sentence, which all authors of this volume were asked to analyze.

After Mary introduced herself to the audience, she turned to a man she had met before.

The example given above, in one sense, is not particularly unnatural in its form; it manifests features usage-based grammarians have found in naturally occurring data. Typological studies have shown that temporal adverbial clauses which express events occurring before the event in the main clause are preposed (e. g. Kortmann, 1991, 138; Diessel, 2005, 463). Due to this iconic principle, which states that the order of linguistic elements in discourse, such as clausal order, tends to follow the temporal order of events, *after*-clauses tend to be expressed before their main clauses, as is the case here. Although this example has a written-like feel, and could be constructed (no data source is given), it also has certain features found in spoken English data. Namely, it has been shown in a usage-based study that in English conversations, when the antecedent of the relative clause functions as an oblique, there tends strongly not to be a relativizer (Fox and Thompson, 2007); that is the case here. Further, syntactic lack of complexity of the antecedent also tends to predict the lack of a relativizer; the head here consists only of an article and a noun. Fox and Thompson suggest that the

10 Infrequently employed forms and functions are important but are not given priority simply because there is a sentiment shared among practitioners to want to capture global patterns. It also seems fair to say that some infrequently observed patterns may simply be errors, which might be worthy of investigation but again are not part of global patterns.

integration of main and relative clause combinations of this type involve a tendency toward monoclausality.

Since Emergent Grammar sees structure as emergent from its context, there is little we can say about this sentence in that respect, since it is separated from its context, even if it in fact were an actually occurring sequence. Further, in Interactional Linguistics, utterances are thought to result from and to reflect the actions they accomplish, but there is little to say about this matter, again since we do not know where this sentence might have occurred. Furthermore, the concept of sentence itself is not very useful for analyzing ordinary spoken data (see, e. g., Miller and Weinert, 1998; Iwasaki and Ono, 2002). Instead, the term 'clause combination' is often used.

In our own work, we have shown that complex clause combinations are rare in spoken language (Laury and Ono, 2010), especially when it comes to multiple embedding of the same type of subordinate clause within another one of the same type. The frequency of occurrence of embedding drops steeply beyond the depth of one, and even that is not the default option; while embedding at the depth of one occurred in 14.6 % of all clauses for Finnish and 14.1 % of all clauses for Japanese, embedding at the depth of two occurred in less than 5 % of all the main clauses (2.7 % for Finnish and 4.4 % for Japanese data). While the example above has no multiple embedding in the sense used in our study, that is, it has two dependent clauses each embedded in the main clause at the depth of one, it is still complex, since the main clause has two dependent clauses. It appears that clause combinations in spoken language are put together at the local level, one by one, in response to various online factors. Longer and more complex combinations involving several clauses may not form a coherent whole, and do not appear pre-planned (Laury and Ono, 2014). In other words, sequences of multiple clauses in spoken language do not result in what might be considered a grammatical sentence but rather form loosely connected sets of clauses which represent some discourse sequence or interactional unit (see, e. g., Auer, 1992; Hopper and Thompson, 2008). This is because speakers in ordinary conversation have difficulty sustaining syntactic projects for a long time, as memory for form is shorter than the memory for content (Auer, 2005, 27). Thus, while possible, clause combinations such as the one above are uncommon in ordinary spoken language. Written language contains more complex clause combinations, but even there, depth of embedding is strictly limited (see, e. g., Karlsson, 2009).

7 Conclusion

We have presented here two central approaches to usage-based grammar, Interactional Linguistics and Emergent Grammar. A central requirement, already implied in the name of the approach, is that serious linguistic study should focus on naturally occurring language, particularly everyday speech, the primary form of language.

We have reviewed the growing body of work in this paradigm and presented key findings, and also discussed the limitations of this approach, pointing out especially that building large scale corpora of naturally occurring speech is the next logical step toward further advancement in the field.

References

Arminen, Ilkka & Minna Leinonen. 2006. Mobile phone call openings: Tailoring answers to personalized summonses. *Discourse Studies* 8(3). 339–368.

Auer, Peter. 1992. The neverending sentence: Rightward expansion in Spoken language. In Miklós Kontra (ed.), *Studies in spoken languages: English, German, Finno-Ugric*, 41–59. Budapest: Hungarian Academy of Sciences, Linguistics Inst.

Auer, Peter. 2005. Projection in interaction and projection in grammar. *Text* 25(1). 7–36.

Auer, Peter. 2009. On-line syntax: Thoughts on the temporality of spoken language. *Language Sciences* 31. 1–13.

Auer, Peter. 2014. Syntactic structures and their symbiotic guests: Notes on analepsis from the perspective of online syntax. *Pragmatics* 24. 533–560.

Barlow, Michael & Suzanne Kemmer (eds.). 2000. *Usage-based models of language*. Stanford, CA: CSLI Publications.

Barth-Weingarten, Dagmar. 2016. *Intonation units revisited: Cesuras in talk-in-interaction*. Amsterdam: John Benjamins.

Bybee, Joan L. 1985. *Morphology: A study of the relation between meaning and form*. Amsterdam: John Benjamins.

Bybee, Joan L. 2007. *Frequency of use and the organization of language*. Oxford: Oxford University Press.

Bybee, Joan L. 2010. *Language, usage and cognition*. Cambridge: Cambridge University Press.

Bybee, Joan L. & Paul J. Hopper (eds.). 2001. *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins.

Chafe, Wallace (ed.). 1980. *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production*. Norwood, NJ: Ablex.

Chafe, Wallace. 1987. Cognitive constraints on information flow. In Russell Tomlin (ed.), *Coherence and grounding in discourse*, 21–51. Amsterdam: Benjamins.

Couper-Kuhlen, Elizabeth & Tsuyoshi Ono. 2007. Incrementing in conversation. A comparison of practices in English, German and Japanese. *Pragmatics* 17(4). 513–552.

Couper-Kuhlen, Elizabeth & Margret Selting. 2018. *Interactional linguistics: Studying language in social interaction*. Cambridge: Cambridge University Press.

Deppermann, Arnulf (ed.). 2013. Conversation analytic studies of multimodal interaction. *Journal of Pragmatics* 46. 1.

Deppermann, Arnulf & Silke Reineke. 2017. Epistemische Praktiken und ihre feinen Unterschiede: Verwendungen von *ich dachte* in gesprochener Sprache. In Arnulf Deppermann, Nadine Proske & Arne Zeschel (eds.), *Verben in interaktiven kontext. B ewegungsverben und mentale verben im gesprochenen Deutsch*, 337–375. Tübingen: Narr.

Diessel, Holger. 2005. Competing motivations for the ordering of main and adverbial clauses. *Linguistics* 43(3). 449–470.

Du Bois, John W. 1980. Beyond definiteness: The trace of identity in discourse. In Wallace Chafe (ed.), *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production*, 203–274. Norwood, NJ: Ablex.

Du Bois, John W. 1985. Competing motivations. In John Haiman (ed.), *Iconicity in syntax*, 343–365. Amsterdam: John Benjamins.

Du Bois, John W. 1987. The discourse basis of ergativity. *Language* 63(4). 805–855.

Ehmer, Oliver & Dagmar Barth-Weingarten (eds.). 2016. Adverbial patterns in interaction. [Special issue. *Language Sciences* 58.

Endo, Tomoko. 2013. Epistemic stance in Mandarin conversation: The positions and functions of *wo juede* (I feel/think). In Yuling Pan & Daniel Kádár (eds.), *Chinese discourse and interaction: Theory and practice*, 12–34. London: Equinox.

Englebretson, Robert. 2003. *Searching for structure: The problem of complementation in colloquial Indonesian conversation*. Amsterdam: John Benjamins.

Etelämäki, Marja & Laura Visapää. 2014. Why blend conversation analysis with cognitive grammar. *Pragmatics* 24(3). 477–506.

Evans, Nicholas. 2007. Insubordination and its uses. In Irina Nikolaeva (ed.), *Finiteness: Theoretical and empirical foundations*, 366–431. New York: Oxford University Press.

Ford, Cecilia E. 1993. *Grammar in interaction: Adverbial clauses in American English conversations*. Cambridge: Cambridge University Press.

Ford, Cecilia E. & Barbara Fox. 1996. Interactional motivations for reference formulation: *He* had. *This* guy had a beautiful, thirty-two O:lds. In Barbara Fox (ed.), *Studies in anaphora*, 145–168. Amsterdam: John Benjamins.

Ford, Cecilia E., Barbara A. Fox & Sandra A. Thompson. 2013. Units or action trajectories?: Is the language of grammatical categories the language of social action. In Beatrice Szczepek Reed & Geoffrey Raymond (eds.), *Units of talk – Units of action*, 13–58. Amsterdam: John Benjamins.

Fox, Barbara. 1987. *Discourse structure and anaphora: Written and conversational English*. Cambridge: Cambridge University Press.

Fox, Barbara, Makoto Hayashi & Robert Jasperson. 1996. Resources and repair: A cross-linguistic study of the syntactic organization of repair. In Elinor Ochs, Emanuel A. Schegloff & Sandra A. Thompson (eds.), *Interaction and grammar*, 185–227. Cambridge: Cambridge University Press.

Fox, Barbara & Sandra A. Thompson. 2007. Relative clauses in English conversation: Relativizers, frequency, and the notion of construction. *Studies in Language* 31(2). 293–326.

Geluykens, Ronald. 1992. *From discourse process to grammatical construction: On left-dislocation in English*. Amsterdam: John Benjamins.

Givón, Talmy. 1979. *On understanding grammar*. New York: Academic Press.

Goodwin, Charles. 1986. Gestures as a resource for the organization of mutual orientation. *Semiotica* 62(1–2). 29–49.

Goodwin, Charles. 2003. Pointing as a situated practice. In Sotaro Kita (ed.), *Pointing: Where language, culture and cognition meet*, 217–241. Mahwah, N.J.: Lawrence Erlbaum.

Goodwin, Charles. 2007. Environmentally coupled gestures. In Susan D. Duncan, Justine Cassell & Elena Terry Levy (eds.), *Gesture and the dynamic dimension of language*, 195–212. Amsterdam: Benjamins.

Goodwin, Marjorie Harness & Charles Goodwin. 1986. Gesture and co-participation in the activity of searching for a word. *Semiotica* 62(1–2). 51–75.

Gries, Stefan. 2009. What is corpus linguistics? *Language and linguistics compass* 3. 1–17.

Günthner, Susanne. 2008. Die Sache ist': Eine Projektorkonstruktion im gesprochenen Deutsch. *Zeitschrift für Sprachwissenschaft* 27(1). 39–72.

Günthner, Susanne. 2011. N be that-constructions in everyday German conversation. In Ritva Laury & Ryoko Suzuki (eds.), *Subordination in conversation*, 11–36. Amsterdam: John Benjamins.

Haakana, Markkku, Minna Laakso & Jan Lindström (eds.). 2009. *Talk in interaction: Comparative dimensions*. Helsinki: Suomalaisen Kirjallisuuden Seura.

Haddington, Pentti. 2012. Movement in action: Initiating social navigation in cars. *Semiotica* 191(1/4). 137–167.
Haddington, Pentti, Lorenza Mondada & Maurice Nevile (eds.). 2013. *Interaction and mobility: Language and the body in motion*. Berlin: Walter de Gruyter.
Haiman, John. 1983. Iconic and economic motivation. *Language* 59(4). 781–819.
Hakulinen, Auli. 1993. The grammar of opening routines. *Yearbook of the Linguistic Association of Finland* 15. 149–170.
Hasegawa, Yoko. 2015. *Japanese: A linguistic introduction*. Cambridge: Cambridge University Press.
Haspelmath, Martin. 2008. Frequency vs. iconicity in explaining grammatical asymmetries. *Cognitive Linguistics* 19(1). 1–33.
Haspelmath, Martin. 2010. Comparative concepts and descriptive categories in cross-linguistic studies. *Language* 86(3). 663–687.
Helasvuo, Marja-Liisa & Aki Kyröläinen. 2016. Choosing between zero and pronominal subject: Modeling subject expression in the first person singular in Finnish. *Corpus Linguistics and Linguistic Theory* 12(2). 263–299.
Hopper, Paul J. 1987. Emergent grammar. In Jon Aske, Natasha Berry, Laura Michaelis & Hana Filip (eds.), *Proceedings of the Thirteenth Annual Meeting of the Berkeley Linguistics Society*, 139–157. Berkeley: Berkeley Linguistics Society.
Hopper, Paul J. 2001. Grammatical constructions and their discourse origins: prototype or family resemblance? In Martin Pütz, Susanne Niemeier & René Dirven (eds.), *Applied cognitive linguistics I: Theory and language acquisition*, 109–129. Berlin: Mouton de Gruyter.
Hopper, Paul J. 2011. Emergent grammar and temporality in interactional linguistics. In Peter Auer & Stefan Pfänder (eds.), *Constructions: Emerging and emergent*, 22–44. Berlin: Mouton de Gruyter.
Hopper, Paul J. & Sandra A. Thompson. 1980. Transitivity in grammar and discourse. *Language* 56(2). 251–299.
Hopper, Paul J. & Sandra A. Thompson. 1984. The discourse basis for lexical categories in universal grammar. *Language* 60(4). 703–752.
Hopper, Paul J. & Sandra A. Thompson. 2008. Projectability and clause combining in interaction. In Ritva Laury (ed.), *Crosslinguistic studies of clause combining: The multifunctionality of conjunctions*, 99–124. Amsterdam: John Benjamins.
Imo, Wolfgang. 2011. On line changes in syntactic gestalts in spoken German. Or: do garden path sentences exist in everyday conversation. In Peter Auer & Stefan Pfänder (eds.), *Constructions: emerging and emergent*, 127–155. Berlin: de Gruyter.
Iwasaki, Shoichi. 2013. *Japanese*. Revised edn. Amsterdam: John Benjamins.
Iwasaki, Shoichi. 2015. A multiple-grammar model of speakers' linguistic knowledge. *Cognitive Linguistics* 26(2). 161–210.
Iwasaki, Shoichi & Tsuyoshi Ono. 2002. "Sentence" in spontaneous spoken Japanese discourse. In Joan L. Bybee & Michael Noonan (eds.), *Complex sentences in grammar and discourse: Essays in honor of Sandra A. Thompson*, 175–202. Amsterdam: Benjamins.
Karlsson, Fred. 2009. Origin and maintenance of clausal embedding complexity. In Geoffrey Sampson, David Gil & Peter Trudgill (eds.), *Language complexity as an evolving variable*, 192–202. Oxford: Oxford University Press.
Kärkkäinen, Elise. 2003. *Epistemic stance in English conversation: A description of its interactional functions, with a focus on I think*. Amsterdam: Benjamins.
Kärkkäinen, Elise & Tiina Keisanen. 2012. Linguistic and embodied formats for making (concrete) offers. *Discourse Studies* 14(5). 587–611.
Keevallik, Leelo. 2003. *From interaction to grammar: Estonian finite verb forms in conversation* (Acta Universitatis Upsaliensis. Studia Uralica Upsaliensia 34). Uppsala: Uppsala University.

Keevallik, Leelo. 2016. Abandoning dead ends: the Estonian junction marker *maitea*'I don't know. *Journal of Pragmatics* 106. 115–128.

Kortmann, Bernd. 1991. *Free adjuncts and absolutes in English: Problems of control and interpretation*. London and New York: Routledge.

Kuno, Susumu. 1973. *The structure of the Japanese language*. Cambridge, MA: MIT Press.

Langacker, Ronald W. 1987. *Foundations of cognitive grammar. Vol. 1, Theoretical prerequisites*. Stanford, CA: Stanford University Press.

Langacker, Ronald W. 1991. *Foundations of cognitive grammar. Vol. 2, Descriptive application*. Stanford, CA: Stanford University Press.

Langacker, Ronald W. 2003. Dynamic conceptualization in grammatical structure. *Research in Language* 1. 53–79.

Langacker, Ronald W. 1999. *Grammar and conceptualization*. Berlin: Mouton de Gruyter.

Laury, Ritva & Tsuyoshi Ono. 2005. Data is data and model is model: You don't discard the data that doesn't fit your model!. *Language* 81. 218–225.

Laury, Ritva & Tsuyoshi Ono. 2010. Recursion in conversation: What speakers of Finnish and Japanese know how to do. In Harry van der Hulst (ed.), *Recursion and human language*, 69–92. Berlin: Mouton de Gruyter.

Laury, Ritva & Tsuyoshi Ono. 2014. The limits of grammar: Clause combining in Finnish and Japanese conversation. *Pragmatics* 24(3). 561–592.

Laury, Ritva & Tsuyoshi Ono (eds.). to appear. *Fixed expressions: Building language structure and action*.

Laury, Ritva, Tsuyoshi Ono & Ryoko Suzuki. to appear. Questioning the clause as a crosslinguistic unit in grammar and interaction. *Studies in Language*.

Lerner, Gene. 1996. On the place of linguistic resources in the organization of talk-in-interaction: "Second person" reference in multi-party conversation. *Pragmatics* 6(3). 281–294.

Levinson, Stephen C. 2006. On the human "interaction engine". In Nick J. Enfield & Stephen C. Levinson (eds.), *Roots of human sociality: Culture, cognition and interaction*, 39–69. Oxford: Berg.

Li, Charles. 1976. *Subject and topic*. New York: Academic Press.

Li, Xiaoting. 2014. *Multimodality, interaction and turn-taking in Mandarin conversation*. Amsterdam: John Benjamins.

Li, Xiaoting & Tsuyoshi Ono (eds.). in press. *Multimodality in Chinese Interaction*. Berlin: Mouton de Gruyter.

Lindström, Jan. 2014. On the place of turn and sequence in grammar: Verb-first clausal constructions in Swedish talk-in-interaction. *Pragmatics* 24(3). 507–532.

Lindström, Jan, Camilla Lindholm & Ritva Laury. 2016. The interactional emergence of conditional clauses as directives: constructions, trajectories, and sequences of actions. *Language Sciences* 58. 8–21.

Linell, Per. 2005. *The written language bias in linguistics: Its nature, origins, and transformations*. New York: Routledge.

Miller, Jim. 1995. Does spoken language have sentences. In Frank R. Palmer (ed.), *Grammar and meaning: Essays in honour of Sir John Lyons*, 116–135. Cambridge: Cambridge University Press.

Miller, Jim & Regina Weinert. 1998. *Spontaneous spoken language: Syntax and discourse*. Oxford: Clarendon Press.

Mondada, Lorenza. 2007. Multimodal resources for turn-taking: Pointing and the emergence of possible next speakers. *Discourse Studies* 9(2). 195–226.

Mondada, Lorenza. 2009. The embodied and negotiated production of assessments in instructed actions. *Research in Language and Social Interaction* 42(4). 329–361.

Nakayama, Toshihide. 2002. *Nuuchahnulth (Nootka) morphosyntax*. Berkeley: University of California Press.
Nakayama, Toshihide & Kumiko Ichihashi-Nakayama. 1997. Japanese *kedo*: Discourse genre and grammaticization. In Ho-min Sohn & John Haig (eds.), *Japanese/Korean Linguistics* 6, 607–618. Stanford: CSLI Publications.
Nilsson, Jenny, Stefan Norrthon, Jan Lindström & Camilla Wide. 2018. Greetings as social action in Finland Swedish and Sweden Swedish service encounters – A pluricentric perspective. *Intercultural Pragmatics* 15(1). 57–88.
Ochs, Elinor, Emanuel Schegloff & Thompson A. Sandra (eds.). 1996. *Grammar and interaction*. Cambridge: Cambridge University Press.
Ono, Tsuyoshi, Ritva Laury & Ryoko Suzuki (eds.). to appear. Usage-based and typological approaches to linguistic units. *Studies in Language*.
Ono, Tsuyoshi & Sandra A. Thompson. 1995. What can conversation tell us about syntax. In Philip W. Davis (ed.), *Descriptive and theoretical modes in the new linguistics*, 213–271. Amsterdam: Benjamins.
Ono, Tsuyoshi & Sandra A. Thompson. 2017. Negative scope, temporality, and clause structure in Japanese and English conversation. *Studies in Language* 41(3). 543–576.
Ono, Tsuyoshi, Sandra A. Thompson & K. K. Luke (eds.). 2012a. Cross-linguistic, multi-modal, and grammaticization perspectives on turn continuation in conversation. *Discourse Processes* 49. 3–4.
Ono, Tsuyoshi, Sandra A. Thompson & Yumi Sasaki. 2012b. Japanese negotiation through emerging final particles in everyday talk. *Discourse Processes* 49. 243–272.
Pekarek Doehler, Simona. 2011. Clause-combining and the sequencing of actions: Projector constructions in French talk-in-interaction. In Ritva Laury & Ryoko Suzuki (eds.), *Subordination in conversation*, Amsterdam: John Benjamins.
Pekarek Doehler, Simona, Elwys De-Stefani & Anne-Sylvie Horlacher. 2015. *Time and emergence in grammar. Dislocation, topicalization and hanging topic in French talk-in-interaction*. Amsterdam: John Benjamins.
Rauniomaa, Mirka & Tiina Keisanen. 2012. Two multimodal formats for responding to requests. *Journal of Pragmatics* 44(6–7). 829–842.
Rossi, Giovanni. 2015. *The request system in Italian interaction*. Nijmegen: Radboud University dissertation.
Sacks, Harvey. 1992. Lectures on conversation. Vols. 1 & 2. Gail Jefferson (ed.), Cambridge: Blackwell.
Sacks, Harvey, Emanuel A. Schegloff & Gail Jefferson. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language* 50(4). 696–735.
Schegloff, Emanuel A. 1986. The routine as an achievement. *Human Studies* 9(2). 111–151.
Schegloff, Emanuel A. 1996. Turn organization: One intersection of grammar and interaction. In Elinor Ochs, Emanuel A. Schegloff & Sandra A. Thompson (eds.), *Interaction and grammar*, 52–133. Cambridge: Cambridge University Press.
Schegloff, Emanuel A. 2006. Interaction: The infrastructure for social institutions, the natural ecological niche for language, and the arena in which culture is enacted. In Nick J. Enfield & Stephen C. Levinson (eds.), *Roots of human sociality: Culture, cognition and interaction*, 70–96. London: Berg.
Schegloff, Emanuel A. 2007. *Sequence organization in interaction: A primer in conversation analysis*. Cambridge: Cambridge University Press.
Selting, Margret. 2013. Verbal, vocal, and visual practices in conversational interaction. In Cornelia Müller, Alan Cienki, Ellen Fricke, Silva Ladewig, David McNeil & Sedinha Tessendorf (eds.),

Body – Language – Communication. An international handbook on multimodality in human interaction. Volume 38.1, 589–609. Berlin: De Gruyter Mouton.
Selting, Margret & Elizabeth Couper-Kuhlen. 2000. Argumente für die Entwicklung einer interaktionalen Linguistik. *Gesprächsforschung – Online-Zeitschrift zur verbalen Interacktion* 1. 76–95. www.gespraechsforschung-ozs.de.
Selting, Margret & Elizabeth Couper-Kuhlen (eds.). 2001. *Studies in interactional linguistics*. Amsterdam: John Benjamins.
Shibatani, Masayoshi. 1990. *The languages of Japan*. Cambridge: Cambridge University Press.
Sidnell, Jack. 2009. *Conversation Analysis: An introduction*. Oxford: Wiley-Blackwell.
Sidnell, Jack & Tanya Stivers (eds.). 2012. *The handbook of Conversation Analysis*. Chichester: Wiley-Blackwell.
Sorjonen, Marja-Leena & Liisa Raevaara. 2014. On the grammatical form of requests at the convenience store: Requesting as embodied action. In Paul Drew & Elizabeth Couper-Kuhlen (eds.), *Requesting in social interaction*, 241–266. Amsterdam: John Benjamins.
Stevanovic, Melisa. 2016. Keskustelunanalyysi ja kokeellinen vuorovaikutustutkimus [Conversation Analysis and experimental interaction research]. In Melisa Stevanovic & Camilla Lindholm (eds.), *Keskustelunanalyysi*, 390–409. Tampere: Vastapaino.
Szczepek Reed, Beatrice & Geoffrey Raymond (eds.). 2013. *Units of talk – Units of action*, 169–206. Amsterdam: John Benjamins.
Thompson, Sandra A. 2002. Object complements and conversation: Towards a realistic account. *Studies in Language* 26(1). 125–163.
Thompson, Sandra A., Barbara A. Fox & Elizabeth Couper-Kuhlen. 2015. *Grammar in everyday talk: Building responsive actions*. Cambridge: Cambridge University Press.
Tomasello, Michael. 2003. *Constructing a language: A usage-based theory of language acquisition*. Harvard University Press.
Tomasello, Michael. 2008. *Origins of Human Communication*. Cambridge, MA: MIT Press.
Tsujimura, Natsuko. 2013. *An introduction to Japanese linguistics*, 3rd edn. Hoboken: John Wiley & Sons Inc.
Zinken, Jörg. 2016. *Requesting responsibility. The morality of grammar in Polish and English family interaction*. Oxford: Oxford University Press.

Géraldine Legendre

10 Optimality-theoretic Syntax

Abstract: OT is a general competition-based theory of generative grammar based on a particular theory of constraint interaction and a theory of the economy of constraint violation. It departs from other generative theories of syntax along three main dimensions: (i) well-formedness constraints on linguistic representations are violable, (ii) constraint universality is logically independent of constraint violability, (iii) the representations which constraints operate on (e. g. building blocks, levels of representation, representational vs. derivational architecture) are orthogonal to that of constraint interaction. Therefore, OT defines a class of grammatical theories that are as diverse as other theories unified under other umbrellas, e. g. Principles and Parameters theories, Unification-Based theories. OT theories share the property that it is unnecessary to stipulate special principles of Economy, Relativized Minimality, Last Resort, and the like because these effects follow as inevitable logical consequences of the general competition-based architecture. OT is used to model grammars, including the interaction of various components of the grammar, but also on-line language processing, acquisition of language, diachronic change, etc. In sum, OT is not a theory of representations, but a theory of the interaction of universal principles in any linguistic domain. (This chapter samples the OT literature and is not intended to be exhaustive).

1 Goals

Optimality Theory (OT) is a development of Generative Grammar which has its roots in Cognitive Science and a fundamental mind/brain question (Smolensky and Legendre, 2006). On the one hand the human *brain* is mathematically a computer consisting of a network of interconnected units (neurons) which performs continuous numerical processing including *optimization*: it maximizes a measure of self-consistency or well-formedness. On the other hand, the human *mind* is a symbol-manipulating computer. As a theory of grammar, OT helps resolve the continuous vs. digital computational tension between the biological and the mental levels by positing that grammatical mental representations are symbolic structures which maximize well-formedness and are optimal (rather than perfect) thanks to an evaluation procedure selecting the best among a set of alternatives.

Acknowledgement: Thanks to the editors of this volume, two fellow contributors, and Paul Smolensky for their feedback on an earlier version of the present chapter. Thanks also to the many audiences that since 1990 have helped me sharpen my own views on the true nature of cross-linguistic variation in syntax and the development of a descriptively and explanatorily adequate theory grounded in fundamental properties of mental computation. And thanks to NSF for an INSPIRE grant (BCS-1344269) which has partially supported the research discussed in this chapter.

https://doi.org/10.1515/9783110540253-010

The very architecture of the theory is grounded in general human cognition. Optimality is a pervasive concept in a variety of cognitive domains including basic memory function (Anderson, 1990), computing the most likely 3D source for a 2D image in visual cognition (Geman and Geman, 1984), learning internal representations in neural networks (Rumelhart et al., 1986), and Bayesian learning in any higher cognition domain (Oaksford and Chater, 1998). These examples are ultimately all formalized in terms of competition plus an algorithm or procedure for choosing an output that optimizes a specified objective function. While the resolution of conflicting information by strict domination hierarchies appears to be a hallmark of grammatical knowledge (Smolensky and Legendre, 2006, 1:41–42) it has also been argued to be a characteristic of decision making, e. g. the Take the Best decision procedure (Gigerenzer and Golstein, 1999; Rieskamp and Hoffrage, 1999).

In OT syntax, grammaticality is tied to a *competition*-based view of grammar, which at the most general level is also found in some usage-based and typological-functional approaches to grammar (e. g., MacWhinney et al., 2014). However, OT (including related theories) constitutes a unique approach to generative grammar that sets it apart from other generative theories, including most versions of Principles and Parameters theories (PPT), Head-driven Phrase Structure Grammar (HPSG), and the pre-OT version of Lexical-Functional Grammar (LFG), to name a few. In the latter theories, satisfaction of the inviolable principles of the theory is formally evaluated for a single sentence at a time, in isolation from alternative sentences. In OT, that decision is arrived at by considering a competition among an appropriately defined set of related sentences, winnowing the worst ones, and declaring the one with the best constraint profile as optimal, hence grammatical. Competition-based generative approaches to syntax outside of OT exist, including early versions of the Minimalist Program (MP; Chomsky, 1991, 1993) where derivations compete subject to transderivational economy constraints (e. g., the Fewest Steps Condition governing syntactic movement), which select the most economical derivation in the reference set (i. e., the set of lexical items that are used in a derivation). One important difference concerns the evaluation of candidate structures: in OT there is an explicit procedure evaluating optimality with respect to a set of ranked, violable constraints; in Early MP there is no single explicit procedure that accommodates inviolable constraints (Müller and Sternefeld, 2001, 27–30). Later versions of MP (e. g., Chomsky, 2000, 2001), however, have eliminated transderivational constraints and along with it the concept of competition-based syntax relying on a candidate/reference set. Still, specific principles of economy, complementarity, and competition *with the comparison built into the principles* (e. g. Last Resort, Procrastinate, Relativized Minimality, etc.) are routinely appealed to in current MP analyses. Building the comparison into the statement of individual principles typically leads to complex statements and is not obviously in line with parsimony but it makes it possible to maintain inviolability (aside from possible parametric variation). OT takes the view that optimality results instead from the *resolution of conflicts* (discussed further below on the basis of linguistic examples) amongst formally simple and violable constraints. The debate ultimately rests on whether competition is the norm (construed as *theorems* that follow from completely general principles of a competition-based grammatical framework in the case of OT) or the exception (e. g. in the specific principles of MP that invoke it).

While OT exploits a domain-general procedure of optimization, it shares with other generative approaches a focus on UG and its domain specificity, and on formal description. In contrast to other approaches, however, OT posits that universal constraints are soft or *violable*, due to its roots in neural computation (where the well-formedness of a network state is the degree to which the state satisfies a set of soft constraints implemented in the network's connections). This in turn means that a surface grammatical structure is typically not perfect (in the sense of satisfying all universal principles of the theory); rather it is merely the best — and what counts as the best varies cross-linguistically, regulated by the central principles of the theory. OT is thus a theory of typology predicting all and only possible languages.

To illustrate the central concept of constraint violability (see Legendre et al., 2016a, 2–10 for a basic discussion of other common optimality effects), consider Germanic object shift. As discussed in Diesing (1996), sentences without object shift in Icelandic (1a) and their object shift counterparts (1b) have different interpretations.

(1) a. Í prófunum svarar hann sjaldan [$_{VP}$ erfiðustu spurningunni]
in exams-the answers he rarely most-difficult question-the
'He rarely answers the most difficult question of an exam'

b. Í prófunum svarar hann erfiðustu spurningunni$_i$ sjaldan [$_{VP}$ t_i]
in exams-the answers he most-difficult question-the rarely
'The most difficult of all questions, he rarely answers on exams'

(1a) has narrow scope interpretation: regardless of which exam the referent of 'he' is taking, he rarely answers whichever question happens to be the most difficult one in that particular exam. In contrast, (1b) has wide scope interpretation: there is one particular question which is more difficult than all others and which appears in all exams, and when he encounters it, he rarely answers it. Vikner (2001) observes that object shift is ungrammatical in compound tense contexts (2b) and the only grammatical sentence without object shift (2a) is actually ambiguous; it has both interpretations.

(2) a. Í prófunum hefur hann sjaldan svarað [$_{VP}$ erfiðustu spurningunni]
in exams-the has he rarely answered most-difficult question-the
'He has rarely answered the most difficult question in exams
'The most difficult of all questions, he rarely answers on exams'

b. *Í prófunum hefur hann svarað erfiðustu spurningunni$_i$
in exams-the has he answered most-difficult question-the
sjaldan [$_{VP}$ t_i]
rarely

As Vikner (2001, 327) puts it, "what matters is not just whether the object has undergone object shift or not, but also whether it could have moved if it had wanted to". Diesing originally proposed that the scope of objects is read off their surface position (Scoping Condition) and took the adjoined adverbial *sjaldan* 'rarely' to mark the left periphery of VP in (1). An object located within VP entails narrow scope interpretation, as in (1a) and (2a). Movement of the object outside of VP results in a wide scope interpretation (1b). An obvious problem for constraint inviolability arises with (2a), where the generalization embodied in the Scoping Condition does not hold. (2a) has a possible wide scope interpretation despite the object being located in VP. In OT terms there is simply no alternative because a constraint violated by object shift in a compound tense context is relatively stronger than the constraint aligning surface position and scope. Thus, a parsimonious account of (1)–(2) together is made available by abandoning the view that Diesing's Scoping Condition is inviolable.

Characterizing the set of possible grammars that comprise UG minimally requires specifying (i) the mental representations that are characteristic of language, (ii) the substantive constraints that distinguish possible from impossible linguistic systems, and (iii) the formal mode of interaction among these constraints. In contrast to other generative theories of syntax OT is not a substantive theory of any particular type of syntactic representation and is not committed to any specific type of constraints (other than violability). This is why optimality-theoretic analyses inspired by different types of substantive theories of syntax can be found in the literature (Legendre et al., 2001), including Government-Binding Theory (e. g., Legendre et al., 1995; Legendre et al., 1998; Grimshaw, 1997; Burzio, 1998), LFG (e. g., Bresnan, 2000; Sells, 2001), Construction Grammar (e. g., Vogel, 2016), and MP (e. g., Speas, 1997, 2001; Müller, 2001; Heck, 2000; Heck and Müller, 2013, 2016). Rather than a theory of syntax per se, OT is a general formal theory of constraint interaction. OT constraints conflict because they are stated in very general terms. They eschew logical complexity because complexity is derivative in an OT system: It is the product of the *interaction* of constraints, not the constraints themselves. In fact, any empirical generalization formulated as a disjunction is the clearest clue to the existence of violable constraints, as pointed out in Speas (1997, 184–185).

Linguistic structures which surface in a language are not inherently grammatical or ungrammatical, rather they are determined to be grammatical through a process of evaluation of a set of candidate surface structures, resulting in selection of an optimal one which is ipso facto grammatical. OT's basic assumptions are the following:

(3) (i) UG is an optimizing system of universal well-formedness constraints on linguistic forms.

(ii) Well-formedness constraints are simple and general. They routinely come into conflict and are (often) violated by surfacing forms.

(iii) Conflicts are resolved through *hierarchical rankings* of constraints. The effect of a given constraint is relative to its ranking, which is determined on a language-particular basis.
(iv) Evaluation of candidates by the set of constraints is based on strict domination. For any two constraints C_1 and C_2, either C_1 outranks C_2 (written $C_1 \gg C_2$) or C_2 outranks C_1. The demands of any constraint C have strict veto power over the (combined) demands of all constraints that C outranks.
(v) Alternative structural realizations of an input compete for the status of being the optimal output of a particular input. The most harmonic output — the one which best satisfies, or minimally violates, the full set of ranked constraints in a given language — is the optimal one. Only the optimal structure is grammatical.
(vi) Every competition yields an optimal output.

Importantly, and in contrast with other generative theories, constraint universality is not construed as constraint inviolability. All constraints are present in the grammars of all languages but they are violable. Furthermore, being low-ranked does not entail having no effect within a language. Rather, the effect of low-ranked constraints in eliminating sub-optimal candidates may be detected in some syntactic contexts but not others, as is demonstrated in the Tools section via examples.

2 Data (and their impact on the theory of constraint interaction)

OT is a meta-theory applicable to all components of the grammar. It has been applied to phonology (e. g., Prince and Smolensky, [1993] 2004; McCarthy and Prince, 1993a), syntax (e. g., Legendre et al., 1993; Samek-Lodovici, 1996; Grimshaw, 1997; Legendre et al., 2001; Burzio, 2010; Legendre et al., 2016a), semantics and pragmatics (Hendriks and de Hoop, 2001; Blutner and Zeevat, 2004; Blutner et al., 2006; H. de Swart, 2010; Hendriks et al., 2010), and the syntax-prosody interface (Samek-Lodovici, 2005, 2015). OT is also relevant to the characterization of performance (e. g., online syntactic processing: Stevenson and Smolensky, 2006; Hoeks and Hendriks, 2011), monolingual and bilingual acquisition of language (Legendre et al., 2002; Legendre et al., 2004; Hendriks and Spenader, 2005/6; Hsin, 2014), diachronic change (Vincent, 2000; Slade, 2003; LaFond, 2003), heritage grammars (Bousquette et al., 2016), code-switching (Bhatt, 1997, 2014; Legendre and Schindler, 2010), discourse (Beaver, 2004), learning algorithms (Tesar and Smolensky, 1998, 2000), and processing algorithms (Ellison, 1994; Tesar, 1994 et seq.; Frank and Satta, 1998; Karttunen, 1998).

OT analyses thus rely on any data used in linguistics and psycholinguistics. These include introspections of the sort that is standard in generative grammar as well as data from corpora, including quantified data, e. g. from CHILDES database (e. g., Legendre et al., 2002, 2004; Hsin, 2014), and experimental data (e. g., Hendriks and Spenader, 2005/6; Stevenson and Smolensky, 2006; Hsin, 2014). OT analyses of the grammar tend to rely on binary grammaticality judgments and yield one candidate that is optimal in Classical OT (Prince and Smolensky, [1993] 2004), which is equated with being grammatical. Graded grammaticality judgments like 'more acceptable' are typically equated with 'more probable' outputs in Stochastic OT and more frequent outputs in Partial Ordering OT (see further discussion below).

Domination is the primary mode of interaction between two constraints C_1 and C_2: C_1 outranks C_2 or C_2 outranks C_1. In Classical OT, strict domination holds over the rankings — any higher-ranked constraint takes absolute priority over any lower-ranked constraint (i. e. a single violation of a higher-ranked constraint is always worse than any number of violations of any number of lower-ranked constraints). Strict domination allows a single optimal output to emerge; it is thus a particular theoretical assumption which further specifies constraint violability.

Contexts of change, e. g. language acquisition or diachronic change, however, are empirical domains where old and new structures often coexist at various stages, hence every comprehensive theory of grammatical phenomena must be able to model optionality tied to a grammar in flux. In OT, this can be modeled by *partial ordering* of constraints whereby the relative ranking of some constraints may be indeterminate in the grammar and 'float', resulting in a *set of strict rankings*, each yielding a potentially different optimal output. Analyses relying on partial ordering include dialectal variation in morphology as in the original proposal (Antilla, 1997), stages of language acquisition (Legendre et al., 2002, 2004), and stages of diachronic change (Slade, 2003).

Consider, for example, the adult sentence 'Georges pushed me' and how the same proposition is expressed by a typical 3-year old child acquiring English. The ensuing characteristic pattern of child 'Optional Infinitives' (Wexler, 1994) is the following: Part of the time the child will produce the adult utterance and part of the time she will produce an uninflected version of it. The child is hypothesized to have two strict rankings which may produce a different optimal output and which she may randomly make use of (it is tempting to say that two rankings equal two grammars but they are not formally equivalent); two or more distinct rankings may in fact result in the same optimal output (Legendre et al., 1993; see discussion below of factorial typologies).

In syntactic production the input is made of the intended meaning, including a Tense specification, and the output corresponds to the best possible expression of that input. In the case of verbal inflection the adult-like output 'Georges pushed me' comes from a ranking whereby linguistically expressing important distinctions such as Tense is paramount (PARSETENSE) but at the cost of building structure (e. g. a functional projection to house Tense under a syntactic analysis or an additional morpheme under a morphological analysis) in violation of Economy of Structure (*STRUCTURE, where * means 'minimize' in this filtering type of constraint (Prince and Smolensky, [1993]/2004)).

Competitions are formally made explicit in table where the optimal candidate is identified by the pointing finger ☞. Constraint ranking is indicated by the left-to-right order, each constraint dominating the ones on its right. Violations of constraints are recorded as * in individual cells; *! are fatal violations for sub-optimal candidates. In syntax, the input to optimization minimally includes the lexical verb and its argument structure, as well as a tense specification.

Table 1 instantiates the adult optimization. Each candidate output is favored by one of the constraints listed in the right-most columns. If ParseTense dominates *Structure, candidate a containing the inflectional morpheme –*ed* encoding past tense is evaluated as optimal, compared to candidate b (without the –*ed* morpheme). That is, candidate a only violates the lower-ranked constraint, *Structure. Candidate b is *less harmonic than* a (i. e. fares worse than a) because it violates the higher-ranked constraint, ParseTense.

Table 1: (Input: $push_V$ (x, y); x = *George*, y = *me*; Tense = past).

	ParseTense	***Structure**
☞ a. George pushed me		*
b. George push me	*!	

If, however, the two constraints are allowed to *float* with respect to one another, two optimizations result, one shown in Table 1 with optimal output candidate a and the other in Table 2 (where *Structure dominates ParseTense) with optimal output candidate b. Legendre et al. (2002, 2004) propose that child grammars have precisely this character of being indeterminate in the sense that they correspond to more than one strict ranking. The resulting pattern is optionality of forms.

Table 2: (Input: push_V (x, y); x = *George*, y = *me*; Tense = past).

	***Structure**	**ParseTense**
a. George pushed me	*!	
☞ b. George push me		*

Assuming the alternative rankings of *Structure and ParseTense are equiprobable, we can equivalently state that output candidates a and b are optimal 50 % of the time each when the two constraints float. In other words, the corresponding pattern of production is a quantitative one: a 50/50 distribution or optionality of two alternative forms. Crucially, when more than two constraints are at play, as is more realistically the case with verbal inflection (Legendre et al., 2002, 2004), Economy of

Structure constraints corresponding to several functional projections (or several morphemes) may be strictly ranked while several PARSE constraints (e. g. Tense, Person agreement, etc.) may float above and under the strictly ranked ones, with quantitative outcomes following a more graded distribution (33 %–66 %; 25–75 %, etc.). Stochastic OT variants have also been developed; these allow for more fine-grained quantitative variation and greater gradience by assigning a numerical probability to each strict ranking (Boersma and Hayes, 2001; see Bresnan and Nikitina, 2009 for an application to syntax).

Economy and optimality essentially exclude optionality. In synchronic grammars optionality is in fact most often *apparent*, hiding important differences in information status, register, dialect, etc., which are formalized in OT in a natural way. An example of apparent optionality in German arising from different optimizations based on different inputs is discussed in the Tools section below. Apparent optionality resulting from different idiolects, registers of a single language is implemented as different constraint rankings. French, for example, exhibits a rich typology of wh-questions reflecting register differences (Colloquial French favors in situ wh with special prosody but Standard French favors either wh-fronting with simple inversion or with complex inversion whereby the lexical subject is doubled with a clitic pronoun (Rizzi and Roberts, 1989)). Register differences are, formally speaking, a subcase of cross-linguistic variation. They correspond to different grammars. Still, it may be, pending actual investigation, that compared with standard cross-linguistic variation, register variation involves minimal constraint re-ranking, possibly involving one constraint only.

Local conjunction of constraints constitutes a secondary mode of constraint interaction in OT (Smolensky, 1993, 1995, 1997). Two constraints C_1 and C_2 may combine to form a new constraint such that when both constraints are violated within a specified local domain, the new constraint is violated: this new constraint is the Local Conjunction (C_1&C_2) of the two original constraints, which it outranks (Legendre et al., 1995, 1998; Aissen, 1999, 2003). For example, wh-structures are well-known to be sensitive to Locality which can be couched in terms of a violable family of MINLINK constraints ('No long chain links'; Legendre et al., 1995, 1998). Referentiality (Rizzi, 1990) ('No non-referential chains') also plays a role whereby more referential/argumental wh-phrases tend to be cross-linguistically more extractable than non-referential/adjunct ones. Legendre et al. (1998) incorporate into an OT analysis of such cumulative effects a local conjunction of two constraints/dimensions, MINLINK and REFERENTIALITYinto $\text{MINLINK}_{\text{REF}}$ ('No long and non-referential chain links') which outranks MINLINK.

Further proposals have sought to relax the strict domination assumption of Classical OT illustrated above. First, two constraints C_1 and C_2 may be *tied* in ranking in a single table, whereby a violation of C_1 in one candidate and a violation of C_2 in another cancel each other: the two constraints are merged and function as a single constraint, yielding a pattern of surface optionality (however, if C_1 and C_2 are high-ranked, lower-ranked constraints tend to become decisive and the outcome is not optionality). In contrast to constraint ties, partial constraint ordering yields a set of rankings, hence multiple grammars/tablex. This set of rankings yields potentially different optimal outputs, hence variation. In no sense are two constraints functioning as one in partial ordering.

A second departure from strict domination involves numerical weightings of constraints (characteristic of Harmonic Grammar), which brings the evaluation closer still to its neural network roots (Legendre et al., 1990; Pater, 2009, 2016; Murphy, 2017). In Harmonic Grammar, combined violations of two constraints C_1 and C_2 with lower weights can gang up and outweigh the violation of another Constraint C_3 with a higher weight than both C_1 and C_2 individually. Historically, OT was introduced as a development of Harmonic Grammar when strict domination was substituted for the prior numerical weighting of constraints (Smolensky and Legendre, 2006). In current work on Gradient Symbolic Computation, numerical differences not only distinguish constraint weights (as in Harmonic Grammar), but in addition both the inputs and outputs of the grammar incorporate numerical distinctions in the degree of presence, or 'activity', of symbols within the linguistic structure. For instance, in the analysis of French liaison of Smolensky and Goldrick (2016), liaison consonants (consonants that alternate with zero) are deficient in their degree of presence in inputs.

3 Tools (with sample analyses)

OT relies on an input-output mapping architecture. An OT grammar comprises several universal components: (i) A set of possible *inputs* (intended meanings in syntax, syntactic structures in semantics/pragmatics), (ii) A mechanism for producing candidate structural descriptions of the input (*Gen* for 'Generator'), (iii) A procedure for evaluating the relative well-formedness (or *Harmony*) of candidate structural descriptions (*H-Eval* for 'Harmony Evaluator'), which depends on a set of well-formedness constraints *Con*.

Gen freely generates all of the types of structures that are present in any of the world's languages by incorporating a small set of very basic inviolable principles, e. g., pertaining to combining words into phrases. The main role of the input to syntax is to determine what competes; what wins is determined by the constraints. OT syntacticians generally agree that the input must specify predicate-argument structure, lexical items, functional features (tense, aspect, etc.), and where relevant, information and illocutionary features, level of argument prominence, prosodic categories, etc. Operator scope must also be included according to Legendre et al. (1995, 1998). For a given input, the grammar generates and evaluates an infinite set of output candidates which represent alternative structural realizations of that input. If the candidate set includes candidates with different LFs then the input to, say, a wh-question must also include target [wh] and operator scope specifications. Properties of the input and the candidate set are to a large extent determined by the underlying substantive theory of syntax. Thus the input in the optimization examples discussed below and the corresponding candidate set borrow extensively from the type of representation assumed in PPT. As noted earlier, the question of the substantive syntactic constraints and representations is independent of the claims made by OT.

Broadly speaking, the input to syntax thus contains all elements relevant to an interpretation while the output is a structured entity which is the surface expression of that interpretation. Optimization run in the opposite direction (from an expression to an interpretation) has been developed to model semantics and pragmatics in OT (Hendriks and de Hoop, 1997, 2001; de Hoop and de Swart, 2000; Blutner and Zeevat, 2004, 2009; Beaver, 2004; Blutner et al., 2006). The realization that competition in form and meaning often hangs together has led to the development of bidirectional OT (Blutner, 2000; Jäger, 2004; Hendriks et al., 2010; Mattausch, 2007). It is particularly relevant to blocking and freezing phenomena (Beaver and Lee, 2004). For applications to the syntax/semantics interface see, for example, de Hoop and Malchukov (2008) and Legendre et al. (2016b).

The issue of whether the grammar is derivational or representational is orthogonal to OT. As emphasized in Prince and Smolensky ([1993] 2004: 95–6) and Legendre et al. (1998, 286–7) OT makes no commitment to a serial or parallel architecture. While many analyses developed in one perspective can be rephrased in the other (e. g., Hale and Legendre, 2004) Gereon Müller and colleagues have motivated a derivational version of OT syntax involving optimization at every step (e. g., Müller, 2001, Heck, 2000, Heck and Müller, 2013, 2016).

OT relies on two types of constraints: *markedness* constraints, like well-formedness constraints in other frameworks, evaluate the inherent goodness of a candidate output, for example its economy of representation (e. g., *Structure); *faithfulness* constraints regulate the input-output mapping and are unique to OT. Input-output faithfulness constraints limit how far candidate outputs may differ from the input. They require the output to express all and only the properties of the input (e. g., ParseTense). Faithfulness and markedness constraints are often conflicting. Preserving a contrast encoded in the input via a feature will typically lead to some cost in markedness, e. g. satisfying ParseTense entails violating *Structure in Tableaux 1 and 2. OT defines a marked structure as one which violates (any number of) markedness constraints. If no faithfulness constraint is violated, then the least marked candidate wins the competition and is declared grammatical. However, trade-offs between markedness and input-output faithfulness are commonplace in syntax and are explored further below.

Faithfulness constraints are crucial to the OT conception of grammar and have played a pivotal role since the theory's inception. In syntax, their role is central to the treatment of ineffable meanings and absolute ungrammaticality, e. g., in multiple wh-questions.

(4) a. *Lisi zhidao shenme shenme shihou?*
Lisi know what what time
'What did Lisi know when?'
(Mandarin, Huang, 1982)

b. *What did Congress know when?*

c. *Koj kakvo na kogo e da ?*
who what to whom has given
'Who gave what to whom?'
(Bulgarian, Rudin, 1988)

d. **Cé aL rinne ciadé?*
who that did what
'Who did what?'
(Irish, McCloskey, 1979)

(4) shows the extent of variation in the presence of two wh-phrases: both are in situ in Mandarin, only one is fronted in English, all are fronted in Bulgarian, and sentences containing two wh-phrases (or more) are ungrammatical in Irish, regardless of their position. Only one wh-phrase may surface, which is fronted.

While ineffability has been construed by some as evidence for 'clash and crash' syntax and inviolable constraints (Pesetsky, 1997) it alternatively points to a genuine role for input-output faithfulness in syntax. Given the input-output architecture, input specifications are only *target* ones, they will be realized only if all faithfulness constraints can be satisfied. In the case of multiple questions, Legendre et al. (1995, 1998) argue in favor of a neutralization analysis schematized in (5), in which different inputs (interpretations) neutralize to one and the same optimal output because specific input features (e. g. [Q]) may be underparsed, yielding structures that are unfaithful to the input under compulsion of one or more higher-ranked constraints. The optimal candidate is close to the input interpretation but not identical, e. g., 'who said something? What was it?' (see Legendre, 2009 for a full analysis). In fact, wh-phrases share a lexical form with indefinite quantifiers in some languages (e. g. Mandarin) but the identity in form is not a guarantee in a given language of input neutralization in multiple wh-structures.

(5) The neutralization approach to ineffability in multiple wh-questions

Input	Optimal output
[Q] (single wh-question)	[Q]
[QQ] (multiple wh-question)	[Q]

OT constraints belong to constraint *families* which instantiate two types of possible interactions between constraints. One type consists of families whose member constraints are not universally ranked with respect to one another; they simply belong to the same family by virtue of their content. Freely rankable constraints in syntax include faithfulness constraints (e. g., PARSE/MAX and FILL/DEP), and markedness constraints such as economy of movement (Grimshaw, 1997; Legendre et al., 1995, 1998), structural constraints respectively governing the appearance of expletive subjects and *do* support (SUBJECT, OBLIGATORY HEADS, Grimshaw, 1997), etc. This is, by far, the most common constraint family type found in OT analyses.

The other type of constraint family consists of sub-hierarchies within a single constraint family whose relative ranking is universally fixed and is characteristic of implicational generalizations in (morpho)-syntax grounded e. g., in animacy distinctions (Human > Animate > Inanimate) or definiteness (Personal pronoun > Proper name > Definite NP > Indefinite specific NP > Non-specific NP). As a result, sub-hierarchies tend to comprise markedness constraints, e. g., the MINLINK family of constraints against long movement proposed in Legendre et al. (1995, 1998), the OPSPEC family of Bakovic (1998), and various prominence hierarchies (e. g., Burzio, 1998; Aissen, 1999, 2003; P. de Swart, 2007). To illustrate with an example outside of the traditional domain of animacy- and definiteness-based implicational generalizations (see Aissen, 1999, 2003 for such examples), Spanish dialects are known to vary as to what type of wh-phrase triggers inversion (i. e. head movement to C) in wh-questions along two separate dimensions (argument vs. adjunct, matrix vs. subordinate clause). In one dialect (e. g., Spanish$_F$, Bakovic, 1998), inversion is required with arguments but not adjuncts in matrix clauses, and with neither in subordinate clauses:

(6) a. *Qué se comió Miguel?*
what ate-3sg Miguel
'What did Miguel eat?'

b. **Qué Miguel se comió?*

c. *Dónde Miguel se fue?*
where Miguel went-3sg
'Where did Miguel go?'

d. *Me pregunto qué Miguel se comió.*
wonder-1sg what Miguel ate-3sg
'I wonder what Miguel ate'

e. *Me pregunto dónde Miguel se fue.*
wonder-1sg where Miguel went-3sg
'I wonder where Miguel went'

Focusing on the argument/adjunct dimension, OPSPEC, a general markedness constraint requiring wh-operators to be in specifier position (Grimshaw, 1997), is individualized for each (non)-argumental type of wh-phrase: core argument, location, manner, reason. These individualized OPSPEC constraints are universally ranked with respect to one another, yielding a *markedness sub-hierarchy*: ARGOPSPEC » LOCOPSPEC »

ManOpSpec » ReasOpSpec. That is, this ranking is fixed and present in all languages. A conflicting economy of movement constraint like Stay (which penalizes wh-fronting and head movement) can be ranked anywhere in the fixed OpSpec sub-hierarchy. All wh-operators of the type whose OpSpec constraint is ranked below Stay will be fronted, because of a scope requirement (OpScope), but to an adjoined, rather than to a specifier position, so as to minimally violate Stay (induced by head movement). Given five possible ways of ranking Stay in the markedness sub-hierarchy, Bakovic's analysis predicts five different grammars or dialects (abstracting away from the additional matrix/subordinate distinction that is included in his analysis). The question of what defines the constraints of a sub-hierarchy in terms of type or content is at present an open question. They are, however, restricted to markedness constraints. Most seem to pertain to cognitively salient categories, including referentiality distinctions (Bakovic, 1998), person/animacy distinctions (Aissen, 1999, 2003), and possibly processing-related constraints such as short movement (Legendre et al., 1995, 1998).

OT syntax aims at describing cross-linguistic variation and explaining it in terms of different resolutions of conflicts among constraints that govern a particular linguistic phenomenon. Formally, grammars of different dialects, languages, stages of early linguistic development, or languages undergoing diachronic change, all rely on different rankings of the same set of relevant universal constraints. But the first step in an optimality-theoretic analysis, which separates it from many other approaches, is to identify the *conflict and trade-offs* that might be giving rise to a particular surface structure, and reason through a hypothesized set of conflicting constraints. OT makes the strong claim that languages cannot differ in their well-formedness criteria but only in which criteria have priority in cases of conflict. Two constraints conflict when satisfying one entails violating the other. The rest of the chapter focuses on illustrating this and other distinctive properties of the theory, one by one, with later comments building on earlier ones. Given the competition-based approach to evaluating structures in OT syntax, it is simply not possible to analyze the sample sentence proposed by the editors of this volume without a full investigation of related structures, which would have expanded the paper much beyond the length limit given; a different strategy is therefore adopted below.

A well-known case of constraint conflict which fosters further discussion of constraint universality under re-ranking concerns the existence of expletive subject *it* in English, with no counterpart in some languages, including Italian.

(7) a. It rains.
    b. Piove.

The relevant constraints in (8) have been slightly modified from Grimshaw and Samek-Lodovici (1998) without altering their basic content. They capture the core ideas of the EPP (Extended Projection Principle, Chomsky, 1982) and the Principle of Full Interpretation (Chomsky, 1991), respectively. Subject is a markedness constraint contributing to the general well-formedness of sentences while Full Interpretation is an input-output faithfulness constraint (equivalent to FILL 'no epenthesis of elements not present in the input'; Prince and Smolensky, [1993] 2004), which bans expletive elements, including subject *it*, *do* support, etc. Both are constraints active in other syntactic domains (Grimshaw, 1997).

(8) a. SUBJECT: The subject surfaces in SpecTP.
b. FULL-INT(ERPRETATION): Lexical items contribute to the interpretation of a structure.

The input to syntactic optimization does not include expletive elements devoid of semantic content. Rather, expletive subject *it* is added by *Gen* to a candidate output in order to satisfy the markedness constraint SUBJECT at the cost of making the candidate output different from the input. The two constraints conflict in the case of weather verbs because the latter do not select for a thematic argument. One option is for a weather verb to surface without a subject, in which case it satisfies FULL-INT (the corresponding structure [*rains*] contains no lexical item which does not fully contribute to its interpretation). This, however, entails a necessary violation of SUBJECT (the structure is specified as a TP whose specifier is not filled). Alternatively, a weather verb may satisfy SUBJECT by surfacing with an expletive subject. This, however, entails a necessary violation of FULL-INT, as expletive *it* does not contribute to the interpretation of the structure. Thus, there is no possible output which satisfies both constraints. The conflict is resolved by hierarchically ranking the constraints.

The input to optimization includes the weather verb and a tense specification, but no argument. The two options — surfacing with a subject and surfacing without — constitute the candidate set of structures to be evaluated by the constraints in *Con* {SUBJECT, FULL-INT}. If FULL-INT outranks SUBJECT, as it does in Table 3, it is less important to satisfy the lower-ranked constraint, SUBJECT, than the higher-ranked one, FULL-INT. The result is that candidate b is better or more harmonic than candidate a with respect to the constraint ranking in Table 3. Hence candidate b emerges from the comparative evaluation as the *optimal* candidate and is thus the only one to be grammatical in Italian.

On this analysis, the grammatical structure, *piove*, ends up violating the universal SUBJECT constraint (its universal status is further examined below). Violations are relative, however, and the one incurred by optimal candidate b is *only* of the lower-ranked constraint. Hence, it is a *minimal* violation. Note that unlike in Table 3 SUBJECT plays a decisive role when the syntax of both null and overt subjects in Italian

Table 3: Italian weather verbs (Input: *piovere*$_V$ [present]).

	FULL-INT	SUBJ
a. EXPL piove	*!	
☞ b. Piove		*

and their relevant information structure properties is compared (see Samek-Lodovici, 1996, Grimshaw & Samek-Lodovici, 1998 for details).

If SUBJECT outranks FULL-INT, the structure containing an expletive subject will be optimal, as is the case in English. As shown in Table 4, the minimal violation incurred by the optimal candidate a is of lower-ranked FULL-INT. Its competitor, b, fares worse because it violates the higher-ranked constraint, SUBJECT. Thus, the different English and Italian patterns result from the two logically possible rankings of the two constraints.

Table 4: English weather verbs (Input: *rain*$_V$ [present]).

	SUBJ	FULL-INT
☞ a. EXPL rains		*
b. Rains	*!	

This cross-linguistic analysis of expletives differs from the traditional PPT analysis in another respect. Whether a given language has expletive subjects or not depends on the relative ranking of SUBJECT and FULL-INT. In other words, it is not the case that some languages have an expletive subject in their lexicon while others don't. It is the syntax that determines whether some universally available element should be drafted by *Gen* to serve as an expletive. What form the expletive takes in a given language is a language-particular issue (see Grimshaw, 1997 for discussion).

The OT analysis of expletive subjects in English and their absence in Italian, based on constraint interaction and minimal constraint violations, in turn contributes to explaining the fact that German has expletive subjects that do not systematically surface. This is particularly the case in impersonal passives lacking a thematic subject, as revealed in the context of question-answer pairs. If the question is a general one about the event (9a), only an answer with *es* (9b), is natural; (9c) is unnatural on a non-presuppositional reading.

(9) a. Was geschah?
‘what happened?’

b. Es wurde schön getanzt.
it was beautifully danced

c. #Schön wurde getanzt.
beautifully was danced

d. *Schön wurde es getanzt.
beautifully was it danced

If, however, the question is about a property (or location) of the dancing itself, then the pattern is a bit more complicated, giving rise to an apparent case of optionality. If the answer simply conveys new information, then (10b) with *es*, is the natural answer. The only difference with (9b) lies in additional stress falling on the adverb *schön* in (10b), represented in caps (prosody-related constraints governing focalization are not included in this discussion of expletive *es*). If the adverb conveys information that is not only new but noteworthy hence unexpected, then the natural answer is (10c) with additional stress on the fronted adverb (native speakers report that (10b) and (10c) are not instances of contrastive stress which requires heavier stressing than new information does). In any discourse situation, the answer with *es* plus fronted adverb (9d–10d) is ungrammatical. Overall, the intuitive explanation for the distribution of expletive *es* is that it surfaces only where absolutely necessary.

(10) a. *Wie wurde getanzt?*
how was danced
'How was the dancing?'

b. *Es wurde SCHÖN getantzt.*
it was beautifully danced

c. *SCHÖN wurde getanzt.*
beautifully was danced

d. **SCHÖN wurde es getanzt.*
beautifully was it danced

The source of the contrast can be attributed to the input by assuming that it encodes information structure features like [new], [noteworthy], etc. (e. g., Choi, 1996; Samek-Lodovici, 1996, 1998). We can further assume, along with the references mentioned above, that elements focalized by virtue of encoding these input features are typically subject to alignment (McCarthy and Prince, 1993a,b) with the edge of a particular syntactic (or prosodic) domain. Because German treats 'new' information differently from 'new and noteworthy' information, it is necessary to assume two (universal) constraints — ALIGN-NEW and ALIGN-NOTEWORTHY which align the focalized element with the left edge of VP and the clause, respectively. Alignment interacts with the two constraints discussed earlier, SUBJECT and FULL-INT. SUBJECT is violated whenever the adverb *schön* (in the absence of subject) fills the relevant position (SpecTP).

To see how the German competitions play out, consider initially two of the relevant inputs in German. When *schön* conveys new and noteworthy information as specified in the input, *SCHÖN wurde getanzt* is optimal. That is, it is less costly to violate SUBJECT than the alignment constraints (see Table 5 below). But when *schön* is not focalized (i. e. no information structure feature is present in the input), the alignment constraints are vacuously satisfied; SUBJECT requires that an expletive subject surface in SpecTP at the cost of violating FULL-INT. The result is *Es wurde schön getanzt* (see Table 7 below). In neither competition can a candidate with both fronting and expletive *es*, **Schön wurde es getanzt*, emerge as a winner. The reason is economy: Either

Table 5: German impersonal passives (Input: *tanzen*$_V$ (x); x = 0; [past]; *schön*$_{Adv}$ [new]).

	AL-NEW$_{\text{VP}}$	*STRUC	SUBJ	FULL-INT
☞ a. [$_{\text{TP}}$ Es wurde [$_{\text{VP}}$ SCHÖN getanzt]]				*
b. [$_{\text{TP}}$ Wurde [$_{\text{VP}}$ SCHÖN getanzt]]			*!	
c. [$_{\text{TP}}$ SCHÖN$_j$ wurde [$_{\text{VP}}$ t$_j$ getanzt]]	*!		*	
d. [$_{\text{CP}}$ SCHÖN$_j$ wurde$_i$ [$_{\text{TP}}$ es t$_i$ [$_{\text{VP}}$ t$_j$ getanzt]]]	*!	*		*
e. [$_{\text{CP}}$ SCHÖN$_j$ wurde$_i$ [$_{\text{TP}}$ t$_i$ [$_{\text{VP}}$ t$_j$ getanzt]]]	*!	*	*	

fronting or the adverb or expletive *es* is enough to produce an optimal candidate. Doing both is overkill. Its formal account relies on economy of structure. Simplifying, only *STRUCTURE violations incurred by CPs are recorded here. Table 5 represents the competition for an input in which *schön* carries the input feature [new].

A TP structure is all that is needed to best satisfy the constraint ranking in Table 5. As we saw earlier in English, expletive subjects result from the basic ranking: SUBJECT » FULL-INT, which also eliminates candidate b. Candidate c is eliminated because it violates ALIGN-NEW: *schön* is aligned with the left edge of TP, not VP. Candidates d and e violate ALIGN-NEW and Economy (*STRUCTURE), hence they also are eliminated. Only candidate a incurs a minimal violation (FULL-INT), hence it is optimal and grammatical.

The fact that German has expletive subjects does not, however, entail that the expletive subject structure is always optimal. This is shown in Table 6 where *schön* carries two input features [new] and [noteworthy].

Table 6: German impersonal passives (Input: *tanzen*$_V$ (x); x = 0; [past]; *schön*$_{Adv}$ [new],[noteworthy]).

	AL-NOTEW$_{\text{CLAUSE}}$	AL-NEW$_{\text{VP}}$	*STRUC	SUBJ	FULL-INT
a. [$_{\text{TP}}$ Es wurde [$_{\text{VP}}$ SCHÖN getanzt]]	*!				*
b. [$_{\text{TP}}$ Wurde [$_{\text{VP}}$ SCHÖN getanzt]]	*!			*	
☞ c. [$_{\text{TP}}$ SCHÖN$_j$ wurde [$_{\text{VP}}$ t$_j$ getanzt]]		*		*	
d. [$_{\text{CP}}$ SCHÖN$_j$ wurde$_i$ [$_{\text{TP}}$ es t$_i$ [$_{\text{VP}}$ t$_j$ getanzt]]]		*	*!		*
e. [$_{\text{CP}}$ SCHÖN$_j$ wurde$_i$ [$_{\text{TP}}$ t$_i$ [$_{\text{VP}}$ t$_j$ getanzt]]]		*	*!	*	

Assuming that each feature is sensitive to its own alignment requirement, candidate c is optimal if ALIGN-NOTEWORTHY outranks ALIGN-NEW. In fact, candidate c is optimal in Table 6 despite the absence of an expletive subject because all its competitors fare worse. TP structures in which *schön* is not fronted (candidates a and b) fatally violate ALIGN-NOTEWORTHY while other competitors succumb to *STRUCTURE (candidates d and e). Note that the optimal candidate c in Table 6 violates SUBJECT, resulting in the focalized adverb appearing in SpecTP, with no violation of *STRUCTURE.

The nature of the competition changes again for an input in which *schön* does not carry any information structure feature (Table 7). ALIGN-NEW and ALIGN-NOTEWORTHY are vacuously satisfied (v. s.). The effect of *STRUCTURE and SUBJECT becomes visible as they eliminate sub-optimal candidates. Once more, nothing is gained from the addition of structure.

Table 7: German impersonal passives (Input: *tanzen*$_V$(x); x = 0; [past]; schön$_{Adv}$).

	AL-NOTEW$_{CLAUSE}$	AL-NEW$_{VP}$	*STRUC	SUBJ	FULL-INT
☞ a. [$_{TP}$ Es wurde [$_{VP}$ SCHÖN getanzt]]	v. s.				*
b. [$_{TP}$ Wurde [$_{VP}$ SCHÖN getanzt]]	v. s.			*!	
c. [$_{TP}$ SCHÖN$_j$ wurde [$_{VP}$ t$_j$ getanzt]]		v. s.		*!	
d. [$_{CP}$ SCHÖN$_j$ wurde$_i$ [$_{TP}$ es t$_i$ [$_{VP}$ t$_j$ getanzt]]]		v. s.	*!		*
e. [$_{CP}$ SCHÖN$_j$ wurde$_i$ [$_{TP}$ t$_i$ [$_{VP}$ t$_j$ getanzt]]]		v. s.	*!	*	

SpecTP can be filled, either with expletive *es* (candidate a) or adverbial *schön* (candidate c). The difference is a violation of SUBJECT, fatally incurred by *schön*. Candidate c —optimal in Table 6— now loses, given the ranking SUBJECT » FULL-INT established earlier. Candidate a, in fact, beats all its competitors, which all violate a constraint that outranks FULL-INT. The outcome of the competition in Table 7 is an instance of the *Emergence of the Unmarked* (McCarthy and Prince, 1994). The unmarked pattern emerges from the effect of low-ranked constraints like SUBJECT when dominating constraints are controlled for (e. g., when the latter are vacuously satisfied). In other words, OT only *demotes* constraints rather than turns them off. In many competitions such demotion will render the relevant constraint inactive but in some (e. g. Table 7) a demoted constraint is crucial to eliminating sub-optimal candidates. A maximally simple and elegant predictive theory can be achieved by treating all constraints as present in every grammar, regardless of ranking.

The comparative discussion of Tableaux 5–7 highlights the fact that the outcome of each competition is indirectly determined by the input. If one adds or removes an input feature, the nature of the competition changes because the input determines which of the constraints are applicable. This was illustrated above with the feature [new] and [noteworthy] activating alignment constraints in two related competitions (out of three).

As stated above, the content of an OT constraint is not responsible for ensuring its universal application. All constraints are universal in the sense that they are present in every language-particular ranking but their relative priority in a given language will be determined by the ranking itself. Thus, all constraints invoked in the tableaux above are equally present in the grammars of English, Italian, and German but their effect is different because their relative rankings vary with respect to one another as well as with respect to other constraints they interact with. Constraint universality is thus logically independent of constraint violability in OT. In comparison, PPT relies on parameters to handle cross-linguistic variation. In particular, having an overt expletive pronoun in a language's lexicon is subject to parametrization. English and German have one, Italian does not. Yet, German does not systematically deploy expletive *es* to satisfy SUBJECT, contrary to what is expected under a system of parameters fixed once and for all in a given language. Inviolable principles and parameters seem to necessarily require ad-hoc solutions, some of which easily come to mind: i) posit a little pro in SpecTP (with *schön* in SpecCP) in grammatical candidate (10c) and ensure the inviolability of SUBJECT— but German does not allow null subjects, ii) stipulate that *es* may only occur in SpecCP position in German in order to eliminate ungrammatical candidate (10d), etc. A mixed pattern, however, is what is to be expected under a system of universal constraints that are activated by features of the input, are ranked for priority in a given language, and can be violated by grammatical structures. From an OT point of view, the occurrence of *es* is tied to a particular constraint interaction that is determined by a particular input. Hence, its occurrence is *context-sensitive*. Finally, comparison between the competitions in Tableaux 5–7 highlights the most fundamental property of constraint violability in OT. Within a single language, the same constraint can be both violated by a grammatical structure in one context and fatal to an ungrammatical structure in another context. For example, SUBJECT is both violated by a grammatical structure (optimal candidate c in Table 6) and fatal to ungrammatical ones (candidates b in Table 5 and b, c in Table 7). ALIGN-NEW is violated by the optimal candidate c in Table 6 and fatally so by sub-optimal candidates (c, d, e) in Table 5. Because the theory predicts that constraints will sometimes be violated in surface forms, such a violation is therefore not an "inconsistency" within the context of OT (cf. Rákosi, 2014).

The core of OT syntax illustrated above instantiates conflict resolution whereby one constraint takes priority over another (independently of its markedness or faithfulness status). A second logical type of conflict resolution can be identified, which amounts to finding a compromise between two conflicting constraints, one of which being gradiently violable, e. g., in the positioning of Balkan clitics subject to Mussafia-Tobler effects (Legendre, 2000a,b). A third logical type of conflict resolution involves 'walking away from it all', with neither of two conflicting constraints opted for. A particular analysis of ineffability, known as the Null Parse analysis (Prince and Smolensky [1993] 2004) appears to instantiate this possibility; see Legendre (2009) for discussion.

4 Evaluation

OT as a theory of UG shares the goal set by Chomsky (1965) for generative syntax, in particular achievement of descriptive and explanatory adequacy. First, OT has all the tools needed to provide a descriptively adequate grammar (both forms and interpretations) of every human language. From a comparative perspective variation constitutes the core of OT while it requires a special type of principle —parametrized ones— in PPT, which is built around the concept of cross-linguistic uniformity of structure. Second, OT meets the criterion of explanatory adequacy by being computationally powerful enough but sufficiently constrained to enable efficient and general learning and processing algorithms (Tesar and Smolensky, 1998, et seq.; Ellison, 1994; Tesar, 1994 et seq.; Frank and Satta, 1998; Karttunen, 1998).

OT has a number of properties which make it a parsimonious approach to syntax, and to the entire grammar encompassing multiple components. With respect to the former, violability makes it possible for constraints to be logically simple. Their content is not responsible for ensuring that they are satisfied by every grammatical structure in every language. One important consequence is the reduction in the level of abstractness of syntactic representations that is tied to constraint inviolability in PPT. Moreover, syntax is just one component of the grammar subject to optimization along with all the others. Under OT, the language faculty has a unified architecture, which is desirable from a general cognitive perspective.

Note that every optimal output is a last resort or the last candidate standing (Grimshaw, 2013). Last resort follows from the very architecture of the theory — competition and constraint interaction. It is never an added principle but rather a theorem of OT.

Some confusion lingers as to the predicted set of grammars in OT, possibly because the effect of *harmonic bounding* on an OT candidate set is not sufficiently appreciated. While it is true that *Gen* takes an input and *in principle* could return a large candidate set, the actual comparative evaluation operates on non-harmonically bounded candidates only. Any candidate which incurs a superset of the violations incurred by a competitor is a harmonically bounded candidate. It can never be optimal under any ranking hence it is universally dis-preferred and is effectively not part of the competition. Examples of harmonically bounded candidates can be found in the sample tableaux discussed above, which were used only to introduce concepts central to OT one-by-one. For example, candidates c and e are harmonically bounded by b in Table 5. Such candidates are not genuine competitors. Another illustration comes from observing that having more structure (e. g. CP) may offset constraint violations (e. g. of a constraint requiring wh-phrases to surface in specCP, due to their operator status) but beyond trying to beat competitors with respect to satisfying a finite set of constraints, adding further structure only results in worse competitors.

One not so infrequent (but unfair) criticism that is leveled against OT syntax is the erroneous claim that the number of constraints is unconstrained (“If you need an additional one to reign in a competition among candidates you just invent one or two or three … and move on”). This could not be further from the truth in a theory of typology. One important consequence of constraint universality pertaining to the more common freely-rankable constraint type is that any new constraint C_n invented on the basis of some phenomenon in, say, German will need to be present in all languages. That is, positing a new constraint in a hierarchy affects the grammar of the particular language but also affects the analyses of all languages given that constraints are held to be universal. Further, to propose a constraint ranking for one language is to claim that all possible re-rankings of those constraints yield *all and only* possible human languages (this is known as a *factorial typology*). Some examples include a factorial typology of basic case marking/grammatical voice systems (Legendre et al., 1993), presentational focus (Samek-Lodovici, 2001), auxiliary selection related to unaccusativity (Legendre, 2007a,b), and person-based auxiliary selection in Italo-Romance dialects (Legendre, 2010).

Finally, 'rankings' should not be confused with 'grammars' or 'languages'. As a general illustration, for a set of N constraints, there are N! (N factorial) rankings, not N! languages, because there are very few candidate outputs to choose from, due to harmonic bounding. As a consequence, many alternative rankings produce the same optimal outputs, considerably reducing the number of grammars/languages predicted. Concretely, Legendre et al. (1993) propose a set of 8 constraints governing the mapping between thematic roles and their morphosyntactic realizations. This effectively predicts about 40.000 rankings (note that deducing the typology predicted by constraints does not involve anything resembling examining all possible rankings.) Yet, as the paper demonstrates, the typology contains 13 possible language types only. Computational tools for determining typologies have been developed, which automate the task for the linguist (e. g. Hayes, 1998; Prince et al., 2016). See the Rutgers Optimality Archive (https://roa.rutgers.edu) for additional OT software and wide-range dissemination of OT papers.

5 Conclusion

Many current theories of grammar claim to be cognitively valid theories but that assertion has different content for different communities, ranging from a focus on the knowledge system and its domain-specific representations to positing psychologically real or domain-general operations to focusing on fundamental aspects of computation in the mind/brain. OT and related theories are mostly concerned with the latter. To the best of our knowledge, they are unique within linguistic theory in their attempt at reconciling two realities that may at first appear to be incompatible: symbolic, largely domain-specific representations of the traditional sort (e. g., abstract linguistic categories, grammatical features, prominence hierarchies, etc.) that are characteristic of natural language systems and form the centerpiece of generative grammar approaches and beyond on the one hand, and on the other, the consensus view in Cognitive Science circles that the mind/brain is a computer performing continuous processing including optimization. OT arguably is rather successful at meeting the challenge by providing a formal competition-based theory of constraint interaction that crosscuts all components of the grammar and places variation at the core of the inquiry, thereby enriching our understanding of how linguistic forms and interpretations are synchronically and diachronically constrained in diverse languages, and throughout the process of language acquisition.

References

Aissen, Judith. 1999. Markedness and Subject Choice in Optimality Theory. *Natural Language and Linguistic Theory* 17. 673–711.

Aissen, Judith. 2003. Differential Object Marking: Iconicity vs. Economy. *Natural Language and Linguistic Theory* 21. 435–483.

Anderson, John R. 1990. *The Adaptive Character of Thought*. Hillsdale, NJ: Lawrence Erlbaum.

Antilla, Arto. 1997. Variation in Finnish Phonology and Morphology. Stanford University, dissertation.

Bakovic, Eric. 1998. Optimality and Inversion in Spanish. In Pilar Barbosa, Danny Fox, Paul Hagstrom, Martha McGinnis & David Pesetsky (eds.), *Is the Best Good Enough? Optimality and Competition in Syntax*, 35–58. Cambridge, MA: MIT Press.

Beaver, David. 2004. The optimization of discourse anaphora. *Linguistics and Philosophy* 27. 3–56.

Beaver, David & Hanjung Lee. 2004. Input-output mismatches in OT. In Reinhart Blutner & Henk Zeevat (eds.), *Optimality Theory and Pragmatics*, 112–153. Houndmills, Basingstoke, Hampshire: Palgrave/Macmillan.

Bhatt, Rakesh M. 1997. Code-switching, Constraints, and Optimal Grammars. *Lingua* 102. 223–251.

Bhatt, Rakesh M. 2014. Argument Licensing in Optimal Switches. In Jeff Macswan (ed.), *Grammatical Theory and Bilingual Code-switching*, 135–158. Cambridge, MA: MIT Press.

Blutner, Reinhard. 2000. Some aspects of optimality in natural language interpretation. *Journal of Semantics* 17. 189–216.

Blutner, Reinhard, Helen de Hoop & Petra Hendriks. 2006. *Optimal Communication*. Stanford, CA: CSLI Publications.

Blutner, Reinhard & Henk Zeevat. 2004. *Optimality Theory and Pragmatics*. Houndmills, Basingstoke, Hampshire: Palgrave/Macmillan.

Blutner, Reinhard & Henk Zeevat. 2009. Optimality-theoretic pragmatics. *ZAS working papers in Linguistics* 51. 1–25.

Boersma, Paul & Bruce Hayes. 2001. Empirical tests of the gradual learning algorithm. *Linguistic Inquiry* 32(1). 45–86.

Bousquette, Joshua, Michael Putnam, Joseph Salmons, Ben Freyd & Daniel Nützel. 2016. Multiple grammars, dominance, and optimization. In Géraldine Legendre, Michael Putnam, Henriette de Swart & Erin Zaroukian (eds.), *Optimality-theoretic syntax, semantics, and pragmatics: From uni- to bidirectional optimization*, 158–176. Oxford: Oxford University Press.

Bresnan, Joan. 2000. Optimal syntax. In Joost Dekkers, Frank van der Leeuw & Jeroen van de Weijer (eds.), *Optimality theory: phonology, syntax and acquisition*, 334–385. Oxford: Oxford University Press.

Bresnan, Joan & Tatiana Nikitina. 2009. The gradience of the dative alternation. In Linda Uyechi & Lian Hee Wee (eds.), *Reality exploration and discovery: Pattern interaction in language and life*, 161–184. Stanford, CA: CSLI Publications.

Burzio, Luigi. 1998. Anaphora and soft constraints. In Pilar Barbosa, Danny Fox, Paul Hagstrom, Martha McGinnis & David Pesetsky (eds.), *Is the Best Good Enough? Optimality and Competition in Syntax*, 93–113. Cambridge, MA: MIT Press.

Burzio, Luigi. 2010. Constraint violability and the Chain Condition. *Revista Virtual de Estudos da Linguagem (ReVEL)* 8(4). [www.revel.inf.br/eng].

Choi, Hye-Won. 1996. Optimizing Structure in Context: Scrambling and Information Structure. Stanford University, dissertation.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chomsky, N. 1982. *Some Concepts and Consequences of the Theory of Government and Binding*. Cambridge, MA: MIT Press.

Chomsky, Noam. 1991. Some notes on economy of derivation and representation. In Robert Freidin (ed.), *Principles and Parameters in Comparative Grammar*, 41–58. Cambridge, MA: MIT press.

Chomsky, Noam. 1993. A minimalist program for linguistic theory. In Ken Hale & Samuel Jay Keyser (eds.), *The View from Building 20*, 1–52. Cambridge, MA: MIT Press.

Chomsky, Noam. 2000. Minimalist inquiries: the framework. In Roger Martin, David Michaels, Juan Uriagereka & Samuel Jay Keyser (eds.), *Step by step: Essays on minimalist syntax in honor of Howard Lasnik*, 89–155. Cambridge, MA: MIT Press.

Chomsky, Noam. 2001. Derivation by phase. In Michael Kenstowicz (ed.), *Ken Hale: a life in language*, 1–54. Cambridge, MA: MIT Press.

de Swart, Peter. 2007. *Cross-Linguistic Variation in Object Marking*. LOT Publications.

Diesing, Molly. 1996. Semantic variables and object shift. *Studies in Comparative Germanic Syntax* 2. 66–84.

Ellison, T. Mark. 1994. Phonological derivation in Optimality Theory. *Proceedings of the 15th International Conference on Computational Linguistics, Volume 2*, 1007–1013.

Frank, Robert & Giorgio Satta. 1998. Optimality Theory and the generative complexity of constraint violability. *Computational Linguistics* 24. 307–315.

Geman, Stuart & Donald Geman. 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6. 721–741.

Gigerenzer, Gerd & Daniel G. Golstein. 1999. Betting on one good reason: The take the best heuristic. In Gerd Gigerenzer, Peter M. Todd & the ABC Research Group (eds.), *Simple Heuristics That Make Us Smart*, 75–95. Oxford: Oxford University Press.

Grimshaw, Jane. 1997. Projection, heads and optimality. *Linguistic Inquiry* 28. 373–422.

Grimshaw, Jane & Vieri Samek-Lodovici. 1998. Optimal Subjects and Subject Universals. In Pilar Barbosa, Danny Fox, Paul Hagstrom, Martha McGinnis & David Pesetsky (eds.), *Is the Best Good Enough? Optimality and Competition in Syntax*, 193–219. Cambridge, MA: MIT Press.

Grimshaw, Jane. 2013. Last Resorts: A Typology of do-support. In Hans Broekhuis & Ralf Vogel (eds.), *Linguistic derivations and filtering: Minimalism and Optimality Theory*, 267–295. London: Equinox.

Hale, John & Géraldine Legendre. 2004. Minimal Links, Remnant Movement, and (Non)-derivational Grammar. In Arthur Stepanov, Gisbert Fanselow & Ralf Vogel (eds.), *Minimality Effects in Syntax*, 177–203. Berlin: Mouton de Gruyter.

Hayes, Bruce. 1998. *Optimality Theory Software*. Available at: http://www.humnet.ucla.edu/humnet/linguistics/people/hayes/otsoft/otsoft.htm.

Heck, Fabian. 2000. Quantifier scope in German and Cyclic Optimization. In Gereon Müller & Wolfgang Sternefeld (eds.), *Competition in syntax* (49), 175–209. Walter de Gruyter.

Heck, Fabian & Gereon Müller. 2013. Extremely local optimization. In Hans Broekhuis & Ralf Vogel (eds.), *Linguistic Derivations and Filtering: Minimalism and Optimality Theory*, 136–165. London: Equinox.

Heck, Fabian & Gereon Müller. 2016. On accelerating and decelerating movement: From minimalist preference principles to harmonic serialism. In Géraldine Legendre, Michael Putnam, Henriette de Swart & Erin Zaroukian (eds.), *Optimality-theoretic syntax, semantics, and pragmatics: From uni- to bidirectional optimization*, 78–110. Oxford: Oxford University Press.

Hendriks, Petra & Helen de Hoop. 1997. On the interpretation of semantic relations in the absence of syntactic structure. In P. Dekker, M. Stokhof & Y. Venema (eds.), *Proceedings of the 11th Amsterdam Colloquium*, 157–162. Amsterdam: ILLC, University of Amsterdam.

Hendriks, Petra & Helen de Hoop. 2001. Optimality-theoretic semantics. *Linguistics and Philosophy* 24. 1–32.

Hendriks, Petra & Jennifer Spenader. 2005/6. When production precedes comprehension: An optimization approach to the acquisition of pronouns. *Language Acquisition* 13(4). 319–348.

Hendriks, Petra, Helen de Hoop, Irene Krämer, Henriette de Swart & Joost Zwarts. 2010. *Conflicts in Interpretation*. London: Equinox.

Hoeks, John & Petra Hendriks. 2011. Optimality Theory and Human Sentence Processing: Towards a Cross-Modular Analysis of Coordination. *Journal of Cognitive Science* 12. 83–128.

de Hoop, Helen & Henriette de Swart. 2000. Temporal adjunct clauses in Optimality Theory. *Italian Journal of Linguistics* 12(1). 107–127.

de Hoop, Helen & Andrej Malchukov. 2008. Case-marking strategies. *Linguistic Inquiry* 39(4). 565–587.

Hsin, Lisa. 2014. Integrated Bilingual Grammatical Architecture: Insights from Syntactic Development. Johns Hopkins University, dissertation.

Huang, C.-T. J. 1982. Logical Relations in Chinese and the Theory of Grammar. MIT, dissertation.

Jäger, Gerhart. 2004. Learning constraint sub-hierarchies. In Reinhart Blutner & Henk Zeevat (eds.), *Optimality Theory and Pragmatics*, 251–287. Houndmills, Basingstoke, Hampshire: Palgrave/MacMillan.

Karttunen, Lauri. 1998. The proper treatment of optimality in computational phonology. *Proceedings of the International Workshop on Finite-State Methods in Natural Language Processing*. ROA 258.

LaFond, Larry. 2003. Historical changes in verb-second and null subjects from Old to Modern French. In D. Eric Holt (ed.), *Optimality Theory and Language Change*, 387–412. Dordrecht: Kluwer.

Legendre, Géraldine. 2000a. Morphological and Prosodic Alignment of Bulgarian Clitics. In Joost Dekkers, Frank van der Leeuw & Jeroen van de Weijer (eds.), *Optimality Theory: Syntax, Phonology, and Acquisition*, 423–462. Oxford: Oxford University Press.

Legendre, Géraldine. 2000b. Optimal Romanian Clitics: A Cross-linguistic Perspective. In Virginia Motapanyane (ed.), *Comparative Studies in Romanian Syntax* (North Holland Linguistic Series 58), 227–264. Oxford: Elsevier.

Legendre, Géraldine. 2007a. Optimizing Auxiliary Selection in Romance. In Raul Aranovich (ed.), *Split Auxiliary Systems: A Cross-linguistic Perspective*, 145–180. Amsterdam/Philadelphia: John Benjamins.

Legendre, Géraldine. 2007b. On the Typology of Auxiliary Selection. *Lingua* 117(9). 1522–1540.

Legendre, Géraldine. 2009. The Neutralization Approach to Ineffability in Syntax. In Curt Rice & Silvia Blaho (eds.), *Modeling Ungrammaticality in Optimality Theory*, 237–266. London: Equinox.

Legendre, Géraldine. 2010. A Formal Typology of Person-Based Auxiliary Selection in Italo-Romance. In Roberta D'Alessandro, Adam Ledgeway & Ian Roberts (eds.), *Syntactic Variation: The dialects of Italy*, 186–200. Cambridge UK: Cambridge University Press.

Legendre, Géraldine & Mary Schindler. 2010. Code Switching in Urban Wolof: A Case for Violable Constraints in Syntax. *Revista Virtual de Estudos da Linguagem (ReVEL)* 8(4). [www.revel.inf.br/eng].

Legendre, Géraldine, Yoshiro Miyata & Paul Smolensky. 1990. Can connectionism contribute to syntax? Harmonic Grammar, with an application. *Proceedings of the Chicago Linguistics Society*, 237–252.

Legendre, Géraldine, William Raymond & Paul Smolensky. 1993. An Optimality-theoretic Typology of Case and Grammatical Voice Systems. *Proceedings of the Nineteenth Annual Meeting of the Berkeley Linguistic Society*, 464–478.

Legendre, Géraldine, Colin Wilson, Paul Smolensky, Kristin Homer & William Raymond. 1995. Optimality and Wh-Extraction. In Jill Beckman, Laura Walsh Dickey & Suzanne Urbanczyck (eds.), *Papers in Optimality Theory* (University of Massachusetts Occasional Papers, UMOP 18), 607–636. Amherst, MA: G.L.S.A., Dept. of Linguistics, University of Massachusetts.

Legendre, Géraldine, Paul Smolensky & Colin Wilson. 1998. When is Less More? Faithfulness and Minimal Links in Wh-Chains. In Pilar Barbosa, Danny Fox, Paul Hagstrom, Martha McGinnis & David Pesetsky (eds.), *Is the Best Good Enough? Optimality and Competition in Syntax*, 249–289. Cambridge, MA: MIT Press.

Legendre, Géraldine, Sten Vikner & Jane Grimshaw (eds.). 2001. *Optimality-Theoretic Syntax*. Cambridge, MA: MIT Press.

Legendre, Géraldine, Paul Hagstrom, Anne Vainikka & Marina Todorova. 2002. Partial constraint ranking in child French syntax. *Language Acquisition* 10(3). 189–227.

Legendre, Géraldine, Paul Hagstrom, Joan Chen-Main, Liang Tao & Paul Smolensky. 2004. Deriving output probabilities in Child Mandarin from a dual-optimization grammar. *Lingua* 114(9–10). 1147–1185.

Legendre, Géraldine, Michael Putnam, Henriette de Swart & Erin Zaroukian (eds.). 2016a. *Optimality-theoretic syntax, semantics, and pragmatics: From uni- to bidirectional optimization*. Oxford: Oxford University Press.

Legendre, Géraldine, Paul Smolensky & Jennifer Culbertson. 2016b. Blocking effects at the lexicon/semantics interface and bi-directional optimization in French. In Géraldine Legendre, Michael Putnam, Henriette de Swart & Erin Zaroukian (eds.), *Optimality-theoretic syntax, semantics, and pragmatics: From uni- to bidirectional optimization*, 276–299. Oxford: Oxford University Press.

MacWhinney, Brian, Andrej Malchukov & Edith Moravcsik (eds.). 2014. *Competing Motivations in Grammar and Usage*. Oxford: Oxford University Press.

Mattausch, Jason. 2007. Optimality, bidirectionality, and the evolution of binding phenomena. *Research on Language and Computation* 5. 103–131.

McCarthy, John & Alan Prince. 1993a. *Prosodic Morphology I; Constraint Interaction and Satisfaction*. Ms. Center for Cognitive Science. Rutgers University.

McCarthy, John & Alan Prince. 1993b. Generalized Alignment. In *Yearbook of Morphology*, 79–153.

McCarthy, John & Alan Prince. 1994. *The Emergence of the Unmarked. Optimality in Prosodic Morphology*. Ms., UMass and Rutgers University.

McCloskey, Jim. 1979. *Transformational Syntax and Model Theoretic Semantics: A Case Study in Modern Irish*. Dordrecht: Reidel.

Müller, Gereon. 2001. Order Preservation, Parallel Movement, and the Emergence of the Unmarked. In Géraldine Legendre, Sten Vikner & Jane Grimshaw (eds.), *Optimality-theoretic Syntax*, 279–313. Cambridge, MA: MIT Press.

Müller, Gereon & Wolfgang Sternefeld. 2001. The rise of competition in syntax: A synopsis. In Gereon Müller & Wolfgang Sternefeld (eds.), *Competition in syntax*, 1–67. Walter de Gruyter.

Murphy, Andrew. 2017. Cumulativity in Syntactic Derivations. University of Leipzig, dissertation.

Oaksford, Mike & Nick Chater. 1998. *Rational Models of Cognition*. Oxford: Oxford University Press.

Pater, Joe. 2009. Weighted constraints in generative linguistics. *Cognitive Science* 33. 999–1035.

Pater, Joe. 2016. Universal Grammar with weighted constraints. In John McCarthy & Joe Pater (eds.), *Harmonic Grammar and Harmonic Serialism*, 1–46. London: Equinox.

Pesetsky, David. 1997. Optimality Theory and syntax: movement and pronunciation. In Diana Archangeli & D. Terence Langendoen (eds.), *Optimality Theory: An Overview*, 134–170. Oxford: Blackwell-Wiley.

Prince, Alan & Paul Smolensky. [1993]. 2004 *Optimality Theory: Constraint Interaction in Generative Grammar*. Oxford: Blackwell. [Technical Report TR-2, Rutgers Center for Cognitive Science, Rutgers University, New Brunswick, NJ, and Technical Report CU-CS-696-93, Department of Computer Science, University of Colorado at Boulder.]

Prince, Alan, Bruce Tesar & Nazarré Merchant. 2016. *OTWorkplace*. https://sites.google.com/site/otworkplace/.

Rákosi, Csilla. 2014. Data and the resolution of inconsistency in Optimality Theory. In András Kertész & Csilla Rákosi (eds.), *The Evidential Basis of Linguistic Argumentation*, 271–307. Amsterdam/Philadelphia: John Benjamins.

Rieskamp, Jörg & Ulrich Hoffrage. 1999. When do people use simple heuristics, and how can we tell. In Gerd Gigerenzer, Peter M. Todd & The ABC Research Group (eds.), *Evolution and cognition. Simple heuristics that make us smart*, 141–167. New York: Oxford University Press.

Rizzi, Luigi. 1990. *Relativized Minimality*. Cambridge, MA: MIT Press.

Rizzi, Luigi & Ian Roberts. 1989. Complex inversion in French. *Probus* 1(1). 1–30.

Rudin, Catherine. 1988. On multiple questions and multiple *wh*-fronting. *Natural Language and Linguistic Theory* 6. 445–501.

Rumelhart, David E., Geoff E. Hinton & Ronald J. Williams. 1986. Learning internal representations by error propagation. In D. Rumelhart, J. McClelland & the PDP Research Group (eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1*, 318–362. Cambridge, MA: MIT Press.

Samek-Lodovici, Vieri. 1996. Constraints on Subjects: An Optimality Theoretic Analysis. Rutgers University, dissertation.

Samek-Lodovici, Vieri. 1998. Opposite Constraints: Left and Right Focus-alignment in Kanakuru. *Lingua* 104(1–2). 111–130.

Samek-Lodovici, Vieri. 2001. Cross-linguistic typologies in OT. In Géraldine Legendre, Sten Vikner & Jane Grimshaw (eds.), *Optimality-theoretic Syntax*, 315–353. Cambridge, MA: MIT Press.

Samek-Lodovici, Vieri. 2005. Prosody-syntax interactions in the expression of focus. *Natural Language and Linguistic Theory* 23. 687–755.

Samek-Lodovici, Vieri. 2015. *The Interaction of Focus, Givenness, and Prosody. A Study of Italian Clause Structure*. Oxford: Oxford University Press.

Sells, Peter. 2001. *Structure, Alignment and Optimality in Swedish*. Stanford: CSLI.

Slade, Benjamin. 2003. How to rank constraints: Constraint conflict, grammatical competition and the rise of periphrastic *do*. In D. Eric Holt (ed.), *Optimality Theory and Language Change*, 337–385. Dordrecht: Kluwer.

Smolensky, Paul. 1993. Harmony, markedness, and phonological activity. Paper presented at the Rutgers Optimality Workshop 1. New Brunswick, NJ.

Smolensky, Paul. 1995. On the internal structure of the constraint component *Con* of UG. Paper presented at UCLA. Los Angeles, CA.

Smolensky, Paul. 1997. Constraint interaction in generative grammar II: Local conjunction, or random rules in UG. Paper presented at the 1997 Hopkins Optimality Theory Workshop & University of Maryland Mayfest, Baltimore, MD.

Smolensky, Paul & Géraldine Legendre. 2006. *The Harmonic Mind: From Neural Computation to Optimality-theoretic Grammar. 2 Vols.* Cambridge, MA: MIT Press.

Smolensky, Paul & Matthew Goldrick. 2016. Gradient Symbolic Representations in Grammar: The case of French Liaison. Rutgers Optimality Archive 1286.

Speas, Margaret. 1997. Optimality theory and syntax: null pronouns and control. In Diana Archangeli & D. Terence Langendoen (eds.), *Optimality theory: An Overview*, 171–199. Oxford: Blackwell-Wiley.

Speas, Margaret. 2001. Constraints on Null Pronouns. In Géraldine Legendre, Sten Vikner & Jane Grimshaw (eds.), *Optimality-theoretic Syntax*, 393–425. Cambridge, MA: MIT Press.

Stevenson, Suzanne & Paul Smolensky. 2006. Optimality in Sentence Processing. In Paul Smolensky & Géraldine Legendre (eds.), *The Harmonic Mind: From Neural Computation to Optimality-theoretic Grammar, Vol. 2*, 307–338. Cambridge, MA: MIT Press.

de Swart, Henriette. 2010. *Expression and interpretation of negation: an OT typology*. Dordrecht: Springer.

Tesar, Bruce. 1994. Parsing in Optimality Theory: A dynamic programming approach. Technical report CU-CS-714-94. Computer Science Department, University of Colorado at Boulder.

Tesar, Bruce & Paul Smolensky. 1998. Learnability in Optimality Theory. *Linguistic Inquiry* 29(2). 229–268.

Tesar, Bruce & Paul Smolensky. 2000. *Learnability in Optimality Theory*. Cambridge, MA: MIT Press.

Vikner, Sten. 2001. The interpretation of object shift and Optimality Theory. In Gereon Müller & Wolfgang Sternefeld (eds.), *Competition in syntax*, 321–340. Walter de Gruyter.

Vincent, Nigel. 2000. Competition and correspondence in syntactic change: Null arguments in Latin and Romance. In Susan Pintzuk, George Tsoulas & Anthony Warner (eds.), *Diachronic Syntax: Models and Mechanisms*, 25–50. Oxford: Oxford University Press.

Vogel, Ralf. 2016. Optimal Constructions. In Géraldine Legendre, Michael Putnam, Henriette de Swart & Erin Zaroukian (eds.), *Optimality-theoretic syntax, semantics, and pragmatics: From uni- to bidirectional optimization*, 55–77. Oxford: Oxford University Press.

Wexler, Ken. 1994. Optional infinitives, head movement, and economy of derivation. In Norbert Hornstein & David Lightfoot (eds.), *Verb Movement*, 305–350. Cambridge UK: Cambridge University Press.

J. Lachlan Mackenzie

11 The Functional Discourse Grammar approach to syntax

Abstract: Functional Discourse Grammar (FDG) is a theory of the organization of Linguistic Expressions as encoding Discourse Acts. It takes a "form-oriented function-to-form" approach, distinguishing between formulation (yielding the Interpersonal and Representational Levels) and encoding (yielding the Morphosyntactic and Phonological Levels). It seeks to achieve pragmatic, cognitive, and typological adequacy, which together define it as belonging to the functionalist paradigm in linguistics. The FDG approach to syntax cannot be understood without continual reference to the Interpersonal and Representational Levels, whose internal structure combines hierarchical layering and configurations of equipollent elements. In the dynamic implementation of FDG, the layering inherent in the formulation levels impacts the creation of syntactic structure, with the hierarchically related elements having prior access to absolute positions and configurational elements being placed thereafter. Languages are distinguished according as they display a predominance of interpersonal, representational or morphosyntactic alignment. Researchers in FDG use a broad range of data types, according to their particular academic settings, traditions and goals. The model differs from many functionally oriented grammars in employing formalization. A full reasoned analysis is given of the sample sentence devised by the editors of the volume for this purpose.

1 Introduction

Functional Discourse Grammar (FDG) differs from other current approaches to syntax in being a theory of the organization of Linguistic Expressions as encodings of Discourse Acts. Its starting point is the observation that human verbal interaction divides into Discourse Acts, each making its own contribution to the ongoing communicative exchange. Some of these may be encoded as Clauses but others may appear as something less or more than a full Clause. Consider example (1), in which FDG recognizes the expression of three Discourse Acts, each with its own function within the overall unit in which they occur, namely a name with a vocative function, a clause with an imperative function, and a sequence of auxiliary and pronoun with an emphatic function:

(1) *Mary, listen to me, will you?*

The three syntactic units together form a Linguistic Expression. The internal structuring of Linguistic Expressions is dealt with at the Morphosyntactic Level of FDG, which

https://doi.org/10.1515/9783110540253-011

will play a central role in this chapter. However, within FDG, it is only one of four levels and not privileged in any way; FDG is thus not "syntactocentric" in the sense made familiar by Culicover and Jackendoff (2005, 21).

FDG arose in the first years of this century in response to debates about how to continue Simon C. Dik's work on Functional Grammar (FG; see Dik, 1997) after his death in 1995 (Mackenzie, 2016a). It was Kees Hengeveld, Dik's successor in Amsterdam and the leading figure in FDG, who outlined the new theory (Hengeveld, 2004). It was presented in its entirety in Hengeveld and Mackenzie (2008), the standard work to which all subsequent publications in FDG refer; Keizer (2015) is an entry-level textbook which applies FDG to English. The theory has been extended, developed and refined in various ways, some of which will be reviewed below.

This chapter offers more than just another introduction to the theory, although it will be necessary, *en route*, to lay out the architecture of FDG and the operation of its various components, levels and layers. Section 2 will present the model's goals and principles of evaluation, concentrating on the Interpersonal Level. Section 3 turns to various FDG tools used to achieve those goals, focusing first on the relevance of the Representational Level for morphosyntax and then presenting the Morphosyntactic Level and characteristic features of the FDG approach to syntax. After Section 4, which briefly discusses the types of data used in FDG work, Section 5 offers an application of the grammar to the sample sentence proposed by the editors for analysis, and the chapter ends with Conclusions that mention significant points of concern for researchers in FDG.

2 Goals and evaluation: three forms of adequacy

The aim of FDG is to furnish researchers with a coherent set of tools for the analysis and comparison of languages from a functional perspective. To achieve this, FDG takes a "form-oriented function-to-form" approach (Hengeveld and Mackenzie, 2008, 39): i. e., for any language under description, FDG seeks to give a complete account of its morphosyntactic and phonological forms, including their ordering and constituency properties, by positing the communicative functions they fulfil and then tracing a pathway from those functions to the forms. This entails, for example, that if a language lacks any systematic formal marking of absolute tense distinctions (i. e. Past, Present, etc.), the functional description will also lack any mention of tense; conversely, if a language does mark tense distinctions systematically, that fact must be registered in the functional description. This is a reflection of a general principle in FDG known as *typological adequacy*. The description of each language should be true to the facts of that language; only then can language comparison be carried out in a valid manner. Universality is never an assumption; at best, it may be a conclusion drawn from empirical work.

The description of the functional side of the function—form equation is theorized in FDG as the operation of *formulation*. Formulation yields two levels of description, the Interpersonal Level and the Representational Level. Together, the make-up of the two levels and the *primitives* on which they draw (i. e. frames, lexemes and operators) specify how a communicative intention and the associated conceptual content are formulated in the language under description. Returning to the case of absolute tense, the instructions for English require speakers to distinguish grammatically between Present and Past but (unlike various other languages, cf. Chafe, 2018, 139–140) not to distinguish grammatically between Past and various degrees of Remote Past. In FDG, this is theorized in terms of two separate components of the overall theory of verbal interaction, a Conceptual Component where the communicative intention and the associated conceptual content are housed, and the Grammatical Component, which contains the operation of formulation. Formulation thus sifts the material coming in from the Conceptual Component and structures it to comply with the requirements of the particular language under description. The output of formulation serves as the input to the other operation carried out in the Grammatical Component, *encoding*. Encoding also yields two levels of description, the Morphosyntactic Level and the Phonological Level, which together account for all the formal properties of the unit under analysis.

The overall structure of FDG's model of verbal interaction is shown in Figure 1. The figure gives pride of place to the Grammatical Component, showing how the operations of formulation and encoding deliver the four Levels, as well as how the grammar is flanked by three other components, the above-mentioned Conceptual Component, but also a Contextual Component (which specifies relevant features of the discursive and situational context) and an Output Component (which converts the Phonological Level description into phonation, writing or gesture). The rectangles feeding into the two operations contain the primitives, i. e. the elements the grammatical rules have to work with. Whereas the general architecture of the grammar is assumed to be applicable to all languages, the primitives are language-specific. This is trivially true of the lexemes and morphemes but also applies to the frames, operators and templates, in the sense that the grammar of each language involves a selection from possible grammatical distinctions. FDG is strongly influenced in its overall architecture by a consensus in the modelling of language production (Levelt, 1989; Konopka and Brown-Schmidt, 2014) but remains a grammar rather than a model of language production. This reflects another goal of FDG, that of achieving *cognitive adequacy* (Butler, 2008a).

Where, then, does the Discourse Act fit into this picture? The Discourse Act is a unit of the interaction between Speaker and Addressee and as such its description appears at the Interpersonal Level. Discourse Acts group into Moves (symbolized as M_n). A Move is defined as either provoking a Reaction (as in (1), repeated here for convenience) or as being a Reaction (imagine Mary's Move in response to (1)):

(1) *Mary, listen to me, will you?*

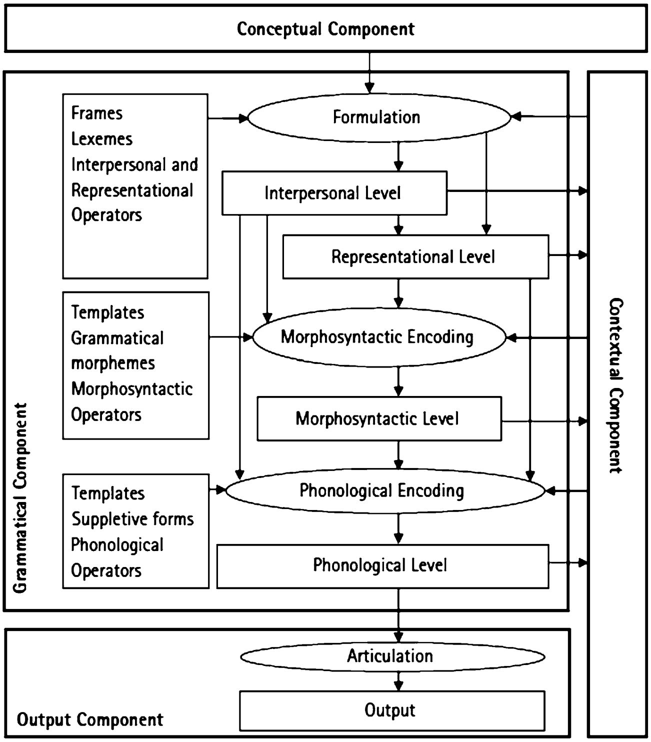


Figure 1: The general layout of FDG.

A Move can consist of a single Discourse Act (symbolized as A_1) or of an – in principle – unlimited number of them; in (1) there are three. The internal structure of Move (1) is shown in (2):[1]

(2) $(M_I: [(A_I), (A_J), (A_K)] (M_I))$

The notation in (2) displays three characteristic, ubiquitous properties of the FDG formalism. The first is *layering*, the nesting of one layer of analysis within a higher one. The Discourse Act layer, represented in (2) by three instances, is nested within the Move layer (here represented as (M_I)). Layering was first proposed within Dik's FG by Hengeveld (1989), drawing inspiration *inter alia* from Foley and Van Valin's (1984) notion of "layered structure", and has become an essential principle of FDG. All four FDG levels are characterized by layering.

The second is *restriction*, the relation represented by the colon (:), which is to be read as "such that". Here the variable M_I is "restricted" to the set of Discourse Acts

1 In the analysis of specific instances, the subscript numerals {1, 2, 3, . . . } are replaced by subscript letters, commencing with I.

{(A_I), (A_J), (A_K)}, and so (2) is to be read as "(M_I . . .) [the occurrence on the left], such that M_I [the occurrence on the right] has the property of being the set {(A_I), (A_J), (A_K) }". The third property exemplified in (2) is *equipollence*, the relation between elements of equal status at the same layer; equipollent elements are placed within a pair of square brackets [. . .]. The various Discourse Acts that compose the Move in (2) are equipollent and thereby form a *configuration*; configurations are treated differently from hierarchical layers when it comes to encoding, as we shall see in Section 3.2 below.

Let us now consider the complete internal structure of the Interpersonal Level, shown in Figure 2, in which, for ease of reading, equipollent elements within the Discourse Act and Communicated Content layers have been arranged vertically rather than horizontally, although in practice they are usually set out horizontally.

(Π M_1:	Move
[(Π A_1: [	Discourse Act
(Π F_1: ILL (F_1): Σ (F_1))	Illocution
(Π P_1: ... (P_1): Σ (P_1))$_S$	Speaker
(Π P_2: ... (P_2): Σ (P_2))$_A$	Addressee
(Π C_1: [	Communicated Content
(Π T_1: [...] (T_1): Σ (T_1)) ... (Π T_n))	Subact of Ascription
(Π R_1: [...] (R_1): Σ (R_1)) ... (Π R_n))	Subact of Reference
] (C_1): Σ (C_1))	Communicated Content
] (A_1): Σ (A_1)) ... (Π A_n))]	Discourse Act
(M_1): Σ (M_1))	Move

Figure 2: The inner structuring of the Interpersonal Level.

It is immediately noticeable that each layer has the same structure, namely (3):

(3) $(\Pi\ v_n\text{: } h\ (v_n)\text{: } \Sigma\ (v_n))$

In (3), h = head, Σ = modifier, Π = operator, while v is for variable, with one of the values {M, A, F, P, C, T, R}. In Figure 2, most of the heads are configurations. A modifier serves to restrict its head: in the case of the Interpersonal Level, typical expressions of modifiers are *briefly* (modifying the head of a Discourse Act), *sincerely* (modifying the head of an Illocution) or *allegedly* (modifying the head of a Communicated Content). While a modifier involves lexical material, an operator has the same functional and scope properties but differs crucially in being encoded by morphological ("grammatical") or phonological material. One aspect of grammaticalization, therefore, involves the historical transfer of units from modifier to operator status (see Hengeveld, 2017). As is apparent in Figure 2, all layers have positions for operators and modifiers.

Figure 2 shows that the head of any Discourse Act (A_1) is a configuration containing up to four elements: an Illocution (F_1), the Speaker (P_1)$_S$, the Addressee (P_2)$_A$, and the Communicated Content (C_1). The Communicated Content, in turn, has as its head a configuration of Subacts. This represents the fact that communication involves the interplay of referential activity (Subacts of Reference (R_1), (R_2), etc.) and of predicational activity (Subacts of Ascription (T_1), (T_2), etc.). In the analysis of the second Discourse Act of (1), the Illocution placeholder ILL is replaced by IMP(erative), and the Communicated Content represents what is imparted, namely a Subact of Reference (R_I) corresponding to *me*, and a Subact of Ascription (T_I) corresponding to *listen* (*to*), cf. (4b). The first Discourse Act, the vocative *Mary*, displays the same general structure, but now without any Communicated Content, since the function is purely geared to drawing the Addressee's attention, cf. (4a). The analysis of the third Discourse Act is more controversial: as a "tag question", its form is clearly dependent on that of the preceding imperative; in actional terms, it serves to deliver an emphatic imperative, the same one (namely (F_J)) as in (A_J). A possible analysis is given in (4c).

(4) a. $(A_I: [(F_I: \text{INTERP } (F_I)) (P_I)_S (P_J: \text{Mary } (P_J))_A] (A_I))$
b. $(A_J: [(F_J: \text{IMP } (F_J)) (P_I)_S (P_J)_A (C_I: [(T_I) (R_I: [+S, -A] (R_I)] (C_I))] (A_J))$
c. $(A_K: [(\text{Emph } F_J) (P_I)_S (P_J)_A] (C_I))] (A_K))$

The entire Interpersonal Level is thus thoroughly actional and shows the hierarchical and equipollent arrangements involved in making Moves, performing Discourse Acts, issuing Illocutions, conveying Communicated Contents and executing Ascription and Reference. The theory's orientation to the Discourse Act entails that the Interpersonal Level is always operational (as, of course, is the Phonological Level): the primacy of the Interpersonal Level is a reflection of FDG's third goal, *pragmatic adequacy*.[2]

The three goals of typological, cognitive and pragmatic adequacy define the space within which FDG operates, allowing for differences in "style" among the various participating researchers while committing them to the same overall principles. Inasmuch as the three "adequacies" make crucial reference to phenomena that go beyond language proper, they also define FDG as belonging to the functionalist paradigm in linguistics, which seeks explanations "in such cognitive domains as memory, attention, processing and prediction, in such social domains as gender, esteem and politeness, or in the spatio-temporal and socio-cultural contexts in which language is used by speakers in their daily lives" (Mackenzie, 2016b, 470).

3 Tools

3.1 The Representational Level

It will be clear that, although essential for grammatical analysis, the Interpersonal Level is not sufficient in itself. As seen in Figure 1, it is complemented by another for-

2 The Representational and Morphosyntactic Levels are bypassed in the case of fixed Expressive and Interactive Discourse Acts such as *Ouch!* and *Congratulations!* respectively (cf. Hengeveld and Mackenzie, 2008, 76–78).

mulation level, the Representational Level, which accounts for all the semantic features of the unit being analysed. FDG distinguishes sharply between the two, with the Interpersonal Level being responsible for rhetoric (the arrangement of Discourse Acts into Moves) and pragmatics (the Illocution and lower layers), while the Representational Level covers semantics.

The inner structure hypothesized for the Representational Level is shown in Figure 3.

$((\pi\ p_1:$	Propositional Content
$\quad [(\pi\ ep_1:$	Episode
$\quad\quad [(\pi\ e_1:$	State-of-Affairs
$\quad\quad\quad (\pi\ f_1: [$	Configurational Property
$\quad\quad\quad\quad (\pi\ f_2: \$_1 \mid \blacklozenge\ (f_2): \sigma\ (f_2))$	Lexical Property
$\quad\quad\quad\quad \ldots$	
$\quad\quad\quad\quad (\pi\ x_1: (\pi\ f_2: \$_2 \mid \blacklozenge\ (f_2): \sigma\ (f_2))\ (x_1): \sigma\ (x_1))_\phi$	Individual
$\quad\quad\quad\quad]\ (f_1): \sigma\ (f_1))$	Configurational Property
$\quad\quad\quad (e_1): \sigma\ (e_1) \ldots (e_n))$	State-of-Affairs
$\quad\quad]\ (ep_1): \sigma\ (ep_1) \ldots (ep_n))$	Episode
$\quad]\ (p_1): \sigma\ (p_1))$	Propositional Content

Figure 3: The inner structuring of the Representational Level.

Even a brief glance at Figure 3 shows that the Representational Level conforms to the same principles as the Interpersonal Level, with each layer showing the pattern presented in (3) above. By convention, lowercase letters are used, but otherwise we see the same use of π to indicate operators and σ for modifiers. The Propositional Content contains a configuration of at least one Episode ($ep_{1\ldots n}$), each of which contains a configuration of States-of-Affairs ($e_{1\ldots n}$). The head of each State-of-Affairs is typically a Configurational Property (f_1). This is exemplified in the Representational Level analysis of (5) as (6):

(5) *The girl is ill.*

(6) $(p_i: (\text{Pres}\ ep_i: (e_i: (f_i: [(f_j: \$_i|\text{ill}_A\ (f_j))\ (1x_i: (f_k: \$_j|\text{girl}_N\ (f_k))\ (x_i))_U]\ (f_i))\ (e_i))\ (ep_i))\ (p_i))$

As will be apparent from (6), the lexemes introduced at the Representational Level appear in the position marked in Figure 3 with a lozenge (⧫) and have their own symbol $\$_{1\ldots n}$. Moreover, the predicate-argument relationship between the Lexical Property 'ill' and the Individual 'the girl' is marked by the presence of the *semantic function* Undergoer on the argument. FDG assumes three basic semantic functions for arguments, A(ctor), U(ndergoer) and L(ocation), which may be further specified in certain languages. For example (Hengeveld and Mackenzie, 2008, 200), positional, allative and ablative Location are all indiscriminately marked by the suffix *–se* in Tariana, while in French the grammatical preposition *à* marks both positional and allative (while *de* marks ablative), and in English there are different grammatical prepositions for each (*at*, *to* and *from* respectively). Again, typological adequacy demands that no more (and no fewer) distinctions be made than are justified by the formal properties found in Linguistic Expressions.

This principle also applies to the FDG approach to parts-of-speech. Languages differ with regard to the number of lexical classes they distinguish, up to – it is hypothesized in FDG – a maximum of four. Whereas in many approaches parts-of-speech are defined in terms of their morphosyntactic properties, the FDG definitions make reference to properties of the Interpersonal and Representational Levels, and to those alone, as seen in (7):

(7) Verb: head of a lexical property that is used as a Subact of Ascription
Noun: head of a lexical property that is used as a Subact of Reference
Adjective: head of a lexical property modifying a lexical property whose head is a Noun
Adverb: head of a lexical property modifying a lexical property whose head is a Verb

Consider example (8):

(8) a. *The bald professor sings well.*
b. $(A_J: [(F_J: \text{DECL}\ (F_J))\ (P_I)_S\ (P_J)_A\ (C_I: [(T_I)\ (R_I)]\ (C_I))]\ (A_J))$
c. $(p_i: (\text{Pres}\ ep_i: (e_i: (f_i: [(f_j: \text{sing}_V\ (f_j): (f_k: \text{well}_{Adv}\ (f_k))\ (f_j))\ (1x_i: (f_l: \text{professor}_N\ (f_l))\ (x_i): (f_m: \text{bald}_A\ (f_m))\ (x_i))_A]\ (f_i))\ (e_i))\ (ep_i))\ (p_i))$

In (8c), *sing* is the head of the lexical property (f_j) that corresponds to the Subact of Ascription (T_I), while *professor* is the head of the lexical property (f_l) that corresponds to the Subact of Reference (R_I). They thus qualify as V(erb) and N(oun) respectively. *Bald* is head of the lexical property (f_m) which modifies an N and *well* is head of the lexical property (f_k) which modifies a V; they therefore qualify as A(djective) and Adv(erb) respectively. Parts-of-speech are marked as subscripts on the respective lexemes.

English is a *differentiated* language in arguably having all four parts-of-speech, but a long tradition of typological work commencing with Hengeveld (1992), and continued within FDG, has shown that languages do not all have specialized lexeme classes for all four functions listed in (7). See Figure 4. *Flexible* languages use a single class of lexemes for more than one of the four; *rigid* languages lack lexemes with certain of the functions. Hengeveld and Mackenzie (2008, 225–227) argue that Warao is a flexible language in allowing the same lexeme to perform the second, third and fourth function, while Garo is a rigid language in not having lexemes specialized for the third or fourth function; in that language "Modifier in (R_1)" is expressed by means of a relative clause.

FDG draws a sharp distinction between *lexemes* and *words*. Lexemes are drawn from the lexicon and are introduced at the Interpersonal or Representational Lev-

Language	Type	Head in (T_1)	Head in (R_1)	Modifier in (R_1)	Modifier in (T_1)
Warao	flexible	Verb	Non-verb		
English	differentiated	Verb	Noun	Adjective	Adverb
Garo	rigid	Verb	Noun		

Figure 4: Flexible, differentiated and rigid languages.

els; these progress to the Morphosyntactic Level, where, if appropriate, they are inflected and appear as words. The Morphosyntactic Level has its own primitives, including "grammatical morphemes": these include affixes such as the agreement suffix –*s* in (8a) but also "grammatical words" (Gw_1). The insertion of grammatical words is typically triggered by operators or functions at the Interpersonal or Representational Level: thus *the* in (8a) is triggered by the operator (+id) "identifiable" on the corresponding Subact of Reference.

FDG discussion of the distinction between lexemes and grammatical words has centred on the analysis of adpositions (Keizer, 2008; Mackenzie, 2013). Those adpositions that express a semantic function, such as English *by* as an expression of Actor, are uncontroversially grammatical. The spatial adpositions of English, however, cannot all be classified as grammatical. Mackenzie (2013) has argued that in English there are precisely five grammatical spatial prepositions, namely {*at*, *to*, *from*, *via*, *toward*(*s*)}, all the others being lexical. Since a Lexical Property has the form (π f_2: $\$_1$|♦ ($f_2$): σ (f_2)) (cf. Figure 3), a defining property of lexemes is that they accept a modifier: spatial adpositions such as {*inside*, *under*, *beside*, . . . } all accept modifiers (*right inside the cave*, *well under the bed, close beside me*, . . .) and thus qualify as lexical, while grammatical adpositions do not. Similar remarks apply to temporal prepositions in English, where only the set {*at*, *to*/*until*, *from*, *for*} are grammatical, all others being lexemes, cf. *shortly after the war*. One advantage of analysing the majority of spatial and temporal adpositions as lexical is that it becomes understandable why very many of the same forms can also occur as "conjunctions" (Hengeveld and Wanders, 2007) or as "adverbs". The conjunctions are analysed as taking an argument whose head is a Configurational Property rather than, as an adposition does, an argument whose head is a Lexical Property; and the adverbs are merely adpositions without such an argument. To solve the terminological dilemma, all three are regarded as belonging to the lexeme class Ad, cf. (9):

(9) a. *before the war*

$(t_i: (f_i: [(f_j: before_{Ad}\ (f_j))\ (1\ ep_i: (f_k: war_N\ (f_k))\ (ep_i))_U]\ (f_i))\ (t_i))_L$

b. *before the war broke out*

$(t_i: (f_i: [(f_j: before_{Ad}\ (f_j))\ (Past\ ep_i: (e_i: (f_k: [(f_l: break_V\_out\ (f_l))\ (1\ ep_j: (f_m: war_N\ (f_m))\ (ep_j))_U]\ (f_k))\ (e_i))\ (ep_i))_U]\ (f_i))\ (t_i))_L$

c. *before*

$(t_i: (f_i: (f_j: before_{Ad}\ (f_j))\ (f_i))\ (t_i))_L$

The sharpness of the distinction between lexical and grammatical items is addressed by Keizer (2007), who confronts the difficulty that categorization in linguistics is rarely an all-or-nothing affair. Applying an inventory of criteria to a range of phenomena in English grammar that have proved hard to classify one way or the other, she concludes that it is necessary to recognize among the primitives of formulation a class of "lexical operators", which appear in the π position but have lexical content. One example Keizer considers is the class of numerals (both cardinal and ordinal), analysed in FDG as operators, i. e. as lacking lexical status. While conceding that they pattern in many ways like grammatical words, she shows that they also display typically lexical properties such as participating in morphological derivation (*two-seater*, *tenfold*, *secondly*, etc.), accepting modification (*approximately three*, *almost twenty*, etc.) and potentially bearing the pragmatic function Focus (*I want THREE, please*). The intermediary category of lexical operator, in which the operator position π is occupied by an item in morphophonemic form, has played an important part in the FDG approach to grammaticalization as a stage between lexical and grammatical status (Hengeveld, 2017).

Keizer (2016) turns to another challenge (for any grammatical theory), the analysis of examples such as those in (10):

(10) a. *He kicked the bucket*.
b. *He hit the ceiling*.
c. *He spilled the beans*.

Although the three clauses are (arguably) morphosyntactically identical, Keizer shows that each represents a different subgroup of idiomatic expressions, with specific functional properties that can be elucidated in FDG. (10a) exemplifies unmotivated, semantically non-decomposable idioms, which are analysed as being associated with a single Ascriptive Act and a single Lexical Property *kick_the_bucket*. (10b) exemplifies motivated but semantically non-decomposable idioms and is associated with a single Ascriptive Act but with a Configurational Property in which the singularity of the Undergoer argument *ceiling* is pre-specified (cf. **He hit the ceilings*); moreover, the occurrence of *the* is specified at the Interpersonal Level (cf. **He hit a ceiling*). (10c), finally, exemplifies motivated and semantically decomposable idioms, associated with two Subacts, one of Ascription ('spill') and the other of Reference ('the beans'); at the Representational Level we find the Configurational Property $(\mathrm{f}_1\text{: } [\mathbf{(f_2\text{: } spill_V\ (f_2))}\ (\mathrm{x}_1)_\mathrm{A}\ \mathbf{(mx_2\text{: } (f_3\text{: } bean_N\ (f_3))\ (x_2))_U}]\ (\mathrm{f}_1))$, in which the elements in bold are pre-specified (cf. **He dropped the beans*, **He spilled a bean*, but ✓*He has been spilling beans all his life*, ✓*the inevitable spilling of* (*the*) *beans*, constructions that motivate the presence of a full Configurational Property). On this basis, Keizer proposes the notion of *partially instantiated frames*, some elements of which are pre-specified, while others are open. To fully understand this proposal, we now must turn to the notion of "frame" in FDG.

For each layer of the Interpersonal and Representational Levels, Hengeveld and Mackenzie (2008) list the frames associated with that layer, that is, the combinations of elements that occur in the language under description. Relevant distinctions captured by frames at the Interpersonal Level include the different inner structures of Discourse Acts (e. g. in terms of the presence/absence of an Addressee – assumed to be absent for Expressives like *Ouch!*) or the possible combinations of Subacts within the Communicated Content (assumed to subtend the thetic–categorial–presentative distinction). At the Representational Level, there are, for example, frames for the various types of Configurational Property permitted by the language, covering such matters as quantitative valency (the number of arguments allowed, both minimum and maximum) and qualitative valency (the semantic functions of those arguments). Hengeveld and Mackenzie (2016) argue that what they call a language's *frameset* is representationally distinct from the inventory of its lexemes. For various reasons connected with the cognitive adequacy of the theory, they propose that lexemes are associated in the speaker's experience with a selection of those frames, but not in a limitative way. One of their aims is to provide space for the creative use of frames by language users, as in *I porched the papers without getting off my bike*, where the speaker uses a lexeme that is most often associated with the frame for a Location, i. e. (11), in a frame for a two-place Configurational Property, i. e. (12) (cf. García Velasco, 2016 on such flexibility in the FDG lexicon):

(11) $(f_1: [(f_2: \blacklozenge_N (f_2)) (l_1)_U] (f_1))$

(12) $(f_1: [(f_2\ \blacklozenge_V (f_2)) (x_1)_A (x_2)_U] (f_1))$

Keizer's (2016) notion of a partially instantiated frame is thus a frame in which at least one element is pre-specified for a particular lexeme, lexical operator or grammatical operator.

The operation of formulation, then, yields two analyses, the Interpersonal Level which covers all aspects of the formal expression of the Discourse Act that reflect its rhetorical and pragmatic properties, and the Representational Level that accounts for all its semantic properties. It is the task of the operation of encoding to convert the hierarchically structured but unordered contents of those two levels into linguistic form. As already mentioned, encoding yields two levels, one morphosyntactic and the other phonological but in this chapter, given the nature of the volume in which it appears, we will focus on the Morphosyntactic Level. Its task is to unify the information incoming from the formulation levels in such a way that the lexemes present in the input are (in relevant languages) inflected and appropriately ordered within a hierarchical morphosyntactic structure and are supported by suitably positioned grammatical morphemes. Given that the Morphosyntactic Level, to which we now turn, is completely dependent upon its input, it would not have made sense to deal with it without considering the constitution of the Interpersonal and Representational Levels. In FDG, which as we have seen delivers four modular analyses for each Discourse Act, the Morphosyntactic Level is neither more central nor less important than the other levels and therefore must be approached against the background of the full model.

3.2 The Morphosyntactic Level

The Morphosyntactic Level, like the two levels presented in Sections 2 and 3.1, is layered. Four syntactic layers are recognized: Linguistic Expression (Le_1), Clause (Cl_1), Phrase (Xp_1), and Word (Xw_1), with further "morphological" layering within the Word; X here indicates the type of Word or Phrase, e. g. (Vw_1) 'Verb word' or (Np_1) 'Noun phrase'. (Note that the variables at this level always take the form of one upper-case and then one or more lower-case letters.) The layering at the Morphosyntactic Level is displayed in Figure 5.

(Le_1:	Linguistic Expression
(Cl_1:	Clause
(Xp_1 :	Phrase
(Xw_1 :	Word
(Xs_1)	Stem
$(Aff_1)^n$	Affix
$(Xw_1))^n$	Word
$(Xp_1))^n$	Phrase
$(Cl_1))^n$	Clause
(Le_1))	Linguistic Expression

Figure 5: The inner structuring of the Morphosyntactic Level.

In keeping with the generally recognized rank scale, Words make up Phrases, Phrases make up Clauses, and Clauses make up Linguistic Expressions (the possibility of multiple, equipollent constituency being symbolized in Figure 5 by the superscript n). However, where the language under analysis has recursive morphosyntactic structure, Phrases may also contain other Phrases or Clauses, and Clauses may contain other Clauses. In addition, FDG allows for Clauses to contain Words and permits Words to contain not only Morphemes (i. e. Stems and Affixes) but also Words, Phrases, and even Clauses. In Bininj Gun-Wok (Evans, 2003, 536), for example, we find examples of Words containing Clauses, as in the following example:

(13) *Ga-ganj-ngu-nihmi-re.*
3-meat-eat-GER-go.PST.PFV
'He goes along eating meat.'

Here the Phrase *ganj* 'meat' is an argument of the incorporated verb *-ngu-*, not of the incorporating verb *-re*. The embedded clausal nature of the incorporated unit is furthermore reflected in the gerundial ending *-nihmi* on the incorporated verb. The extent to which recursion applies is an empirical question to be examined per language. The possible existence of languages without recursivity has been extensively discussed of late (Everett, 2005; Nevins et al., 2009); FDG is set up to recognize languages with any degree of recursivity, from none to the maximum.

As will be clear from the name of the Morphosyntactic Level, FDG shares with various other approaches the position that there is no essential difference between syntax and morphology. The template structure to be described below also applies within the Phrase and within the Word, and phenomena such as dummies and agreement are equally prevalent among the constituent parts of the Word. This is the basis for not distinguishing between a Syntactic and a Morphological Level.

It should be stressed that the categories applied at the Morphosyntactic Level (Word, Phrase, etc.) only apply there and relate to formal, strictly morphosyntactic categories. The Lexical Properties introduced at the formulation levels carry, as we have seen, a subscript indicating the part-of-speech to which they belong. In English, which has the lexeme classes Verb, Noun, Adjective and Ad(verb), there is generally a good correspondence between these lexeme classes and the Word classes (Nw_1), (Vw_1), ($Adjw_1$) and ($Ad(v)w_1$). In Warao, identified in Figure 4 as a highly flexible language, the same lexeme (of the class Non-Verb) may be used as Noun, Adjective or Adverb, as seen in the following data from Romero-Figueroa (1997, 49, 50, 119):

(14) a. *yakera* (Nw_1)
goodness
'goodness'

b. *Hiaka yakera* ($Adjw_1$) *auka saba tai nisa-n-a-e.*
garment goodness daughter for 3SG.F buy-SG-PUNCT-PST
'She bought a beautiful dress for her daughter.'

c. *Oko kuana* ($Advw_1$) *yaota-te arone yakera* ($Advw_2$) *nahoro-te, ...*
1PL hardness work-NONPST although goodness eat-NONPST
'Although we work hard and eat well, ...'

It is clear that the syntactic context in which *yakera* occurs – after the Noun word in (14b) and before the Verb word in (14c) – gives *a priori* justification for analysing it as an Adjective word in the first context and an Adverb word in the second (as indeed also applies to *kuana* in (14c)).

Although languages are symbolic constructs and therefore in principle could tolerate a maximally arbitrary relation between function and form, what we observe in reality is a large degree of homology: this observation justifies all functional approaches to grammar, including FDG. The relation between the Morphosyntactic Level and the formulation levels is seen by Hengeveld and Mackenzie (2008) as being governed by three principles that maximize parallelism across the levels, namely *iconicity*, *domain integrity*, and *functional stability*. These have recently been subsumed by Hengeveld and Leufkens (2018) under a more general principle of Transparency, which they define very broadly as "a one-to-one relation between units of meaning and units of form" (2018, 141).

Iconicity can be illustrated by the correspondence between, on the one hand, how Moves and Discourse Acts are analysed at the Interpersonal Level and Propositions and Episodes at the Representational Level and, on the other, the order in which they are expressed. The possibility of adding modifiers indicating the position of a unit in a discourse sequence (*firstly*, *secondly*, …) or the role of a proposition in a debate (say, as a consequence (*therefore*) or a counter-argument (*however*)), as well as operators on States-of-Affairs that allude to relative positioning in a temporal sequence (Anterior, Simultaneous, Posterior, …), all demonstrate that ordering in physical reality and mental experience should be reflected at the formulation levels. For this reason (15a) is more natural than (15b), despite their being synonymous in truth value:

(15) a. *The game began at 7.30 and ended in a draw.*
b. ?*The game ended in a draw and began at 7.30.*

As with all three principles to be discussed here, iconicity can be overridden by other communicative factors. Consider the following examples:

(16) a. *The game, which began at 7.30, ended in a draw.*
b. *The game, which ended in a draw, began at 7.30.*

FDG analyses non-restrictive relative clauses of the type shown in (16) as Dependent Discourse Acts (unlike restrictive relatives, which belong to the same Discourse Act as the antecedent). In (16a), iconicity is respected, because the beginning of the game is mentioned before its conclusion; (16b), despite being anti-iconic, is also an acceptable form because iconicity is overruled by the speaker's focus on the starting time of the game.

The principle of domain integrity refers to the cross-linguistic preference for units that belong together at the Interpersonal and/or Representational Levels to be juxtaposed at the Morphosyntactic Level. For example, modifiers should ideally be placed in expression next to the heads that they modify; functions and operators should be expressed by elements that are close to the morphosyntactic units to which they apply; and the realization of one Subact of Reference, for example, should not be interrupted by that of another Subact of Reference.

Again, this principle applies as a default, but many languages show instances where Domain Integrity is overridden by other communicative strategies. Here are some simple examples from English:

(17) a. *Are you going into town?*
b. *What are you looking at?*

(18) *I am now going into town.*

(19) *The guy has arrived who's going to fix my lock.*

In (17a) the integrity of the Verb phrase (Vp) *are going* is violated by the clause-initial placement of *are* to signal an Interrogative Illocution. In (17b) there is in addition to the non-integrity of the Vp a violation of the integrity of the Adposition phrase (Adp) *at what*, given the clause-initial placement of the question-word in English, which is justified by its Focus status. In (18) the Vp *am going* loses integrity by being interrupted by an Adverb phrase (Advp) *now*. And in (19) the integrity of the Np *the guy who's going to fix my lock* is broken by the placement of the bulky relative clause in clause-final position. Note *en passant* that FDG contains nothing resembling transformational rules that move constituents; a corollary is that there are no empty nodes, traces, D-structure, S-structure or any of the paraphernalia of those theories of syntax that assume movement rules.

The principle of functional stability, finally, requires that constituents with the same interpersonal or representational specification be located in the same position relative to other categories. In Turkish, for example, a Focus-bearing constituent is recognizably placed in immediately pre-verbal position, with the tense-bearing verb position also being fixed, namely as P^F. Within the Word layer, the principle of functional stability is of particular importance in the sense that in complex words the relative order of meaning-bearing elements will strongly tend to be fixed. Consider the analysis of words such as the following from Turkish (Lewis, 1967, 124), where the suffix *-miş* occurs twice in succession, once as a resultative and once as an evidential suffix:

(20) *Gel-miş-miş-∅.*
come-RES-NONVIS.PST-3SG
'S/he came, apparently.'

The position closer to the stem marks off the first occurrence of the suffix as corresponding to an f_1-operator while the position further from the stem reflects the presence of an evidentiality operator at (p_1).

These considerations bring us naturally to the matter of positions in the FDG approach to linearization. Bear in mind that there is no sooner-to-later sequencing at the formulation levels, except in the broad sense discussed with relation to iconicity above. It is a task of the Morphosyntactic Level to impose what linguists refer to as left-right ordering. To achieve this, it draws on its primitives, which include a language-specific set of templates for each of the layers. These contain one to four *absolute positions*, namely initial (P^I), second (P^2), middle (P^M), and final (P^F) position. These four are claimed to be cross-linguistically relevant but do not necessarily all apply in every language or at every layer. Which absolute positions are relevant for a specific language's syntax and its layers is determined empirically, on the methodological assumption that the fewest possible should be applied. For reasons of space, we will focus here on the ordering of the elements of the Clause.

As shown in Figure 6, once an absolute position (shown in bold) has been occupied, the template can be expanded with further relative positions.

$\mathbf{P}^{\mathbf{I}}$	P^{I+1}	P^{I+2}	…							
	$\mathbf{P}^{\mathbf{2}}$	P^{2+1}	P^{2+2}	…						
	…		P^{M-2}	P^{M-1}	$\mathbf{P}^{\mathbf{M}}$	P^{M+1}	P^{M+2}	…		
							…	P^{F-2}	P^{F-1}	$\mathbf{P}^{\mathbf{F}}$

Figure 6: Absolute and relative positions at the Morphosyntactic Level.

Figure 6 states that if we find evidence for a position P^I, it may be expanded to the right by further positions (P^{I+1}, etc.). P^F may of course only be expanded to the left, while P^M may be expanded in either direction. P^2, finally, can only be expanded to the right and if it is present, P^I cannot be expanded to the right.

The morphosyntax of main clauses in German requires a template that contains P^I, P^2, P^M and P^F, cf. (21):

(21) *Sie* (P^I) *ist* (P^2) *gestern* (P^M) *an-ge-komm-en* (P^F).
3S.F AUX.3S yesterday on-PTCP-come-PTCP
'She arrived yesterday.'

German declarative main clauses reserve P^2 for finite verbs (in this case a finite auxiliary). This entails, since P^2 cannot expand leftwards, that *sie* is placed in P^I. *Angekommen*, like non-finite verbal elements generally in German, goes to clause-final position. *Gestern*, introduced as a modifier at the Episode layer of the Representational Level, is in P^M (as we shall see, elements originating in the hierarchical part of formulation structure have prior access to absolute positions). The access of *ist* to an absolute position is also justified by the fact that it marks tense, corresponding to an operator at the same Episode layer. Notice that whatever word order variations on (21) occur, *ist* remains in the P^2 position. Consider the following alternatives, which differ in the distribution of pragmatic functions at the Interpersonal Level:

(22) a. *Gestern* (P^I) *ist* (P^2) *sie* (P^{2+1}) *angekommen* (P^F).
b. *Angekommen* (P^I) *ist* (P^2) *sie* (P^{2+1}) *gestern* (P^F).

The discussion of (21) has used such expressions as "are placed first", "goes to" or "prior access". This is justified by FDG's approach to constituent order, in which the various operations that lead to morphosyntactic structure occur in sequence (García Velasco et al. 2012). The order of operations does not reconstruct the incremental left-to-right/sooner-to-later processing of the utterance (indeed, given the leftwards expansion of P^M and P^F, this would not be thinkable); rather, the order of *dynamic implementation* indicates "the sequence of steps that the analyst must take in understanding and laying bare the nature of a particular phenomenon" (Hengeveld and Mackenzie, 2008, 2).

Within dynamic implementation, a strict distinction is made between the placement of hierarchical and configurational elements. In FDG, unlike most if not all other approaches to morphosyntax, those elements that originate in the configurationally structured parts of the formulation levels are not ordered until all the elements of the hierarchically structured parts have been placed in the template. This is a reflection of the observation first articulated in Hengeveld (1989) that one major function of morphosyntactic order is to reflect relations of relative scope (similar considerations underlie the “cartographic” research programme in syntax; cf. Shlonsky, 2010 for an introduction). As a result, elements that correspond to modifiers, operators or functions at the formulation levels are placed first, after which the configurationally organized elements go to the remaining places. The hierarchical elements therefore have prior access to the absolute positions, as was seen to apply to *ist* and *gestern* in (21). This means that, unlike approaches in which the placement of the arguments is primary (e. g. Subject–Verb–Object) and adverbials are then grouped around or between these elements, FDG places gives pride of place to adverbials (being hierarchically ordered modifiers), and the arguments take whatever positions remain.

It is the hierarchically highest elements of the formulation levels whose counterparts are positioned earliest at the Morphosyntactic Level. In addition, the entire Interpersonal Level is dealt with before the Representational Level. This downward progression through the hierarchies is shown in Figure 7.

Interpersonal Level	M ➢ A ➢ C ➢ R ➢ T
	↓
Representational Level	p ➢ ep ➢ e ➢ f^c ➢ f^l

Figure 7: Ordering of expression of hierarchical elements in FDG.

We can deduce from Figure 7 that, in hierarchical ordering, modifiers, operators, and functions at the Move layer of the Interpersonal Level (M) are the first to be placed, while the last to be placed are modifiers, operators, and functions at the Configurational Property (f^c) and Lexical Property (f^l) layers at the Representational Level.

One justification for this approach is the observation that discourse markers (more precisely, modifiers of the Discourse Act) typically take a peripheral position in the syntax of the clause. They therefore have first dibs on P^I or P^F, inevitably relegating any elements placed later to a less peripheral position, namely P^{I+n} (if available) or P^{F-n}. In (27), for example, *so* as a Discourse Act modifier precedes the Communicated Content-modifying adverb *unfortunately*, both from the Interpersonal Level but at respectively higher and lower layers, and the latter precedes the Episode-modifying *today*, originating at the Representational Level:

(23) *So* (P^I) *unfortunately* (P^{I+1}) *today* (P^{I+2}) *I* (P^{I+3}) *have* (P^M) [*a terrible hangover*] (P^{M+1}).

The result is another manifestation of the principle of iconicity discussed above: scope correlates with relative position. Note again that the finite verb *have*, carrying the absolute tense from the Episode layer, occupies P^M; the subject *I* and the object *a ter-*

rible hangover (both originating in the Configurational Property) end up in relative positions.

The approach sketched here also accounts for the well-documented tendency for Topic and Focus elements to occupy peripheral positions in relevant languages. Topic and Focus are pragmatic functions of Subacts and thus impact morphosyntactic structure after the highest layers of the Interpersonal Level but before any Representational Level phenomena. In Tzotzil (Aissen, 1992), if both a Topic and a Focus constituent are present, the Topic occurs in the initial position and the Focus in the post-initial position. The language is otherwise predicate-initial, so that the verb, which carries Tense, a representational feature, is pushed back to third position, P^{I+2}:

(24) [*A ti prove tzeb-e*] (P^{I}) *sovra* (P^{I+1}) *ch'ak'bat* (P^{I+2}). (Aissen, 1992, 51)
TOP DEF poor girl-TOP leftovers was.given
'It was leftovers that the poor girl was given.'

If there is only either a single Topic or a single Focus present, the predicate occurs in post-initial position; only if there is neither a Topic nor a Focus does it occur in initial position. Languages like this, in which morphosyntactic phenomena are determined by properties of the Interpersonal Level analysis, are said to show *interpersonal alignment*.

The placement of constituents with respect to their semantic functions is illustrated for Turkish in (25) (Kornfilt, 1997, 90):

(25) *Hasan* (P^{I}) *kitab-ı* (P^{F-2}) *Ali-ye* (P^{F-1}) *ver-di* (P^{F}).
Hasan.NOM book-ACC Ali-DAT give-PST.3SG
'Hasan gave the book to Ali.'

In Turkish, the placement of the accusative and dative Nps – unlike the nominative Np, which is a Subject – is determined by their semantic functions (as stated at the Representational Level). The order shown in (25) may be changed as a result of the assignment of a pragmatic function (which of course takes precedence over semantic functions), but absent such pragmatic functions the Recipient precedes the predicate and is itself preceded by the Undergoer, occupying P^{F-n} positions relative to the absolute final position of the predicate. Languages in which morphosyntactic phenomena are determined by properties of the Representational Level analysis (e. g. semantic functions, or animacy) are said to show *representational alignment*.

Finally, languages in which morphosyntactic structure neutralizes pragmatic or semantic distinctions are said to display *morphosyntactic alignment*. One prominent way in which such neutralization is manifested is in the assignment of the *syntactic functions* Subject and Object (and the like); another is when word order is (co-)determined by complexity or "weight". English is a relevant example: firstly, Subject and Object can be assigned to elements independently of their Topic or Focus status, as in (26) and (27); secondly, Subject and Object neutralize the semantic distinction between Actor and Undergoer, and between Undergoer and Location (recipient) respectively, as in (28) and (29):

(26) a. (How did you react?) *I*$_{\text{TopSubj}}$ *objected*$_{\text{Foc}}$.
b. (So who objected?) *I*$_{\text{FocSubj}}$ *objected*.

(27) a. (Who did they award a prize?) *They awarded my doctoral student*$_{\text{FocObj}}$ *a prize*.
b. (What did they award your doctoral student?) *They awarded her*$_{\text{Obj}}$ *a prize*$_{\text{Foc}}$.

(28) a. *The stranger*$_{\text{ASubj}}$ *shouted*.
b. *The stranger*$_{\text{USubj}}$ *fell*.

(29) a. *The midfielder*$_{\text{ASubj}}$ *passed the ball*$_{\text{UObj}}$ *to me*$_{\text{L}}$.
b. *The midfielder*$_{\text{ASubj}}$ *passed me*$_{\text{LObj}}$ *the ball*$_{\text{U}}$.

Syntactic functions are thus assigned only in languages in which pragmatic and/or semantic considerations are neutralized; this entails that subject and object are not regarded as universal features of language structure. In practice, individual languages tend to display a mixture of the three types of alignment, but one of the three is typically dominant.

Although, as we have seen, hierarchy is not absent from the FDG approach to syntax (cf. Figure 5 above), the emphasis – as is apparent in Figure 6 – is on linearity. The hierarchical relations that are so prominently present in formulation are expressed at the Morphosyntactic Level in the relative order of constituents, and the operation of encoding does not duplicate those hierarchical relations. This stance brings FDG closer to Culicover and Jackendoff's Simpler Syntax (2005; this volume), in which they propose "relatively flat" (2005, 108), multiple-branching structures, which they defend as being analogous to other offshoots of the generative paradigm such as Lexical Functional Grammar (Bresnan, 2001; Dalrymple and Findlay this volume), Head-Driven Phrase Structure Grammar (Pollard and Sag, 1994; Müller and Machicao y Priemer this volume) or indeed the more practical grammar of English edited by Huddleston and Pullum (2002) and "the view of syntax generally adopted in psycholinguistic research" (Culicover and Jackendoff, 2005, 107).

Summarizing this overview of the main tenets of the FDG approach to syntax, we may say that:

– syntax is not viewed autonomously but as an aspect of the encoding of the two formulation levels;
– syntax is seen as being subject to the same principles as morphology, hence the assumption of a Morphosyntactic Level;
– morphosyntax is governed by the three principles of iconicity, domain integrity and functional stability, jointly known as transparency;

- a distinction is drawn between absolute and relative positions;
- in dynamic implementation, hierarchically related elements have prior access to absolute positions, with configurational elements being placed thereafter;
- languages are distinguished according as they display interpersonal, representational or morphosyntactic alignment;
- hierarchy is not absent from the Morphosyntactic Level, but the emphasis is on flat, linear structure.

4 Data

Although the central principles of the model and the implementation of those principles across four layered levels of analysis are shared by all members of the international FDG community, the style of argumentation and the types of data used to provide evidence for that argumentation can vary quite widely in accordance with the particular academic settings, traditions and goals of individual researchers. The theory's orientation to understanding grammar as an instrument of action and interaction has, for the majority, encouraged the use of corpus data to clarify the embedding of each Discourse Act in its context. This is not to say that FDG is used for purposes of discourse analysis: the name "Functional Discourse Grammar" has frequently been misunderstood to imply a "functional grammar of discourse", but as will be clear from this chapter, this is not the aim of the model.

A flashpoint for disagreements among FDG researchers has been the role and extent of the Contextual Component in the overall model, which, as shown in Figure 1, is conceived of as (a) receiving and storing the analyses from the four Levels of the Grammatical Component and (b) having the potential to influence the operations of formulation and encoding. This rather restricted conception of the Contextual Component (defended by Hengeveld and Mackenzie, 2014) has been challenged by Butler (2008b), Cornish (2013), and Connolly (2014), who from different angles argue that context impinges on grammatical processes in ways that go far beyond Hengeveld and Mackenzie's "conservative stance" (Butler, 2008b, 238). In addition, Mackenzie (2014) and Giomi (2014) have maintained that FDG, in aligning with language production models, is too strongly oriented to the speaker and have proposed an alternative view of the Contextual Component as shared by speaker and hearer in interaction. The evidence for these challenges and counterarguments has been drawn from corpus data and/or from transcriptions of dialogue, mostly in English.

Corpus data, often coupled to work with informants, is the basis for the application of FDG to the description of poorly described and/or minority languages, such as Genee's (2009) work on Blackfoot, Wolvengrey's thesis (2011) on Plains Cree, or Hengeveld and Fischer's grammar of A'ingae (in prep.). In these cases, the researchers are obliged to put together their own corpora, rather than making use of already existing collections of texts or transcriptions. Corpus work also underlies the substantial body of work contributed by Brazilian colleagues, working on oral Brazilian Portuguese, which is subject to considerable variation as well as existing in a situation of diglossia with the "cultivated norm" of the standard language; Pezatti (2014) is the fullest application of FDG to the syntax of any language published to date.

FDG work on language typology, by contrast, is strongly dependent upon "second-hand" data from published grammars, themselves often written in theoretical frameworks that do not share the model's functional principles and pay scant attention to discourse factors. Much of the work for which FDG is best known – on parts-of-speech systems, on aspect—tense—modality, on adpositions, on transparency, on grammaticalization (see various references above) – has drawn most of its data from such language descriptions, generally employing the sampling criteria developed by Rijkhoff et al. (1993). And finally, as in the present chapter, authors have not hesitated from time to time to call on their personal knowledge of their own language or languages very familiar to them to illustrate their arguments.

The extent to which the various types of evidence converge has not been examined for FDG, but the detailed proposals for the Grammatical Component do act as a brake on excessive divergence and generate an exchange of analyses within the research group that is valid and insightful.

5 Sample analysis

FDG differs from many functionally oriented grammars in employing formalization; accordingly, a formal analysis can be given for sentence (30) proposed by the editors, as shown in Figure 8. The analysis involves various details some of which may be controversial; I therefore present it as "my analysis" rather than "the FDG analysis".[3]

(30) *After Mary introduced herself to the audience, she turned to a man she had met before.*

At the Interpersonal Level, the example sentence is treated as a Move (M_I); without further co-text, it cannot be said whether this Move contains more than the Discourse Act (A_I). The decision to identify a single Discourse Act here anticipates the phonological form, in which there is a rising and then a falling Intonational Phrase. If the formulation had been as in (31), with two falling Intonational Phrases and a longer pause, then analysis as two Discourse Acts would have been indicated:

(31) *Mary introduced herself to the audience. Then she turned to a man she had met before.*

3 I am grateful to Kees Hengeveld for his perspicacious comments on an earlier draft.

Interpersonal Level: (M$_I$: (A$_I$: [(F$_I$: DECL (F$_I$)) (P$_I$)$_S$ (P$_J$)$_A$ (C$_I$: [(T$_I$) (+id R$_I$: /mɛərɪ/$_N$ (R$_I$)) (+id R$_J$)] (C$_I$)) (C$_J$: [(T$_J$) (R$_I$) (–id R$_K$)] (C$_J$))] (A$_I$)) (M$_I$))

Representational Level: (p$_i$: (Past ep$_i$: (Sim e$_i$: (f$_i$: [(f$_j$: /tɜːn/$_V$ (f$_j$)) (1x$_i$)$_A$ (1x$_j$: (f$_k$: /mæn/$_N$ (f$_k$)) (x$_j$)): (Past ep$_j$: (Ant e$_j$: (f$_l$: [(f$_m$: /miːt/$_N$ (f$_m$)) (x$_i$)$_A$ (x$_j$)$_U$] (f$_l$)) (e$_j$): (t$_i$: (f$_n$: /bɪˈfɔː/$_{Ad}$ (f$_n$)) (t$_i$))$_L$ (e$_j$)) (ep$_j$))] (x$_j$))$_L$] (f$_i$)) (e$_i$)) (ep$_i$): (t$_j$: (f$_o$: [(f$_p$: /ˈɑːftər/$_{Ad}$ (f$_p$)) (Past ep$_k$: (Sim e$_k$: (f$_q$: [(f$_r$: /ɪntrəˈdjuːs/$_V$ (f$_r$)) (1x$_i$)$_A$ (x$_i$)$_U$ (1x$_k$: (f$_s$: /ˈɔ:dɪəns/$_N$ (f$_s$)) (x$_k$))$_L$] (f$_q$)) (e$_k$)) (ep$_k$))] (f$_o$)) (t$_j$))$_L$ (ep$_i$)) (p$_i$))

Morphosyntactic Level: (Le$_i$: [(Cl$_i$: [(Adp$_i$: (Adw$_i$: /ˈɑːftər/$_{Ad}$ (Adw$_i$)) (Adp$_i$)) (Np$_i$: (Nw$_i$: /mɛərɪ/$_N$ (Nw$_i$)) (Np$_i$)) (Vp$_i$: (Vw$_i$: [Vs$_i$: /ɪntrəˈdjuːs/$_V$ (Vs$_i$)) (Af$_i$: PAST (Af$_i$))] (Vw$_i$)) Vp$_i$)) (Gw$_i$: /hɜːˈself/ (Gw$_i$)) (Adp$_j$: [(Adw$_j$: /tʊ/ (Adw$_j$)) (Np$_j$: [(Gw$_j$: DEF (Gw$_j$)) (Nw$_j$: /ˈɔ:dɪəns/$_N$ (Nw$_j$))] (Np$_j$)) (Adp$_j$))] (Cl$_i$)) (Cl$_j$: [(Np$_k$: (Nw$_k$: /ʃɪ/ (Nw$_k$)) (Np$_k$)) (Vp$_j$: (Vw$_j$: [Vs$_j$: /tɜːn/$_V$ (Vs$_j$)) (Af$_j$: PAST (Af$_j$))] (Vw$_j$)) (Vp$_j$)) (Adp$_k$: [(Adw$_k$: /tʊ/ (Adw$_k$)) (Np$_l$: [(Gw$_k$: INDEF (Gw$_k$)) (Nw$_l$: /mæn/$_N$ (Nw$_l$)) (Cl$_k$: [(Np$_m$: (Nw$_m$: /ʃɪ/ (Nw$_m$)) (Np$_m$)) (Vp$_k$: [(Vw$_k$: /hæd/ (Vw$_k$)) (Vw$_l$: [/miːt/ + PAST] (Vw$_l$))] (Vp$_k$)) (Adp$_l$: (Adw$_l$: /bɪˈfɔː/$_{Ad}$ (Adw$_l$)) (Adp$_l$))] (Cl$_k$))] (Np$_l$))] (Adp$_k$))] (Cl$_j$))] (Le$_i$))

Morphosyntactic Level Positional Analysis:

Cl$_i$:	P^I	(Adp$_i$: (Adw$_i$: /ˈɑːftər/$_{Ad}$ (Adw$_i$)) (Adp$_i$))
	P^{I+1}	(Np$_i$: (Nw$_i$: /mɛərɪ/$_N$ (Nw$_i$)) (Np$_i$))
	P^M	(Vp$_i$: (Vw$_i$: [Vs$_i$: /ɪntrəˈdjuːs/$_V$ (Vs$_i$)) (Af$_i$: PAST (Af$_i$))] (Vw$_i$)) Vp$_i$))
	P^{M+1}	(Gw$_i$: /hɜːˈself/ (Gw$_i$))
	P^{M+2}	(Adp$_j$: [(Adw$_j$: /tʊ/ (Adw$_j$)) (Np$_j$: [(Gw$_j$: DEF (Gw$_j$)) (Nw$_j$: /ˈɔ:dɪəns/$_N$ (Nw$_j$))] (Np$_j$)) (Adp$_j$))
Cl$_j$:	P^{pre}	(Cl$_i$)
	P^I	(Np$_k$: (Nw$_k$: /ʃɪ/ (Nw$_k$)) (Np$_k$))
	P^M	(Vp$_j$: (Vw$_j$: [Vs$_j$: /tɜːn/$_V$ (Vs$_j$)) (Af$_j$: PAST (Af$_j$))] (Vw$_j$)) (Vp$_j$))
	P^{M+1}	(Adp$_k$: [(Adw$_k$: /tʊ/ (Adw$_k$)) (Np$_l$: [(Gw$_k$: INDEF (Gw$_k$)) (Nw$_l$: /mæn/$_N$ (Nw$_l$)) (Cl$_k$: [(Np$_m$: (Nw$_m$: /ʃɪ/ (Nw$_m$)) (Np$_m$)) (Vp$_k$: [(Vw$_k$: /hæd/ (Vw$_k$)) (Vw$_l$: [/miːt/ + PAST] (Vw$_l$))] (Vp$_k$)) (Adp$_l$: (Adw$_l$: /bɪˈfɔː/$_{Ad}$ (Adw$_l$)) (Adp$_l$))] (Cl$_k$))] (Np$_l$))] (Adp$_k$))
Cl$_k$	P^I	(Np$_m$: (Nw$_m$: /ʃɪ/ (Nw$_m$)) (Np$_m$))
	P^M	(Vp$_k$: [(Vw$_k$: /hæd/ (Vw$_k$)) (Vw$_l$: [/miːt/ + PAST] (Vw$_l$))] (Vp$_k$)))
	P^F	(Adp$_l$: (Adw$_l$: /bɪˈfɔː/$_{Ad}$ (Adw$_l$)) (Adp$_l$))

Phonological Level: (U$_i$: [(r IP$_i$: [(PP$_i$: /ˈɑːftə/ (PP$_i$)) (PP$_j$: /ˈmɛərɪ/ (PP$_j$)) (PP$_k$: /ɪntrəˈdjuːsthɜːˈself/ (PP$_k$)) (PP$_l$: /tʊðɪˈɔ:dɪəns/ (PP$_l$))] (IP$_i$)) (f IP$_j$: [(PP$_m$: ʃɪˈtɜːnd (PP$_m$)) (PP$_n$: /tʊəˈmæn/ (PP$_n$)) (PP$_o$: /ʃɪhædˈmet /(PP$_o$)) (PP$_P$: /bɪˈfɔː/ (PP$_P$))] (IP$_j$))] (U$_I$))

Possible Output Component effects: (PP$_k$: /ɪntrəˈdjuːstəˈself/ (PP$_k$)); (PP$_l$: /təðɪˈɔ:dɪəns/ (PP$_l$)); (PP$_o$: /ʃɪdˈmet /(PP$_o$))

Figure 8: my analysis of example (30).

Instead, what is proposed in Figure 8 is to distinguish two Communicated Contents within the Discourse Act, (C$_I$) and (C$_J$). The presence of a Communicated Content can be tested by attempting to insert such adverbs as *reportedly*, which have been shown (e. g. Hengeveld, 2013, 18) to scope over Communicated Contents, but not Discourse Acts. The occurrence of such attested examples as (32), in which *reportedly* takes only the respective *after*-clause in its scope, encourages us to regard the analysis in Figure 8 as correct in this respect.

(32) *After she reportedly saw the man drop the dog, the woman rushed down the stairs and stabbed him in the thigh.*[4]

4 https://zh-cn.facebook.com/ABC10News/posts/2141686055873865, accessed 6 September 2018.

Another feature of the Interpersonal Level analysis concerns the status of *herself* in the first clause. With (R_I) referring to *Mary* and (R_J) referring to the audience, the analysis sees *herself* as not being referential but as forming part of the Subact of Ascription (T_I); i. e. *introduce oneself*, as used here, forms a single predicate of self-naming. In (33), by contrast, the reflexive pronoun is referential, with each of the Individuals introduced being analysed as a Subact of Reference:

(33) *After Mary introduced herself, her band and her manager to the audience, she …*

Returning to Figure 8, we note that the full valency of *introduce* is present at the Representational Level in the Configurational Property (f_q), triggering the presence of the reflexive pronoun at the Morphosyntactic Level.

At the Representational Level, the sample sentence is analysed as a single Propositional Content (p_i), consisting of an Episode (ep_i) corresponding to *she turned to a man she had met before*, which is modified by the Time $(t_j)_L$, i. e. (at a time) *after Mary introduced herself to the audience*, which itself contains an Episode (ep_k). Within (ep_i), corresponding to the restrictive relative clause *she had met before*, there is another Episode (ep_j). These are all Episodes because they display absolute time (namely Past); States-of-Affairs (e_i) and (e_k) are marked as Sim(ultaneous) but the State-of-Affairs (e_j) is marked as Ant(erior), reflecting the occurrence of the pluperfect *had met*, i. e. the meeting is anterior to the Past time identified by the Episode. Another salient property of the representational analysis is the treatment of the "conjunction" *after* and the "adverb" *before* as both belonging to the part-of-speech Ad, as explained in Section 3.1 above: whereas *after* (f_p) takes an Episode (ep_k) as its argument, *before* (f_n) has no argument.

Figure 8 gives two analyses for the Morphosyntactic Level. The first presents the sentence as a Linguistic Expression (Le_i), consisting of two clauses (Cl_i) and (Cl_j), the latter containing the relative clause (Cl_k). The "flatness" of the FDG approach to syntax is apparent throughout, for example in (Cl_i), which is treated as a configuration of Ad phrase (Adp_i), Noun phrase (Np_i), Verb phrase (Vp_i), Grammatical word (Gw_i) and an Ad phrase (Adp_j) which does show (minimal) internal hierarchical structuring. Note that Grammatical words (Gw_1) are introduced here, for example (Gw_i), already available in its phonological form, or (Gw_j), where DEF is a placeholder for the ultimate phonological representation as /ðiː/ or /ðə/. Affixes are also introduced here: the PAST suffix in English is again introduced as a placeholder, with its ultimate form as /d/, /t/ or /ɪd/ to be determined at the Phonological Level; for *met*, the phonological form of the single Verb word (Vw_l) will be drawn from a set of primitives at the Phonological Level.

The second analysis shows the position occupied by the various Clauses and Phrases. Here the initial clause (Cl_i) occupies position P^{pre}, reserved for elements of a Linguistic Expression that precede the P^I position, here held by the Subject of the main clause *she*. Within (Cl_i), P^I is occupied by (Adp_i: (Adw_i: /ˈɑːftər/$_{Ad}$ (Adw_i)) (Adp_i)), realizing not only the head of the Configurational Property (f_o) that heads the time modifier (t_j) but crucially also the semantic function L(ocation) assigned to (t_j) – it is this function that justifies giving this Episode modifier access to P^I. As a result, the Subject *Mary* is relegated to P^{I+1}, still in the initial domain but now in a relative position. The analysis of the relative clause shows how the FDG Morphosyntactic Level makes no use of empty positions. The matter of the object of *met* is resolved at the Representational Level, not here; similarly, the absence of a relative pronoun *whom* or complementizer *that* is dealt with at the language-specific interface between the Representational and Morphosyntactic Levels.

The Phonological Level analysis only shows the highest layers assumed here: U (Utterance), IP (Intonational Phrase), PP (Phonological Phrase); note the use of small caps symbols. The operators 'r' and 'f' on (IP_i) and (IP_j) indicate rising and falling contours respectively, reflecting the Interpersonal Level analysis of the Discourse Act as containing two Communicated Contents. The output of the Phonological Level (in practice much more detailed) is sent to the Output Component, at which various phonetic reduction processes may occur, e. g. delivering (PP_o: /ʃɪdˈmet /(PP_o)) 'she'd met'.

6 Conclusions

To conclude this chapter, it may be useful to consider a couple of issues that are currently preoccupying the community of FDG researchers. One of these concerns the interfaces between the various levels of analysis. While there is general agreement that FDG is devoted to formalizing function-to-form mappings and that the above-mentioned transparency principles favour bijective relations between formulation and encoding, there is also a desire to avoid duplicating information across levels. We already saw one example of possible duplication at the Morphosyntactic Level, where the limited layering to some extent recapitulates the hierarchies of the Interpersonal and Representational Levels. The possibility that the FDG approach to syntax should place even more emphasis on linearity is germane to the controversy sparked off by Frank et al. (2012), who question the need for hierarchy in syntax from the viewpoint of language use.

Another concern follows from our commitment to formalizing our analyses. In practice, this appears to have driven a regrettable wedge between FDG and many other types of usage-based, cognitive and/or constructionist work, despite shared adherence to the principles of functionalism and overlapping fields of interest. The FDG community has reacted by directing its publication policy towards generalist journals such as *Linguistics*, *Language Sciences* or *Pragmatics* in order to enter the mainstream. Nevertheless, it has continually proved necessary to explain the foundations of the theory before progressing to the presentation of analyses. It is hoped that the present chapter will contribute to increasing familiarity with FDG among linguists.

References

Aissen, Judith. 1992. Topic and focus in Mayan. *Language* 51. 43–80.

Bresnan, Joan. 2001. *Lexical Functional Syntax*. Oxford: Blackwell.

Butler, Christopher S. 2008a. Cognitive adequacy in structural-functional theories of language. *Language Sciences* 30. 1–30.

Butler, Christopher S. 2008b. Interpersonal meaning in the noun phrase. In Daniel García Velasco & Jan Rijkhoff (eds.), *The noun phrase in Functional Discourse Grammar*, 221–261. Berlin: Mouton de Gruyter.

Chafe, Wallace. 2018. *Thought-based linguistics*. Cambridge: Cambridge University Press.

Connolly, John H. 2014. The Contextual Component within a dynamic implementation of the FDG model: Structure and interaction. *Pragmatics* 24. 229–248.

Cornish, Francis. 2013. On the dual nature of the Functional Discourse Grammar model: Context, the language system/language use distinction, and indexical reference in discourse. *Language Sciences* 38. 83–98.

Culicover, Peter W. & Ray Jackendoff. 2005. *Simpler syntax*. Oxford: Oxford University Press.

Dik, Simon C. 1997. *The theory of Functional Grammar. 2 volumes*. Ed. by Kees Hengeveld. Berlin & New York: Mouton de Gruyter.

Evans, Nicholas 2003. *Bininj Gun-Wok: A pan-dialectal grammar of Mayali, Kunwinjku and Kune. 2 vols*. Canberra: Australian National University.

Everett, Daniel L. 2005. Cultural constraints on grammar and cognition in Pirahã: Another look at the design features of human language. *Current Anthropology* 46. 621–646.

Foley, William & Robert Van Valin Jr. 1984. *Functional syntax and universal grammar*. Cambridge: Cambridge University Press.

Frank, Stefan L., Rens Bod & Morten H. Christiansen. 2012. How hierarchical is language use? *Proceedings of the Royal Society B: Biological Sciences* 279. 4522–4531.

García Velasco, Daniel. 2016. A flexible lexicon for Functional Discourse Grammar. *Linguistics* 54(5). 907–945.

García Velasco, Daniel, Kees Hengeveld & J. Lachlan Mackenzie. 2012. Epilogue: Dynamic morphosyntax in Functional Discourse Grammar. *Language Sciences* 34(4). 491–498.

Genee, Inge. 2009. What's in a morpheme? Obviation morphology in Blackfoot. *Linguistics* 47(4). 913–944.

Giomi, Riccardo. 2014. Grammar, context and the hearer: A proposal for an addressee-oriented model of Functional Discourse Grammar. *Pragmatics* 24(2). 275–296.

Hengeveld, Kees. 1989. Layers and operators in Functional Grammar. *Journal of Linguistics* 25(2). 127–157.

Hengeveld, Kees. 1992. Parts of speech. In Michael Fortescue, Peter Harder & Lars Kristoffersen (eds.), *Layered structure and reference in a functional perspective*, 29–55. Amsterdam: John Benjamins.

Hengeveld, Kees. 2004. The architecture of a Functional Discourse Grammar. In J. Lachlan Mackenzie & María de los Ángeles Gómez-González (eds.), *A new architecture for Functional Grammar*, 1–21. Berlin & New York: Mouton de Gruyter.

Hengeveld, Kees. 2013. Parts-of-speech system as a basic typological determinant. In Jan Rijkhoff & Eva van Lier (eds.), *Flexible word classes: Typological studies of underspecified parts of speech*, 31–55. Oxford: Oxford University Press.

Hengeveld, Kees. 2017. A hierarchical approach to grammaticalization. In Kees Hengeveld, Heiko Narrog & Hella Olbertz (eds.), *The grammaticalization of tense, aspect, modality, and evidentiality: A functional perspective*, 13–38. Berlin: de Gruyter Mouton.

Hengeveld, Kees & Rafael Fischer. In prep. *A grammar of A'ingae* (*Cofán*/*Kofán*).

Hengeveld, Kees & Sterre Leufkens. 2018. Transparent and non-transparent languages. *Folia Linguistica* 52(1). 139–175.
Hengeveld, Kees & J. Lachlan Mackenzie. 2008. *Functional Discourse Grammar: A typologically-based theory of language structure*. Oxford: Oxford University Press.
Hengeveld, Kees & J. Lachlan Mackenzie. 2016. Reflections on the lexicon in Functional Discourse Grammar. *Linguistics* 54(5). 1135–1161.
Hengeveld, Kees & Gerry Wanders. 2007. Adverbial conjunctions in Functional Discourse Grammar. In Mike Hannay & Gerard Steen (eds.), *Structural-functional studies in English grammar: In honour of Lachlan Mackenzie*, 211–227. Amsterdam: John Benjamins.
Huddleston, Rodney & Geoffrey Pullum (eds.). 2002. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.
Keizer, Evelien. 2007. The lexical-grammatical dichotomy in Functional Discourse Grammar. *Alfa: Revista de Lingüística* 51(2). 35–56.
Keizer, Evelien. 2008. English prepositions in Functional Discourse Grammar. *Functions of Language* 15(2). 216–256.
Keizer, Evelien. 2015. *A Functional Discourse Grammar for English*. Oxford: Oxford University Press.
Keizer, Evelien. 2016. Idiomatic expressions in Functional Discourse Grammar. *Linguistics* 54(5). 981–1016.
Konopka, Agnieszka E. & Sarah Brown-Schmidt. 2014. Message encoding. In Matthew Goldrick, Victor Ferreira & Michele Miozzo (eds.), *Oxford handbook of language production*, 3–20. Oxford: Oxford University Press.
Kornfilt, Jaklin. 1997. *Turkish*. London: Routledge.
Levelt, Willem J. M. 1989. *Speaking*. Cambridge MA: MIT Press.
Lewis, Geoffrey L. 1967. *Turkish Grammar*. Oxford: Clarendon.
Mackenzie, J. Lachlan. 2013. Spatial adpositions between lexicon and grammar. In J. Lachlan Mackenzie & Hella Olbertz (eds.), *Casebook in Functional Discourse Grammar*, 67–93. Amsterdam & Philadelphia: John Benjamins.
Mackenzie, J. Lachlan. 2014. The Contextual Component in a dialogic FDG. *Pragmatics* 24(2). 249–273.
Mackenzie, J. Lachlan. 2016a. A first history of Functional Grammar. In Carlos Assunção, Gonçalo Fernandes & Rolf Kemmler (eds.), *History of linguistics 2014: Selected papers from the 13th International Conference on the History of the Language Sciences (ICHoLS XIII)*, Vila Real, Portugal, 25–29 August 2014, 233–245. Amsterdam: John Benjamins.
Mackenzie, J. Lachlan. 2016b. Functional linguistics. In Keith Allan (ed.), *Routledge handbook of linguistics*, 470–484. London: Routledge.
Nevins, Andrew, David Pesetsky & Cilene Rodrigues. 2009. Pirahã exceptionality: A reassessment. *Language* 85. 355–404.
Pezatti, Erotilde Goreti. 2014. *A ordem das palavras em português [Word order in Portuguese]*. São Paulo: Parábola: Editorial.
Pollard, Carl & Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. Chicago: University of Chicago Press.
Rijkhoff, Jan, Dik Bakker, Kees Hengeveld & Peter Kahrel. 1993. A method of language sampling. *Studies in Language* 17. 169–220.
Romero-Figueroa, Andrés. 1997. *A reference grammar of Warao*. Munich & Newcastle: Lincom Europa.
Shlonsky, Ur. 2010. The cartographic enterprise in syntax. *Language and Linguistics Compass* 4(6). 417–429.
Wolvengrey, Arok. 2011. *Semantic and pragmatic functions in Plains Cree Syntax*. Utrecht: LOT.

Stefan Müller and Antonio Machicao y Priemer

12 Head-Driven Phrase Structure Grammar

Abstract: Head-Driven Phrase Structure Grammar is a constraint-based theory. It uses features and values to model linguistic objects. Values may be complex, e. g. consist of feature values pairs themselves. The paper shows that such feature value pairs together with identity of values and relations between feature values are sufficient to develop a complete linguistic theory including all linguistic levels of description. The paper explains the goals of researchers working in the framework and the way they deal with data and motivate their analyses.

The framework is explained with respect to an example sentence that involves the following phenomena: valence, constituent structure, adjunction/modification, raising, case assignment, nonlocal dependencies, relative clauses.

1 General remarks

Head-Driven Phrase Structure Grammar (HPSG) was developed in the 1980s by Carl Pollard and Ivan Sag. Ivan Sag was one of the developers of Generalized Phrase Structure Grammar (Gazdar, Klein, Pullum and Sag, 1985), which strongly influenced HPSG. Carl Pollard worked in Categorial Grammar (Pollard, 1984), from where some ideas were brought into the new framework. Categorial Grammar is an approach that puts a lot of emphasis on lexical information and by adopting a lexical view a lot of the shortcomings of GPSG (Jacobson, 1987; Müller, 2016, Section 5.5) were avoided. With the advent of HPSG, research on GPSG came to an end almost entirely and most researchers switched to HPSG. Since 1993 annual conferences have been held rotating between the US, Europe, and Asia. The HPSG online bibliography at https://hpsg.hu-berlin.de/HPSG-Bib/ lists over 1500 papers published in journals, books, or conference proceedings.

The following sections deal with data, goals, tools and evaluation as they are used in the HPSG research community. These sections have a rather general character with the exception of Section 4, which deals with some of the formal foundations of HPSG. The formal foundations will be the basis of the analysis of an example sentence in Section 6.

Acknowledgements: We thank Bob Borsley for comments on an earlier version of this paper and Bob Levine and Steve Wechsler for discussion. Special thanks go to the editors of this volume and Sam Featherston for detailed comments on earlier versions of this paper. The students of the first author's MA class on grammar implementation also read this paper. We thank Jakob Wünsch for comments.

https://doi.org/10.1515/9783110540253-012

2 Data

There is no dogma concerning the kind of evidence that should be used in work within HPSG. Most of the early practitioners of HPSG were coming from a Chomskyan research tradition and hence, worked mainly based on introspection. Since HPSG was well-formalized from the beginning it was used in grammar implementations and these implementations were used in research prototypes or applications. These were built with respect to naturally occurring data like spoken dialogues for appointment scheduling in the Verb*mobil* project, which was running from 1992–2000 (Wahlster, 2000). Languages that are covered in Verb*mobil* are German (Müller and Kasper, 2000), English (Flickinger et al., 2000), and Japanese (Siegel, 2000). Other projects were set up with the goal of parsing parts of Wikipedia (Flickinger et al., 2010). So, computationally oriented researchers were working with corpus data right from the beginning. While more theoretically oriented papers started out with introspectively produced data, many researchers started using data from the web and those who were reflecting deeper about the possibility of reproducing the results of search queries and about accessibility of their data sources used corpora like the British National Corpus, Tiger (Brants et al., 2004) or COW (Schäfer and Bildhauer, 2012). Nevertheless, introspection might be a useful guide but it alone is not sufficient since certain phenomena are just not accessible through introspection. See Müller (2007) and Meurers and Müller (2009) on wrong claims about particle verbs, extraposition and subjacency, and apparent multiple frontings in German that were based on introspective data.

On the other hand, attested examples alone are not sufficient either. For instance the following example from a newspaper is ungrammatical since it contains both *das* and *dem*, where only *das* would be grammatical:

(1) * *Dagegen hatte nach dem Bekanntwerden des ersten Berichtsentwurfs*
there.against had after the release of.the first report.draft
nicht nur das dem Umweltbundesamt protestiert.[1]
not only the the Umweltbundesamt protested
Intended: 'Not just the Umweltbundesamt protested against this after the release of the first draft of the report.'

One can speculate how this error came about. Probably, it is due to some reformulation of the text. Unless such questions are the focus of inquiry, examples like (1) are not of interest and have to be filtered out. So a mix of methods is required: corpus data and introspection or corpus data and experiments. Since (1) is a clear case, introspection is sufficient here.

1 taz, 22.06.2017, p. 9 (The taz is a Berlin-based nation-wide newspaper.)

Apart from corpus data as used for example by Bildhauer (2011), all available evidence from experiments (e. g., speaker judgments, eye tracking data, and results from speeded reading tasks) are seen as relevant data. Since HPSG is a performance-compatible competence theory (Sag and Wasow, 2011), it can be paired with a performance model and hence experimental data can feed back into theory development. The performance model can be seen as an additional layer of constraints referring to the linguistic knowledge and hence certain structures may be ruled out for performance reasons rather than being ruled out by the theory itself. An example of the latter is the question of why there are no languages that form questions by reversing the order of words in a string, a question often asked in Mainstream Generative Grammar. The answer of those working in HPSG is that our short-term memory is just not large enough to do such complex computations, a fact that is independent of our linguistic knowledge. A linguistic competence theory does not have to explain the non-existence of such languages. To take another, more relevant example, consider Sag et al. (2007, 228), who argue that Subjacency and the Complex NP Constraint should not be part of a competence grammar.

While some HPSG analyses are developed without testing their psycholinguistic plausibility, the analyses do make certain predictions that are open for testing. For example, Wittenberg and Piñango (2011) and Wittenberg et al. (2014) examined various analyses of complex predicates and found that Goldberg's analysis (2003) makes wrong predictions while the one in Müller (2010) is compatible with the psycholinguistic findings.

3 Goals

The goal of research is similar to the goals of Construction Grammar and hence also includes many of the goals of Mainstream Generative Grammar (GB, Minimalism and variants thereof): we want to understand language as a cognitive system, we want to understand which properties are common to all languages and how languages may vary, we want to understand how natural language can be acquired and processed. The explorations are not limited to a core grammar in the Chomskyan sense since it is believed that the so-called periphery interacts in interesting ways with what is thought of as the core. There is interesting research on idioms (Soehn and Sailer, 2008; Sag, 2007; Richter and Sailer, 2009; Kay et al., 2015) and in fact the distinction between core and periphery does not play an important role in theorizing (Müller, 2014b).

Research in HPSG is not limited to syntax. Many papers address semantic phenomena and make the syntax-semantics interface explicit. HPSG theories are declarative statements about language and this linguistic knowledge can be used in various ways. One way is to find the meaning of a given utterance (parsing) and the other way is to find the phonological or orthographic representation for a given meaning (production or generation). Apart from work on syntax and semantics, there is work on phonology (Bird and Klein, 1994; Bird, 1995; Orgun, 1996; Höhle, 1999; Klein, 2000; Alexopoulou and Kolliakou, 2002), morphology (Riehemann, 1997; Crysmann and Bonami, 2016), information structure (Engdahl and Vallduví, 1996; Kuhn, 1996; Wilcock, 2005; De Kuthy, 2002; Paggio, 2005; Bildhauer, 2008; Bildhauer and Cook, 2010) and dialogue (Schlangen et al., 2003; Ginzburg and Cooper, 2004).

Since the work is formalized it can be implemented and used in computer applications. One field for applications is machine translation (Oepen et al., 2007), another one information extraction. For further details see the webpage of the DELPH-IN consortium.[2]

4 Tools

HPSG is a model-theoretic approach (Pullum and Scholz, 2001; Richter, 2007) and hence belongs to the family of constraint-based theories.[3] Linguistic theories are sets of constraints (mostly feature-value pairs) that constrain the number of linguistic objects licensed by the theory. Theories like HPSG are surface oriented, that is, there is no underlying structure from which another representation is derived. A sentence like (2a) is analyzed directly involving the words that are visible. That is (2a) is not derived from (2b). Neither is (2b) derived from (2c):

(2) a. *This book, Kim was given as a present.*
  b. *Kim was given this book as a present.*
  c. *Somebody gave Kim this book as a present.*

Of course, Chomsky (1957) was right in pointing out that simple phrase structure grammars are inappropriate for modeling linguistic phenomena since they cannot account for the fact that these sentences are related. HPSG has means of capturing these relations, but they do not involve different levels like Deep and Surface Structure, and they do not involve transformations of complete sentential structures (or their equivalent in Minimalist theories). Rather than employing a passive transformation or more general transformations that account for passive, lexical rules are used deriving passive participles (like *given* in (2a)) from word stems (see Section 4.10). Similarly, extraction phenomena like the fronting of *this book* in (2a) are modeled by establishing a relation between the fronted element and the head on which the fronted element depends, but this does not involve movement in the literal sense (see Section 4.9). In what follows, we nevertheless use terms like *extraction* and *scrambling* since these terms are

2 http://www.delph-in.net/wiki/index.php/Background, 2018-06-23.

3 These theories are sometimes also called unification-based theories but constraint-based is the more general term.

established in the literature. But as will become clear these phenomena are not described with respect to several trees but it is always just one linguistic structure that is assumed for one utterance.

The fact that HPSG is surface-oriented and transformationless is a huge advantage when it comes to psycholinguistical plausibility (Fodor, Bever and Garrett, 1974, 320–328). HPSG just assigns complex categories to linguistic objects (for instance 'complete nominal projection in the accusative') and states the dependency between elements (for instance head and its argument). These entities and the relations are directly observable. No statements regarding the processing order of constraints are made in HPSG theories. HPSG theories are just declarative statements about linguistic objects. This has the advantage that the linguistic knowledge can be paired with various processing models and processing regimes. The linguistic knowledge can be used for parsing utterances and for generation/production as well. HPSG grammars can also account for fragmentary input (Pullum and Scholz, 2001, Section 3.2). This is not the case for some of its competitors, e. g., all those models that see languages as (infinite) sets that are enumerated by a grammar (Chomsky and Miller, 1963, 283). See Sag and Wasow (2011), Müller (2018c) and Wasow (2019) for further discussion.

4.1 Features, values, structure sharing and relational constraints

HPSG makes use of a very small number of descriptive tools: features that have values of a certain type. Values of features may be complex. For instance, an AGREEMENT feature may have a complex value providing information about case, gender and number. Types are organized in hierarchies, which makes it possible to capture generalizations by positing abstract types and more specific subtypes (see also Section 4.6). Values of features can be identified with values of other features (structure sharing, explained below in more detail) or they can be related to other features by relational constraints (explained below). As will be shown in the remainder of the paper, this is sufficient to express everything one has to say about language: roots, stems, words, lexical rules, phrases can all be described using feature-value pairs.

HPSG assumes feature structures as models of linguistic objects.[4] These feature structures are described by feature descriptions, which are also called *attribute value matrix* (AVM). Such AVMs consist of feature value pairs. The values can be atomic (e. g., *sg* and *pl*) or feature descriptions. Every feature structure is of a certain type. Types are written in italics. They are ordered in hierarchies with the most general type at the top of the hierarchy and the most specific types at the bottom. Figure 1 shows an example hierarchy for the type *number* and its subtypes.

4 Feature structures are usually depicted as graphs (Pollard and Sag, 1994, 16–17; Richter, 2007). Due to space limitations we do not give an example here but provide feature descriptions only, which are used to formulate theories about possible feature structures.

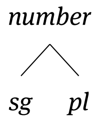


Figure 1: Subtypes of *number* in a grammar of English.

Types in a model of a linguistic object are maximally specific, that is, a noun in a model of an actual utterance has a NUMBER value that is *sg* or *pl*. The linguist develops theories that describe possible feature structures. In contrast to feature structures, feature descriptions can be partial. For instance it is not necessary to specify a NUMBER value for the word *sheep* since *sheep* can be used both in singular and in plural NPs.

(3) a. *one sheep*
  b. *two sheep*

(4) is an example of a complex AVM. It shows a description of a referential index as it is used in the semantic representation of *man*:

(4) $$\begin{bmatrix} \textit{ref} & \\ \textsc{per} & \textit{3} \\ \textsc{num} & \textit{sg} \\ \textsc{gen} & \textit{mas} \end{bmatrix}$$

There are nouns like *cousin* that are underspecified with respect to their gender. They could either have male or female gender as is shown by examples with coreferring pronouns:

(5) *I met Peter's cousin. She/he is very tall.*

There are two ways to specify this. The first is to use a disjunction (*fem* ∨ *mas*) and the second is to use a common supertype for *fem* and *mas*. While both solutions are equivalent when it comes to models, descriptions without disjunctions are often more compact, which is why representation in the latter is preferred over the former. Figure 2 shows the type hierarchy for gender with a special type for objects that can be either

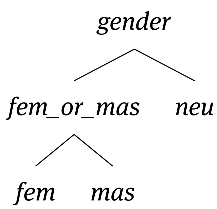


Figure 2: Subtypes of *gender* in a grammar of English.

fem or *mas*. The value of GEN in the description of the referential index of *cousin* is *fem_or_mas*. (See Müller (2016, Section 14.3) for alleged problems for model-theoretic syntax with cases like this.)

One very important part of the formalism is *structure sharing*. It is used to express that information in feature structures is identical, that is, token-identical rather than just type identical. Structure sharing is indicated by boxed numbers in feature descriptions. An identical number at several places in an AVM expresses the fact that the respective values are identical.

To give an example of structure sharing, let us consider case agreement in German noun phrases:

(6) a. *der* *Mann*
the.NOM man.NOM ∨ DAT ∨ ACC

b. *des* *Mannes*
the.GEN man.GEN

The determiner has to agree with the noun in case. *Mann* can be nominative, dative or accusative. *Mannes* is in the genitive. The form of the determiner has to be *des* if the noun is genitive. So the case specification for *Mann* would be *nom* ∨ *dat* ∨ *acc* or an equivalent type. One could be tempted to suggest that the noun *Mann* has a case value that is *nom* ∨ *dat* ∨ *acc* and that it has to be combined with a determiner that has the CASE value *nom* ∨ *dat* ∨ *acc*. The following AVM depicts this without assuming anything about the theory that will be developed later:

$$(7)\quad \begin{bmatrix} \text{DETERMINER} & \begin{bmatrix}\text{CASE} & \mathit{nom} \vee \mathit{dat} \vee \mathit{acc}\end{bmatrix} \\ \text{NOUN} & \begin{bmatrix}\text{CASE} & \mathit{nom} \vee \mathit{dat} \vee \mathit{acc}\end{bmatrix} \end{bmatrix}$$

But a specification of the values as in (7) is not sufficient since when an actual determiner is chosen (*der* or *des*), the case of the complete NP is unambiguously determined, but this is not reflected in (7). With the setting above, we would get the following for inserting the determiner *der*:

$$(8)\quad \begin{bmatrix} \text{COMBINATION} & \begin{bmatrix}\text{CASE} & ??\end{bmatrix} \\ \text{DETERMINER} & \begin{bmatrix}\text{CASE} & \mathit{nom}\end{bmatrix} \\ \text{NOUN} & \begin{bmatrix}\text{CASE} & \mathit{nom} \vee \mathit{dat} \vee \mathit{acc}\end{bmatrix} \end{bmatrix}$$

The disjunctive specification of the determiner is resolved to nominative, but the other disjunction is not affected. Furthermore it is unclear what the case of the whole combination would be. If the case value of the whole NP is determined by the head (the noun alone), it would be *nom* ∨ *dat* ∨ *acc* and this would mean that the whole phrase *der Mann* could be used as a dative or an accusative object. Obviously, this is not what is wanted. What is needed instead is that the case of the determiner is token-identical to the case of the noun (agreement) and to the case of the complete phrase (projection of feature values). This is ensured by structure sharing:

(9)
$$\begin{bmatrix} \text{COMBINATION} & \begin{bmatrix}\text{CASE } \boxed{1}\end{bmatrix} \\ \text{DETERMINER} & \begin{bmatrix}\text{CASE } \boxed{1}\end{bmatrix} \\ \text{NOUN} & \begin{bmatrix}\text{CASE } \boxed{1}\end{bmatrix} \end{bmatrix}$$

With such a setting the case of the NP *der Mann* is nominative and the one of *des Mannes* is genitive, as expected. Note also that the case of *die Frau* is *nom* ∨ *acc* since *Frau* is compatible with all four cases and *die* is *nom* ∨ *acc*. Depending on the governing verb this disjunction can be resolved to either *nom* or *acc*.

While structure sharing is the most important expressive means in HPSG there is one extension of the basic formalism that plays a crucial role in most HPSG analyses: *relational constraints*. Relational constraints are used to relate several values in a feature structure to each other. The relational constraint that is used most often in HPSG is *append* (⊕), which is used to concatenate two lists. The Schema 1, which will be discussed in Section 4.7, is an example for an application of such a constraint.

4.2 Descriptive levels, feature geometry and modularization

The following AVM is the description of the word *man*. The grouping of features – the so-called feature geometry – is the one of Pollard and Sag (1994) and Sag (1997).[5] (10) shows parts of the lexical item for *man*:

(10) lexical item for *man*:

$$\begin{bmatrix} \textit{word} & \\ \text{PHONOLOGY} & \langle \text{m,æ,n} \rangle \\ \text{SYNTAX-SEMANTICS} & \begin{bmatrix} \textit{synsem} & \\ \text{LOCAL} & \begin{bmatrix} \textit{local} & \\ \text{CATEGORY} & \begin{bmatrix} \textit{category} & \\ \text{HEAD} & \textit{noun} \\ \text{SPR} & \langle \text{DET} \rangle \end{bmatrix} \\ \text{CONTENT} & \ldots \begin{bmatrix} \textit{man} & \\ \text{INST} & x \end{bmatrix} \end{bmatrix} \\ \text{NONLOCAL} \ldots & \end{bmatrix} \end{bmatrix}$$

5 There are various HPSG variants around that differ mainly in the way features are grouped. One example is Sign-Based Construction Grammar (Sag, 2012). For a discussion of SBCG's feature geometry see Müller (2018c, Section 10.6.2).

The first feature value pair describes the phonological form of the word. The value of PHONOLOGY is a list of phonemes. For reasons of readability usually the orthographic form is given in HPSG papers and phonological structure is omitted, but see Bird and Klein (1994), Bird (1995), Orgun (1996), Höhle (1999), Klein (2000), and Alexopoulou and Kolliakou (2002) for phonological analyses. The second feature is SYNTAX-SEMANTICS (SYNSEM) and its value is a description of all properties of a linguistic object that are syntactically and semantically relevant and can be selected by other heads. Information that is locally relevant (LOCAL) is distinguished from information that plays a role in non-local dependencies (NONLOCAL, see Section 4.9). Syntactic information is represented under CATEGORY (CAT) and semantic information under CONTENT (CONT). The example shows the HEAD value, which provides information about all syntactic aspects that are relevant for the external distribution of a maximal projection of a lexical head. In particular the part of speech information (*noun*) is represented under HEAD. As well as information regarding the head features, valence information also belongs under CAT. The example shows the SPR feature, which is used for the selection of a specifier (see Section 4.3 for details on valence).

The AVM in (10) shows a description of phonological, morpho-syntactic, and semantic aspects of a word. But of course other aspects can be and have been described by feature value pairs as well. For instance Engdahl and Vallduví (1996), Kuhn (1996), Wilcock (2005), De Kuthy (2002), Paggio (2005), Bildhauer (2008), and Bildhauer and Cook (2010) show how information structure can be modeled in HPSG in general and how the interaction between phonology, syntax, semantics and information structure can be captured in a constraint-based setting. For general discussion of interfaces between the linguistic levels of description see Kuhn (2007).

As is clear from looking at (10), information about the descriptive linguistic levels is represented in one structure. This makes it possible to connect syntax and semantics (see Section 4.3), phonology and information structure (Bildhauer, 2008), and syntax and information structure (Bildhauer and Cook, 2010), phonology and semantics (Halliday, 1970) and whatever other descriptive levels have to be connected. Since all information is in the same structure, HPSG is compatible with psycholinguistic findings that tell us that all available information is processed in parallel (Tanenhaus et al., 1995). This sets HPSG apart from other models like GB and Minimalism that assume that there are post-syntactic modules like Phonological Form and Logical Form. Minimalism has a conception that differs from the GB architecture (Richards, 2015, 812, 830) but it is psycholinguistically as implausible as the GB model for the same reason. Language processing is not a bottom up combination with shipping complete phrases or phases to the interfaces. For a discussion of psycholinguistics from an HPSG perspective see Wasow (2019).

In principle, all the information in (10) could be provided in a simple, unstructured list of feature value pairs (as is done in Minimalism, for instance). However, having phonology, syntax, semantics and information structure in different parts of the structure provides a cleaner representation. Furthermore, the specific groupings of information are motivated by the need to share this information (see Section 4.1 on structure sharing). As will be shown in Section 4.7, the SYNSEM value of an argument will be structure shared with the respective representation in the valence list of its head. Similarly the information under LOCAL is shared between an extracted element and the place where it is missing (see Section 4.9). Everything under CAT (valence, part of speech, case, ...) is identified in symmetric coordinations (Pollard and Sag, 1994, 202) and finally, all the head features under HEAD are projected from head daughters to their mothers. In principle, it would be possible to share single values of an unstructured list of features but having these groupings allows for a more general treatment: nouns share other properties with their projections than verbs. A noun has a certain case and the respective value is relevant for the whole nominal projection. Similarly the verb form (*fin* = *finite*, *bse* = infinitive without *to*, *inf* = infinitive with *to*, *pas* = passive) is relevant for the maximal verbal projection. (11) shows two example HEAD values for nouns and verbs:

(11) a. $\begin{bmatrix} \textit{noun} \\ \text{CASE} \;\; \textit{gen} \end{bmatrix}$ b. $\begin{bmatrix} \textit{verb} \\ \text{VFORM} \;\; \textit{pas} \end{bmatrix}$

With such an encoding of information it is possible to provide general constraints for head information without making recourse to specific features and their values. See Section 4.5 on the Head Feature Principle.

4.3 Valence and linking

The previous subsections dealt with the basic formal apparatus that is used in HPSG and made some general remarks about foundational assumptions. In what follows, we will look at the analysis of the example sentence in (12) to explain further basic assumptions.[6]

(12) *After Mary introduced herself to the audience, she turned to a man that she had met before.*

As was said in the introduction, HPSG is a lexicalist theory. So, much information about the combinatorial potential of a head is represented in its lexical item. For instance the verbs in (12) can take two NPs, an NP and a PP, or two NPs and a PP. The required form of the arguments is described in a list, the so-called argument structure list (ARG-ST list). (13) provides some prototypical examples:

(13)

	ARG-ST
a. meet	⟨ NP, NP ⟩
b. turn	⟨ NP, PP[*to*] ⟩
c. introduce	⟨ NP, NP, PP[*to*] ⟩

6 This sentence was provided as a shared task by the editors of this volume.

NP and PP[*to*] are abbreviations. They stand for AVMs describing fully saturated nominal or prepositional objects. Square brackets are used in such abbreviations for specification of some values like case of NPs or the form of the preposition in PPs.

The elements on the argument structure list are ordered according to their obliqueness (Keenan and Comrie, 1977). That is, subjects are ordered before primary objects before secondary objects before obliques. Since the order is fixed one can use this list to establish the particular linking patterns between arguments and the semantic roles they fill. For example, the first NP in the ARG-ST list of *turn* is linked to the agent (ARG1 in (14)) and the PP is linked to the person or object that is turned to (ARG2 in (14)).

(14) Linking for the verb *turn*:

$$\begin{bmatrix} \text{CAT}|\text{ARG-ST} & \left\langle \text{NP}_{\boxed{1}}, \text{PP}[\textit{to}]_{\boxed{2}} \right\rangle \\ \text{CONT}|\text{RELS} & \left\langle \begin{bmatrix} \textit{turn} \\ \text{ARG1} & \boxed{1} \\ \text{ARG2} & \boxed{2} \end{bmatrix} \right\rangle \end{bmatrix}$$

We do not use feature names like AGENT and THEME in the AVMs but rather more abstract features like ARG1. This is compatible with Dowty's approach (1991), which uses proto-roles.

For SVO languages it is useful to distinguish the arguments that are realized before the verb from those that follow it. The respective elements are represented in different valence lists. We assume that the subject (the first element of the ARG-ST list) is represented in a list called SPECIFIER list and that all other arguments are represented as elements of the COMPLEMENTS list.[7] These mappings are language-dependent. For example, for German, all arguments of finite verbs are mapped to COMPS (Pollard, 1996, 295–296, Kiss, 1995, 80).[8] (15) shows the respective mapping for *turn*:

(15) Mapping from ARG-ST to SPR and COMPS for verbs like *turn*:

$$\begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{SPR} & \langle \boxed{1} \rangle \\ \text{COMPS} & \langle \boxed{2} \rangle \\ \text{ARG-ST} & \langle \boxed{1}\,\text{NP}, \boxed{2}\,\text{PP}[\textit{to}] \rangle \end{bmatrix} \end{bmatrix}$$

7 Borsley (1987) argues for a SUBJ, a SPR, and a COMPS feature. We follow Borsley (1989) and the German tradition and assume a head feature SUBJ that is used for control and raising constructions (Pollard, 1996, 295–296, Kiss, 1995, Section 3.1.1, Müller, 1999b, Section 1.7). Under this assumption, SUBJ is not a valence feature, a head cannot be combined with anything in SUBJ. We assume that those subjects that can be combined with their head are either in SPR or in COMPS, depending on the language (Müller, 2018b).

8 German is a V2+SOV language. Hence, orders in which the subject appears in front of the verb are analyzed as extraction structures in which the subject is fronted. (Note: we are using the term *fronted* here since this is a handy way to describe German. In the analysis there is no underlying sentence from which something is fronted. See Section 4.9 on nonlocal dependencies.)

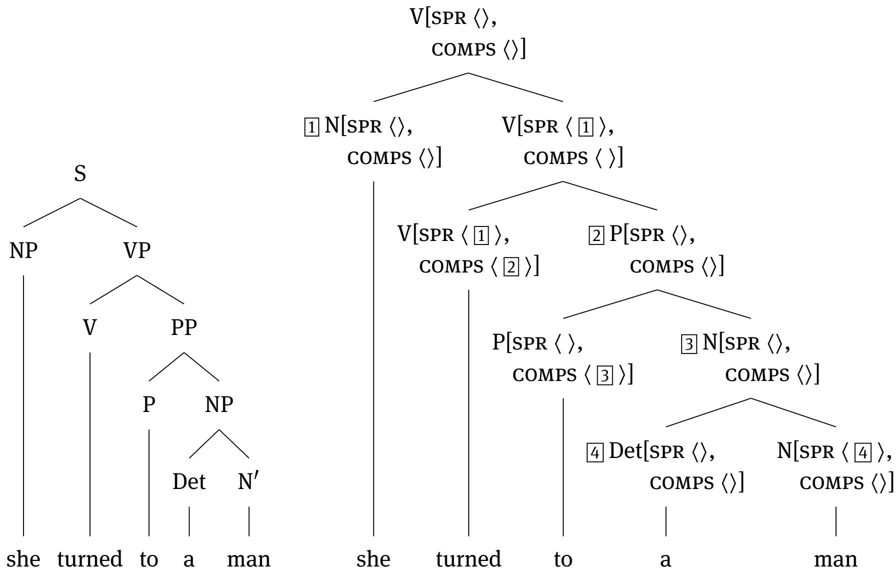


Figure 3: Analysis of *she turned to a man*.

The structure sharing $\boxed{1}$ indicates that the first element of the ARG-ST list is identical to the element in SPR and the $\boxed{2}$ indicates that the second element in ARG-ST is identical to the element in the COMPS list. SPR is used as well for nouns selecting a determiner (see (10)). Figure 3 shows how *she turned to a man* is analyzed. The left figure shows the analysis with traditional abbreviations and the right one shows the individual features for valence and part of speech. An N′ is a linguistic object of category *noun* that does not select any complements but needs a specifier. A preposition is a linguistic object that selects an NP, that is, a linguistic object of category *noun* that is complete as far as valence is concerned. The abbreviation VP corresponds to a verbal projection of a verb with an empty COMPS list and one element in the SPR list. If the element in SPR is saturated as well, we get an S. Looking at the analysis of *a man* we see that the description in the SPR list of the noun is identified with the determiner that is combined with the noun ($\boxed{4}$). Elements combined with a lexical head selecting them are not represented in the valence list of the resulting mother node. All other elements in valence lists are passed up. For instance *turned* selects both for a PP and an NP. The PP is combined with the verb first so that the description of the PP is not contained in the mother node (but see Section 6.3 for a modification). But the NP requirement is passed up: *turned to a man* selects an NP via its SPR list. After combination with *she* we get a fully saturated phrase, a maximal projection, that is, something with empty SPR and COMPS list.

Note that this representation of valence avoids unary projections as they are common in $\overline{\text{X}}$ theory: the pronoun *she* is just an NP without the intermediate projections from N^0 to $\overline{N}$ to NP. Similarly, an intransitive verb in English can be treated as a VP.[9]

In this paragraph and throughout the paper, the combination of items is described in a bottom-up way, but it is important to note that all statements are purely declarative, which means that there is no order in which constraints have to apply.

In the introduction, it was mentioned that in HPSG everything is done with feature value pairs, that is, without trees and phrase structure rules. How this is achieved will be explained in the next subsection.

4.4 Constituent Structure

While other theories that are similar to HPSG in using feature value pairs for describing complex categories use special phrase structure rules to model constituent structure (Bresnan and Kaplan, 1982), HPSG uses feature descriptions also for constraints on constituent structure. For example, the structure for *a man* can be represented by using features whose values correspond to the daughters in the tree. (16) shows parts of the structure for *a man*:

(16) $$\begin{bmatrix} \textsc{phon} & \langle\, a, man \,\rangle \\ \textsc{dtrs} & \left\langle \begin{bmatrix}\textsc{phon} & \langle\, a \,\rangle\end{bmatrix}, \begin{bmatrix}\textsc{phon} & \langle\, man \,\rangle\end{bmatrix} \right\rangle \end{bmatrix}$$

Note that HPSG differs from many theories in that the phonology is represented at the mother nodes. So HPSG is not like other theories where just the leaves in a tree are concatenated. Rather every linguistic object has its own PHON value. This makes it possible to specify constraints on phonology that are dependent on structure without assuming that these phonological constraints are somehow external or post-syntactic.

In addition to what is given in (16), structure sharing is used to point to the daughter that contains the head:

(17) $$\begin{bmatrix} \textsc{phon} & \langle\, a, man \,\rangle \\ \textsc{head-dtr} & \boxed{1} \\ \textsc{dtrs} & \left\langle \begin{bmatrix}\textsc{phon} & \langle\, a \,\rangle\end{bmatrix}, \boxed{1}\begin{bmatrix}\textsc{phon} & \langle\, man \,\rangle\end{bmatrix} \right\rangle \end{bmatrix}$$

The so-called head daughter is the daughter that contains the head. In the case of *a man* this is simply the noun *man* but for [[*introduced herself*] *to the audience*] the head daughter would be *introduced herself* since this phrase contains the head *introduced*.

9 Pollard and Sag (1994) and Ginzburg and Sag (2000, 34, 364) assume that a lexical verb is projected to the VP level in any case. See Müller (2013b, 935) for some discussion.

Both flat (Pollard and Sag, 1987, 1994; Sag, 1997; Ginzburg and Sag, 2000) and binary branching (Kiss, 1995; Meurers, 1999a; Kathol, 2001; Müller, 2002, 2018b) structures have been suggested in the literature. We will assume binary branching structures in this paper. The basic combinatorial schemata, which are introduced in Section 4.7, are similar to forward and backward application in Categorial Grammar (Ajdukiewicz, 1935; Steedman and Baldridge, 2006) and to Merge in Minimalism (Chomsky, 1995). See Müller (2013b) for a detailed comparison and some remarks concerning the history of ideas. The reason for assuming binary branching structures is that this makes it possible to assume the same set of schemata for head-argument combinations for many languages (if not for all) and hence allowing to capture crosslinguistic generalizations. Note though that the representation of daughters is sufficiently general to allow for flat structures. The N-P-N construction, suggested by Jackendoff (2008) to analyze phrases like *student after student*, is an example where we would assume flat structures with more than two daughters (Bargmann, 2015; Müller, 2019).

4.5 Principles: Implicational constraints

HPSG publications often contain prose statements stating principles. One such principle is the Head Feature Principle which says that in headed structures the head features of the head daughter are identical to the head features of the mother. This principle is formalized by an implicational constraint:

(18) *headed-phrase* ⇒

$$\begin{bmatrix} \text{SYNSEM}|\text{LOC}|\text{CAT}|\text{HEAD}\ \boxed{1} \\ \text{HEAD-DTR}|\text{SYNSEM}|\text{LOC}|\text{CAT}|\text{HEAD}\ \boxed{1} \end{bmatrix}$$

(18) will be explained in more detail below, what is important here is the formal aspect: HPSG can formulate implicational constraints that are to be interpreted as logical, that is, if the left-hand side of the implication holds, the right-hand side must hold as well. In (18) the left-hand side is a single type (*headed-phrase*) but in principle a complex description could be used at the left-hand side as well.

As mentioned above, types are organized in hierarchies (see also Section 4.6). An implication such as (18) holds for all structures of type *headed-phrase* and this includes of course all subtypes of *headed-phrase*.

Note that the constraint above does not entail that all structures are headed, it only states that certain constraints must hold for structures of type *headed-phrase*. The implicational constraint does not say anything about structures of another type.

4.6 Inheritance hierarchies

In Section 4.1, we explained type hierarchies and showed how they may be useful for specifying features and leaving them underspecified. But this is not the only advantage of using types in a theory. Types play an important role for capturing generalizations. Figure 4 shows a type hierarchy for the subtypes of *sign*. *sign* is the most general type for feature structures of linguistic objects. Signs can be of type *stem*, *word* or *phrase*.

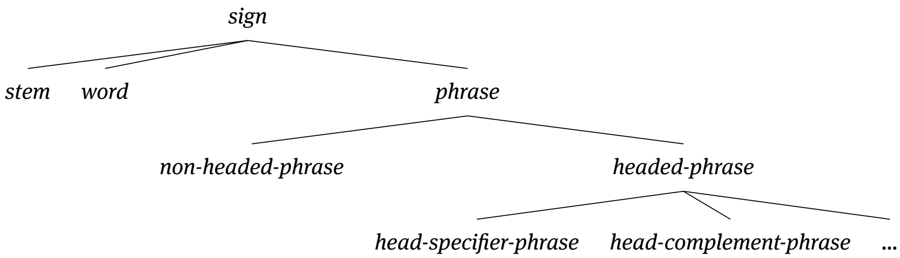


Figure 4: Type hierarchy for *sign*: all subtypes of *headed-phrase* inherit constraints.

Types are associated with features. For instance, feature structures of type *sign* always have a PHON value and a SYNSEM value. Signs of type *phrase* have a DTRS value in addition to the features introduced by *sign* and phrases that are of type *headed-phrase* have a HEAD-DTR in addition to everything that *phrase* has. Besides the introduction of features at certain types, values of features can be specified. So type hierarchies can be used to capture generalizations over certain linguistic objects. They can be used to capture lexical generalizations as well as to classify phrases. Construction Grammar is well-known for using inheritance hierarchies for capturing generalizations (Goldberg, 1995; Croft, 2001) but HPSG uses inheritance since the very first HPSG paper (Flickinger, Pollard and Wasow, 1985). The early work on inheritance was work about the lexicon since HPSG is a lexicalist framework but later this was extended to phrasal types (Sag, 1997, 2010).

4.7 Head-Argument Schemata

With the kind of representation of constituent structure introduced in Section 4.4, we can now formalize the treatment of valence depicted in Figure 3. The following schema licenses Head-Complement structures:

Schema 1 (Head-Complement Schema [preliminary])
head-complement-phrase $\Rightarrow$

$$\begin{bmatrix} \text{SYNSEM|LOC|CAT|COMPS } \boxed{1} \\ \text{HEAD-DTR| } \boxed{2} \text{ SYNSEM|LOC|CAT|COMPS } \left\langle \boxed{3} \right\rangle \oplus \boxed{1} \\ \text{DTRS } \left\langle \boxed{2}, \begin{bmatrix} \text{SYNSEM } \boxed{3} \end{bmatrix} \right\rangle \end{bmatrix}$$

The schema expresses constraints on structures of type *head-complement-phrase*.[10] It states that the COMPS list of the head daughter is split into two parts: a list with one element ($\langle \boxed{3} \rangle$) and the remainder of the list ($\boxed{1}$).[11] The COMPS list contains descriptions of the syntactic and semantic properties of arguments. One such description is identified with the SYNSEM value of the second element in the DTRS list, which is the non-head daughter. The remainder of the list ($\boxed{1}$) is identified with the COMPS list of the whole phrase. The SPR list of the head daughter is not affected in head-complement phrases and hence the SPR value of the mother node is identical to the SPR value of the head daughter. This is not shown in the schema. Since this passing on of the SPR value holds for further types of phrases (e. g., head-adjunct phrases), the constraints are formulated as constraints on a supertype from which *head-complement-phrase* inherits.

The head daughter ($\boxed{2}$) is identified with the first element of the DTRS list. The second element of the DTRS list corresponds to one element of the COMPS list of the head daughter ($\boxed{3}$). The DTRS list is assumed to be ordered in the way the elements are serialized. Therefore, the PHON value of the mother node is the concatenation of the PHON values of the daughters. In English, the complements of verbs, nouns and adjectives follow the head, but in languages like German and Dutch the complements follow nouns and prepositions but they precede adjectives and verbs. One of the cases where one could claim a head-final order in English is the postposition *ago* (cf. the preposition *in*):

(19) a. *one year ago*
b. *in one year*

In order to account for the two possibilities (19a) vs. (19b), one could simply state another version of the Head-Complement Schema or one could assume a more abstract representation of the schema, one that is neutral with respect to serialization of the daughters. The more abstract version is provided as Schema 2.

Schema 2 (Head-Complement Schema)
head-complement-phrase ⇒

$$\begin{bmatrix} \text{SYNSEM|LOC|CAT|COMPS } \boxed{1} \\ \text{HEAD-DTR|SYNSEM|LOC|CAT|COMPS } \langle \boxed{2} \rangle \oplus \boxed{1} \\ \text{NON-HEAD-DTRS } \left\langle \begin{bmatrix} \text{SYNSEM } \boxed{2} \end{bmatrix} \right\rangle \end{bmatrix}$$

This schema does not constrain the order of the daughters. Since head daughter and non-head daughter are represented as values of two different features nothing is said

10 The same kind of implicational constraint is used for stating principles, but although principles and schemata look similar, principles usually are defined as constraints on more general types. For example, the type *headed-phrase* is a supertype of *specifier-head-phrase*, *head-complement-phrase*, *filler-head-phrase* and so on. See Sag (2010, 533) for an elaborate type hierarchy of English clause types.
11 The alternative is to combine a head with all of its complements in one go (Ginzburg and Sag, 2000, 33–34). The result of this alternative is a flat structure (see Section 4.4).

about the order. One can then assume two different subtypes: one in which the head daughter is the first element in the DTRS list and the non-head daughter the second one and another subtype in which the head daughter is the second element in the DTRS list and the non-head daughter is the first one. The latter version is used to analyze orders like (19a).[12]

In Section 4.3, two valence features SPR and COMPS were introduced. We provided the schema for head-complement phrases above and a parallel schema for specifier-head combinations is given as Schema 3:

Schema 3 (Head-Specifier Schema)
head-specifier-phrase ⇒

$$\begin{bmatrix} \text{SYNSEM|LOC|CAT|SPR } \boxed{1} \\ \text{HEAD-DTR|SYNSEM|LOC|CAT} \begin{bmatrix} \text{SPR} & \boxed{1} \oplus \langle \boxed{2} \rangle \\ \text{COMP} & \langle\rangle \end{bmatrix} \\ \text{NON-HEAD-DTRS} \left\langle \begin{bmatrix} \text{SYNSEM } \boxed{2} \end{bmatrix} \right\rangle \end{bmatrix}$$

The last element of the SPR list is realized as the non-head daughter. The remaining list is passed up to the mother node. Note that the non-head daughter is taken from the end of the SPR list, while the non-head daughter in head-complement phrases is taken from the beginning. For heads that have exactly one specifier this difference is irrelevant, but in the analysis of object shift in Danish suggested by Müller and Ørsnes (2013b), the authors assume multiple specifiers and hence the difference in order of combination is relevant.

The COMPS values of mother and head daughter are identified in specifier-head phrases, as it is the case for the SPR value in head-complement phrases. The respective constraints are inherited from a supertype and are not given here. The head daughter's COMPS value is specified to be the empty list. This ensures that all complements are combined with the head first to form a $\overline{\text{N}}$ or a VP and the specifier(s) are combined with the respective projection after the combination with the complements.

Since specifiers always precede their heads, we could also have provided the schema with reference to DTRS rather than mentioning NON-HEAD-DTRS.

With the two schemata above and appropriately specified lexical items, we almost have everything that is needed for the analysis depicted in Figure 3. What has not been explained yet is how the part of speech information at the mother node of a tree is constrained. This is ensured by the Head Feature Principle, which was introduced in Section 4.5. Since the part of speech information is part of the HEAD value of signs (see (11)), it is passed up to mother nodes in syntactic structures. Depending on the part of speech, other information is passed up to the mother along the head path. For

12 Due to space limitations scrambling cannot be explained here. In order to account for languages with freer constituent order, the order in which items from the COMPS list are combined with their head is relaxed. For details see Müller (2016) or Müller (2015a).

example, verbs project information about the form of the verb, that is, the maximal projection of a verb contains information on whether the verb inside the sentence is finite, a perfect participle or some other form of the verb. Information about the form of the preposition is also passed up in order to make it selectable by governing verbs. Similarly, for languages with case inflection at nouns the case information is projected to the NP node.

4.8 Adjunction

Our example sentence in (12) contains two adjuncts: the adverbial clause *after Mary introduced herself to the audience* is attached to the remainder of the sentence and *before* is attached to *met*. The basic technique to describe head-adjunct combinations is similar to what we saw about head-argument combinations. Adjuncts are functors and have a special feature (MODIFIED, MOD) whose value is a description of heads they can combine with. For instance, *before* can modify VPs. The MOD feature is a head feature, that is, it is projected along the head path. (20) gives the lexical item of *before*:

(20) CAT value of the adverb *before*:

$$\begin{bmatrix} \text{HEAD} & \begin{bmatrix} \textit{adv} \\ \text{MOD} \;\; \text{VP} \end{bmatrix} \\ \text{SPR} & \langle\rangle \\ \text{COMPS} & \langle\rangle \end{bmatrix}$$

This lexical item for *before* should be contrasted with the preposition *before* as it is used in (21):

(21) *She met a man before the meeting.*

The *before* that is used in (21) selects for an NP and only after the combination of *before* and *the meeting* the complete phrase may modify a VP.

(22) CAT value of the preposition *before*:

$$\begin{bmatrix} \text{HEAD} & \begin{bmatrix} \textit{prep} \\ \text{MOD} \;\; \text{VP} \end{bmatrix} \\ \text{SPR} & \langle\rangle \\ \text{COMPS} & \langle \text{ NP } \rangle \end{bmatrix}$$

Since the information about the item that can be modified is part of the HEAD value, it is ensured that this information is also present at projections of *before*, that is, the PP *before the meeting* has the same MOD value as the preposition *before*.

Figure 5 shows the analysis of the example in (23):

(23) *She met him before.*

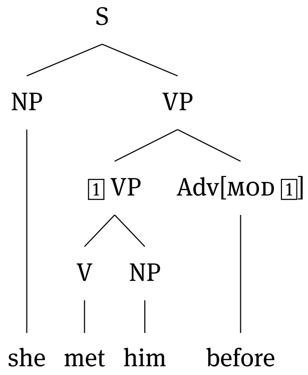


Figure 5: Analysis of sentence with adjunct.

The MOD value of the adverb is identified with the head daughter. Head-adjunct structures are licensed by the following schema:

Schema 4 (Head-Adjunct Schema)
head-adjunct-phrase ⇒

$$\begin{bmatrix} \text{HEAD-DTR}|\text{SYNSEM } \boxed{1} \\ \text{NON-HEAD-DTRS} \left\langle \begin{bmatrix} \text{SYNSEM}|\text{LOC}|\text{CAT} \begin{bmatrix} \text{HEAD}|\text{MOD} & \boxed{1} \\ \text{SPR} & \langle\rangle \\ \text{COMP} & \langle\rangle \end{bmatrix} \end{bmatrix} \right\rangle \end{bmatrix}$$

This schema enforces the identity between the MOD value of the non-head daughter with the SYNSEM value of the head daughter. The adjunct (the non-head daughter) has to be completely saturated. Without such a requirement the theory would admit strings like (24), in which the preposition *in* entered a head-adjunct structure without being completely saturated.

(24) * *Mary worked in.*

Heads cannot take (further) arguments if their valence features have the empty list (⟨⟩) as their value. Similarly, it has to be ensured that certain words or phrases cannot be used as modifiers. For example, a pronoun like *he* or a complete NP like *the man* does not modify anything. In order to make sure that such lexical items and phrases do not enter head-adjunct structures, their MOD value or rather the MOD value of their head is specified as *none*. Since *none* is incompatible with any *synsem* object, words like *he* and phrases like *the man* are incompatible with the requirements for non-head daughters in head-adjunct phrases.

4.9 Extraction: Modeling nonlocal dependencies as sequences of local dependencies

The analysis of unbounded dependencies is inherited from GPSG and originally due to Gazdar (1981). We want to explain it with reference to the examples in (25):

(25) a. *This man, she met before.*
  b. *This man, I think that she met before.*

In what follows we assume an empty category-based analysis of nonlocal dependencies.[13] Figure 6 shows the analysis of (25a): the position of the object is taken by a trace. The trace is basically a joker that can fulfill whatever is required in a phrase. However, the trace has special properties. It passes up information about the missing element (NP in the example).[14] This is indicated by the slash ('/') in the figure. The information is passed on to higher nodes until it is finally bound off by some other element being compatible with the properties that are passed on. This element is called *filler*.

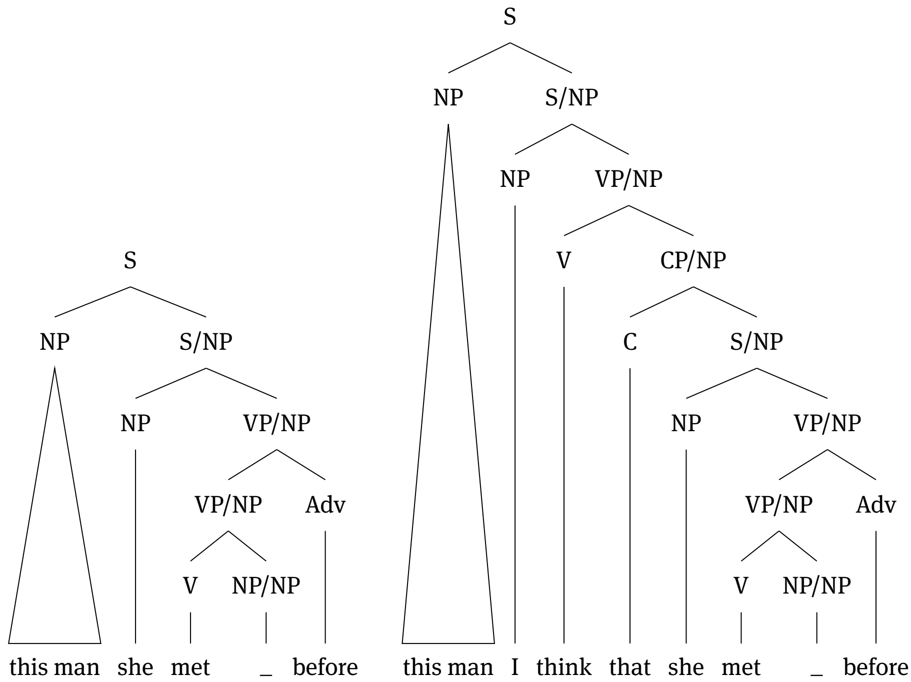


Figure 6: Analysis of nonlocal dependencies as local passing of information.

13 For traceless analyses of extraction see Bouma, Malouf and Sag (2001). A detailed discussion of this analysis can be found in Levine and Hukari (2006).
14 Again the analysis is explained bottom up for explanatory purposes only.

A complete sentence is a verbal projection that has fully saturated valency lists and no element in the SLASH list.

The lexical item for a trace is given in (26).[15] Further NONLOC features are introduced below. Their values are the empty list, therefore they are not provided here to enhance readability.

(26) Lexical item for a trace (adapted from Pollard and Sag, 1994, 164):

$$\begin{bmatrix} \text{PHON} & \langle\rangle \\ \text{SYNSEM} & \begin{bmatrix} \text{LOC} & \boxed{1} \\ \text{NONLOC} & \begin{bmatrix} \text{INHER}|\text{SLASH} & \langle \boxed{1} \rangle \\ \text{TO-BIND}|\text{SLASH} & \langle\rangle \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

This basically says: the word has no phonological material, i. e., nothing is pronounced. Whatever is locally required is compatible with the trace. The trace is a man without qualities, it does what it is told. But: whatever it is that is required by the syntactic context (NP, PP, an adjunct), the LOCAL value of the trace is identified with the element in the INHERITED|SLASH value ($\boxed{1}$). The SLASH value is passed up the tree by the Nonlocal Feature Principle, which states that the lists of nonlocal features at a mother node are the concatenation of the nonlocal features of the daughters minus those elements that have been bound off (those that are listed under TO-BIND in a daughter, usually in the head daughter). Figure 7 shows this in detail. The boxed

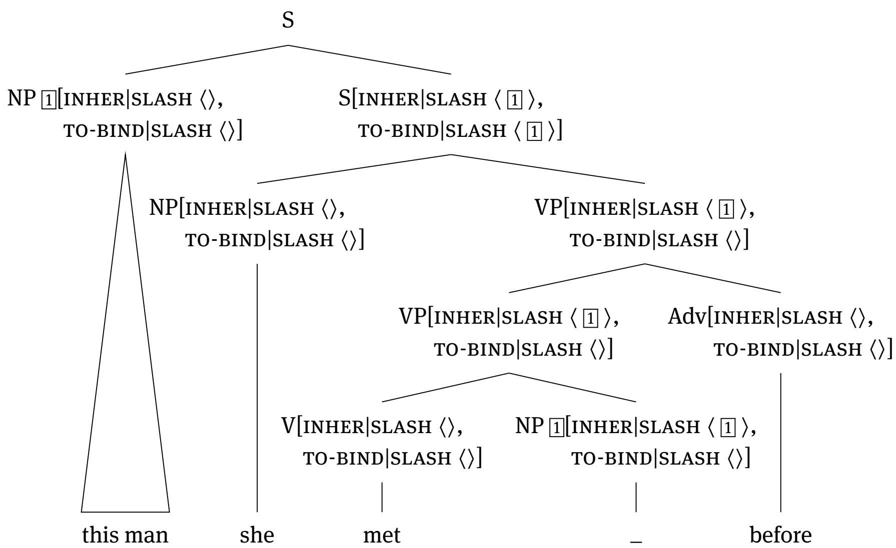


Figure 7: Analysis of nonlocal dependencies as local passing of information.

15 Pollard and Sag (1994) use sets as the value of nonlocal features. The mathematical formalization behind sets is very complicated (Pollard and Moshier, 1990). We use lists in what follows.

number after a category symbol (for example the $\boxed{1}$ following the NP) refers to the LOCAL value of the respective linguistic object while boxes in front of category symbols as in earlier figures refer to SYNSEM values. All normal words have empty SLASH values, but the trace contributes a SLASH element. The respective lists are concatenated and since no word or schema has a specified TO-BIND|SLASH value, nothing is bound off until *she met before* is combined with the filler *this man*. This special filler-head combination is licensed by the following schema:

Schema 5 (Head-Filler Schema)
head-filler-phrase ⇒

$$
\left[\begin{array}{ll}
\text{HEAD-DTR } \boxed{1} & \left[\text{SYNSEM} \left[\begin{array}{ll} \text{LOC|CAT} & \left[\begin{array}{l} \text{HEAD} \left[\begin{array}{l} \textit{verb} \\ \text{VFORM } \textit{fin} \end{array}\right] \\ \text{SPR } \langle\rangle \\ \text{COMPS } \langle\rangle \end{array}\right] \\ \text{NONLOC} & \left[\text{TO-BIND|SLASH } \langle \boxed{2} \rangle\right] \end{array}\right]\right] \\
\text{DTRS} & \left\langle \left[\text{SYNSEM} \left[\begin{array}{l} \text{LOC } \boxed{2} \\ \text{NONLOC|INHER|SLASH } \langle\rangle \end{array}\right]\right], \boxed{1} \right\rangle
\end{array}\right]
$$

The LOCAL value that was passed up through the tree is identified with the LOCAL value of the non-head daughter. The specification of the INHER|SLASH value of the non-head daughter makes sure that nothing is extracted out of the filler. The head daughter is specified to be a finite clause with all arguments saturated.

There is a Nonlocal Feature Principle that ensures that the NONLOC values of the mother are the concatenation of the NONLOC values of the daughters minus the elements in TO-BIND. Since the TO-BIND|SLASH value of the head-daughter is non-empty, there has to be an INHER|SLASH value at the head daughter since something has to be bound off. The respective element in INHER|SLASH has to be $\boxed{2}$. INHER|SLASH of the head daughter may contain further elements, but $\boxed{2}$ has to be in there and it will be bound off and not contained in the INHER|SLASH value of the mother node.

We used the empty element in (26) in the analysis of nonlocal dependencies and we want to close this subsection with a general comment on empty elements in HPSG. Empty elements are usually frowned upon within the HPSG community and some researchers do not use them at all or at least avoid them for certain kinds of phenomena (Sag and Fodor, 1994; Bouma, Malouf and Sag, 2001, Section 3.5), but there is no dogmatic ban on empty elements as for instance in Construction Grammar. For instance, suggestions to assume empty elements can be found in Bender (2000), Sag, Wasow and Bender (2003, 464), Borsley (1999, 2009) and Alqurashi and Borsley (2013). Personally, we agree with CxG views that there is a language acquisition problem with empty elements but we think this holds only for those empty elements that cannot be motivated by language-internal evidence. For instance, facts about object agreement in Basque are not accessible to learners of English and hence they could not learn an AgrO projection (as assumed for instance by Chomsky, 1995, 7, 59–60). But learners of English have evidence that they can leave out determiners in plural noun phrases. As it has been shown elsewhere, there are situations in which grammars using empty elements capture the generalizations regarding omissible elements more directly than grammars without empty elements (Müller, 2014a; 2016, Chapter 19).

4.10 Roots, words, lexical rules

We showed in Section 4.3 how valence information is represented and in Section 4.7 how this valence information determines which kinds of trees are licensed. One important tool of HPSG has not been mentioned yet: lexical rules. Lexical rules are used to relate lexical objects. For languages with inflection, roots are the minimal objects described by linguists. These roots are related to inflected forms by lexical rules. Lexical rules are also used for derivational morphology. For example, the German adjective *lesbare* 'readable' is derived from the root *les-* 'read' by appending the suffix *-bar* '-able' (Müller, 2003). The resulting adjectival stem *lesbar* 'readable' is inflected by adding the suffix *-e*.

Lexical rules are also used for modeling valence alternations. The following simplified lexical rule accounts for the passive:

(27) $$\begin{bmatrix} \textit{stem} & \\ \text{HEAD} & \textit{verb} \\ \text{ARG-ST} & \langle \text{NP} \rangle \oplus \boxed{1} \end{bmatrix} \mapsto \begin{bmatrix} \textit{word} & \\ \text{HEAD} & \begin{bmatrix} \textit{verb} & \\ \text{VFORM} & \textit{pas} \end{bmatrix} \\ \text{ARG-ST} & \boxed{1} \end{bmatrix}$$

The lexical rule in (27) basically maps a verb selecting for at least one NP to a participle passive that does not select for the subject but for all other arguments. The remaining arguments will be mapped to SPR and COMPS. English requires that $\boxed{1}$ starts with another NP or a sentential argument so that this NP (or sentential argument) can be put into the SPR list. Of course more has to be said about case assignment and so on but the basic explanation given above may be sufficient to understand the concept of a lexical rule. For details on passives in English see Pollard and Sag (1987) and on passives in Germanic languages in general see Müller and Ørsnes (2013a), Müller (2018b).

Lexical rules are doing some of the work that was done with transformations. Dowty (1978) called them lexically governed transformations. There is a crucial difference though: while transformations relate actual trees that have to be generated by the grammar before being able to function as inputs to transformations, lexical items license classes of trees. So lexical rules relate lexical items that license different classes of trees rather than relating trees directly. This difference is important when considering the psycholinguistic plausibility of theoretical models (Bresnan, 1978).

4.11 Grammatical functions

HPSG refers to NPs, PPs, and so on but does not refer to subjects, objects, and other grammatical functions as descriptive primitives of the theory. An exception is the SUBJ feature, which refers to subjects, but this is needed for separating subjects from other arguments in SVO languages[16] and for allowing access to information about non-expressed subjects of non-finite verbs in control and raising structures (see also footnote 7).

As has been often mentioned in the literature, it is not trivial to define grammatical functions in a crosslinguistically valid way. For instance, in German, subject (as far as NPs are concerned) can be defined as equivalent with non-predicative nominative (Reis, 1982). But in Icelandic, there are quirky case subjects and nominatives can be objects (Zaenen, Maling and Thráinsson, 1985). The verb agrees with the nominative element independent of its grammatical function. Hence, the term *subject-verb agreement* is inappropriate for Icelandic. Rather one should talk about *nominative-verb agreement*. As already mentioned above, subjects in SVO languages are serialized differently from objects and hence there has to be a syntactic differentiation between subjects and objects. A further difference is controlability. Both of these criteria have been used in the grammar of Icelandic to identify subjects. Using the SUBJ feature for modeling these properties is one way to deal with the data, but it does not presuppose a unified concept of subject that applies to all subjects cross-linguistically. Nevertheless, the individual criteria that have been suggested for subjecthood may apply within languages and of course they are modeled in HPSG (position before the verb in SVO languages, controlability, agreement, omitability in imperatives, etc.). However, these criteria may cluster differently from language to language. This is entirely unproblematic since it is not necessary in HPSG to assign grammatical functions to constituents.

4.12 Levels of representation

HPSG does not assume a Deep Structure from which a Surface Structure is derived by transformations as in Government & Binding (Chomsky, 1981), but there is something similar to Deep Structure: the ARGUMENT-STRUCTURE list (see Section 4.3). This list is a representation that contains all arguments of a head in a certain order.[17] This argument structure list can be used for linking valence requirements to semantic roles

16 We use the SPR feature for subjects in this paper, but see Ginzburg and Sag (2000) for the use of SUBJ as valence feature. Some versions of HPSG do not distinguish between subjects and complements in the valence list of heads (Sag, 2012; see Müller, 2018c, Section 10.6.2.3 for discussion). These versions use a feature XARG to make one argument (usually the subject) accessible for control and also question tag formation (Bender and Flickinger, 1999).

17 Butt (1995, 27) points out the LFG's f-structure corresponds to Deep Structure.

(Wechsler et al., 2019) as well as for Binding Theory (Pollard and Sag, 1992; Branco, 2019). So, things that are done in Deep Structure trees in GB, namely assigning theta roles to arguments at a certain position, are done within lexical items in HPSG.

4.13 Summary

This brief sketch of the formal foundation and basic tools mentioned many essential concepts that are used in HPSG. Of course a lot more could be and has been said about the properties of the formalism, but this introductory article is not the place to discuss them in detail. However, it cannot be emphasized enough that it is important that the formal details are worked out. The interested reader is referred to the work of Shieber (1986), Pollard and Sag (1987, Chapter 2), Johnson (1988), Carpenter (1992), King (1994, 1999), Pollard (1999) and Richter (2004, 2007). The work of King, Pollard, and Richter reflects current assumptions, that is, the model theoretic view on grammar that is assumed nowadays.

The following section deals with criteria for viewing an HPSG analysis as successful and after this section we fill in the missing pieces required for the analysis of the given example sentence.

5 Evaluation

Since HPSG is well-formalized it is always clear what a certain analysis stands for and what predictions it makes. It is therefore possible to test the analysis against collected data. This can be done in two ways: either by thinking about the consequences of an analysis or – more systematic and more reliable – by computer implementations that can be run against test suites (Oepen and Flickinger, 1998). Test suites are data collections of either hand-made or otherwise available sentences or phrases. Hand-made test suites can contain ungrammatical strings, which are especially valuable since they can be used for testing for overgeneration of linguistic theories. Systematic testing is important since it is often the case that one believes to have found a real simplification of one's theory. Running the test suite can show us the one example out of several thousands not covered by the simpler grammar, or the additional examples getting unwanted structures/readings. In pencil and paper work, these examples could easily escape our attention. Another way to test implemented grammars is to let them generate strings for a given meaning. The results are often surprising. They reveal aspects of the grammar nobody ever thought about before.

During a workshop on *Progress in linguistics* at the Freie Universität Berlin in 2013 the first author suggested that proponents of linguistic theories should work out grammar fragments of reasonable size and provide lists of covered and rejected example sentences (see Müller and Ørsnes (2015) and Müller (2017) for examples). When theories are developed further it can be checked whether the amount of covered data stays constant or is getting bigger. This is a possible way to evaluate the success of theoretical work. While some of the frameworks currently available changed their fundamental assumptions frequently, this is not the case for HPSG. There were no radical breaks during the past 33 years. This ensures that the amount of data that is covered by HPSG theories steadily increases.

There are always many ways to write a grammar for a given data set. In order to decide which grammar is the best one, one can compare these grammars to grammars of other languages. Let's say there are two ways of describing phenomenon X. If we have one grammar that can account for the data and is compatible with what we know about another language or other languages then we choose this grammar over the other (Müller, 2015b). Similarly, a simplicity metric can also be applied language internally: we chose the grammar that has to postulate fewer theoretical entities: fewer features, fewer empty elements. This is nothing special though. It is common scientific practice, also known as Occam's Razor. It should be noted though that the most compact description of linguistic knowledge is not necessarily the best one. There is ample evidence that a lot of linguistic knowledge is just stored as chunks in the human brain. So even though we could write a more compact grammar that derives chunks from their components rather than storing them, we do not necessarily do this since the empirical domain is not just a set of generated sentences. It is important how linguistic knowledge is represented in the brain and how it is used by speakers of the language.

6 Sample analysis

As was mentioned in Section 4.3, the shared task for authors of this volume is to analyze (12) – repeated here as (28) for convenience:

(28) *After Mary introduced herself to the audience, she turned to a man that she had met before.*

Some details concerning the analysis of (28) were already discussed in Section 4 by explaining the tools used in HPSG. But since some of the phenomena manifested in (28) are less foundational, we decided to put their analysis in a separate section.

In order to explain the HPSG analysis of the sentence, one has to explain how valence information is encoded and how it is linked to semantic representations (Section 4.3), and how the internal structure of basic sentences is licensed (Section 4.7). In order to account for the attachment of the adjunct *after Mary introduced herself to the audience* to the main clause and to explain how the relative clause *that she had met before* attaches to the noun it modifies, we have to explain how adjunction works (Section 4.8).

The analysis of relative clauses like *that she had met before* in the above example involves nonlocal dependencies. The respective schemata and mechanisms for establishing nonlocal dependencies are explained in Section 4.9. In Section 6.1, we provide a special schema for relative clauses, which is needed in addition.

Analyses of complex tenses are required to treat periphrastic forms like *had met*. Auxiliaries are raising verbs and hence we introduce the analysis of raising in Section 6.2. Case assignment and agreement play a role even in simple sentences. These phenomena are dealt with in Sections 6.3 and 6.4, respectively.

In order to account for the binding of the pronoun *herself* to *Mary*, something would have to be said about Binding Theory (Pollard and Sag, 1992). Due to space limitations Binding Theory will not be covered in this overview.[18] Section 6.5 sketches the semantics that is used.

The analysis of (28) is implemented in the TRALE system (Meurers, Penn and Richter, 2002; Penn, 2004) as part of the BEngl grammar (Müller, 2009, 2012, 2018b,a). This grammar is developed as part of the CoreGram project (Müller, 2015b). The grammars that are developed in this project share a common core of grammatical constraints. There are grammars for German, Danish, Persian, Maltese, Mandarin Chinese, English, French, and some toy fragments of other languages. The detailed analysis of (28) including semantics is available at https://hpsg.hu-berlin.de/~stefan/Pub/current-approaches-hpsg.html. Due to space limitations we can not explain the analysis in full detail here, especially with respect to semantics. The reader is referred to other sources in what follows. Some of the analyses deviate from theory variants that may be more common in the literature. When this is the case, reasons for deviating will be mentioned or discussed in footnotes.

6.1 Relative clauses

The analysis of relative clauses builds on the analysis of nonlocal dependencies already introduced in Section 4.9. The type of relative clause that is relevant in the analysis of (12) consists of an extracted phrase containing a relative pronoun and a clause from which it is extracted. As the bracketing in (29) shows, the relative pronoun *who* can be realized far away from where it would be usually realized in a declarative sentence.

(29) *the man who*$_i$ [*I* [*believe* [*that* [*Mary* [*was* [*introduced* [*to* $\_{}_i$]]]]]]]

Relative clauses are special in comparison to the nonlocal dependencies explained in Section 4.9 in that there are further conditions imposed on the fronted phrase: the

18 Binding Theory is assumed to operate on the list of all arguments of a head, the so-called ARG-ST list (see Section 4.3). An anaphor has to be coindexed with a less oblique element on the ARG-ST list if there is any. For the reflexive pronoun *herself* this means that it has to be coreferential with the subject of *introduce*. For further details see Pollard and Sag (1992) and Branco (2019).

fronted element has to contain a relative pronoun. The fronted element can be a relative pronoun as in our example and in (30a) the relative pronoun may be embedded as in (30b-c):

(30) a. *who* (*she met before*)
b. *whose sister* (*she met before*)
c. *a friend of whose sister* (*she met before*)

This requirement for a relative pronoun to be present can be analyzed as another nonlocal dependency and we can use similar techniques to establish the dependency. The referential index of the relative pronoun is shared with an element in the INHER|REL list of the relative pronoun:

(31) SYNSEM value of the relative pronoun *that*:

$$\left[\begin{array}{ll} \text{LOC} & \left[\begin{array}{ll} \text{CAT} & \left[\begin{array}{ll}\text{HEAD} & \textit{noun}\\ \text{SPR} & \langle\rangle \\ \text{COMPS} & \langle\rangle\end{array}\right] \\ \text{CONT} & \left[\text{IND}\ \boxed{1}\right]\end{array}\right] \\ \text{NONLOC} & \left[\text{INHER}\ \left[\text{REL}\ \langle \boxed{1} \rangle\right]\right]\end{array}\right]$$

In order to enhance readability, the INHER|SLASH and the TO-BIND values are not given, since their value is the empty list.

The lexical entry of the relative pronoun *whose* is rather similar, the only difference being the part of speech for *whose*, which is *det* rather than *noun*. Figure 8 shows the analysis of the noun phrase *a friend of whose sister*.

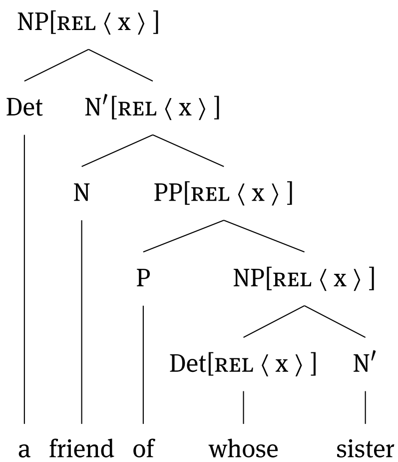


Figure 8: Representation of the referential index of a relative pronoun in the NONLOC|INHER|REL list.

A relative clause differs from an ordinary clause in that it can modify a noun. As the following examples show, ordinary clauses cannot do this:

(32) a. * *The man* [*she had met that before*] *laughs.*
b. * *The man* [*she had met him before*] *laughs.*

This means that the MOD value of verbs – which are the heads of clauses – has to be *none*. But for relative clauses we must have a representation where the relative clause has an $\overline{\mathrm{N}}$ as the MOD value. There are at least three ways to achieve this: one can assume that the MOD value of verbs is a default that can be overridden in the case of relative clauses (Sag, 1997), one can assume that relative clauses are the projection of an empty relativizer (Pollard and Sag, 1994, 213–217), or one can assume that relative clauses are unheaded (Müller, 1999b,a). In the latter proposal a relative phrase is combined with a clause from which it is missing and the result is a relative clause with an appropriately specified MOD value. The latter proposal has the advantage that the special headless construction can also contribute the appropriate semantics. The relative clause *that she had met before* then behaves like other postnominal modifiers and can be adjoined to an $\overline{\mathrm{N}}$. As Sag (1997) points out, an analysis with the verb as head would require different meanings for the verb when used in ordinary clauses in comparison to verbs that are used in relative clauses. Rather than assuming verbs with nominal meanings for modification of $\overline{\mathrm{N}}$s, Sag assumes that relative clauses have verbal meaning and that there is a special schema that licenses the combination of $\overline{\mathrm{N}}$s and relative clauses. The approach suggested here does not need such an additional schema.

The Relative Clause Schema is given as Schema 6:

Schema 6 (Relative Clause Schema)
relative-clause ⇒

$$\begin{bmatrix} \text{SYNSEM} & \begin{bmatrix} \text{LOC} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \textit{rc} \\ \text{MOD} \;\; \overline{\mathrm{N}}_{\boxed{1}} \end{bmatrix} \\ \text{SPR} & \langle\rangle \\ \text{COMPS} & \langle\rangle \end{bmatrix} \end{bmatrix} \end{bmatrix} \\ \text{DTRS} & \left\langle \begin{bmatrix} \text{SS} & \begin{bmatrix} \text{LOC}\ \boxed{2} \\ \text{NLOC}|\text{INH} \begin{bmatrix} \text{REL} & \langle \boxed{1} \rangle \\ \text{SLASH} & \langle\rangle \end{bmatrix} \end{bmatrix} \end{bmatrix}, \begin{bmatrix} \text{SS} & \begin{bmatrix} \text{LOC}|\text{CAT} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \textit{verb} \\ \text{VFORM} \;\; \textit{fin} \end{bmatrix} \\ \text{SPR} & \langle\rangle \\ \text{COMPS} & \langle\rangle \end{bmatrix} \\ \text{NLOC}|\text{TO-BIND} & \begin{bmatrix} \text{REL} & \langle \boxed{1} \rangle \\ \text{SLASH} & \langle \boxed{2} \rangle \end{bmatrix} \end{bmatrix} \end{bmatrix} \right\rangle \end{bmatrix}$$

The schema combines a finite clause containing a gap ($\boxed{2}$) with an appropriate filler, that is, with a filler that has the same LOCAL value as the gap, namely $\boxed{2}$. The filler has to contain a relative pronoun, the referential index of which ($\boxed{1}$) is identified with the referential index of the $\overline{\mathrm{N}}$ that is modified by the relative clause. This schema is really similar to the Filler-Head Schema: the first daughter is the filler that binds off an unbounded dependency. One difference is that the Filler-Head Schema has one of the daughters designated as head daughter while the Relative Clause Schema is unheaded. Both schemata inherit from the same supertype, which states the constraints regarding a filler and the phrase from which it is extracted.

This schema does not include information about semantics but a treatment of semantics can be found in Müller (2013a, Section 11.2.2) and Müller (1999a, Section 2.7). For further details on semantics see also Section 6.5.

6.2 Raising

Our example sentence involves the auxiliary verb *have*. Auxiliaries are raising verbs (Pollard and Sag, 1994, 143; Sag, Wasow and Bender, 2003, 396; Sag et al., 2020): they do not care for the type of the subject of the embedded verb since they do not assign semantic roles to it. In languages that allow for subjectless constructions (as for instance German) auxiliaries may embed subjectless verbs (Kiss, 1995, 87). The trick to describe raising predicates is that arguments may belong to several heads at the same time. So, a raising predicate takes as its subject whatever its complement requires as a subject. This is done with structure sharing (the tool that does most of the work in HPSG, see Section 4.1). (33) shows the lexical item for the finite form *had* of the auxiliary verb *have*:

(33) CAT value for *had*:

$$\left[\begin{array}{ll} \text{HEAD} & \left[\begin{array}{l} \textit{verb} \\ \text{VFORM} \;\; \textit{fin} \end{array}\right] \\ \text{SPR} & \langle \boxed{1} \rangle \\ \text{COMPS} & \left\langle \left[\text{LOC|CAT} \left[\begin{array}{ll} \text{HEAD} & \left[\begin{array}{l} \textit{verb} \\ \text{VFORM} \;\; \textit{perf} \end{array}\right] \\ \text{SPR} & \langle \boxed{1} \rangle \\ \text{COMPS} & \langle\rangle \end{array}\right] \right] \right\rangle \end{array}\right]$$

The auxiliary selects for a VP (something with an empty COMPS list and one element in the SPR list). The element in the specifier list is identified with the element in the SPR list of the auxiliary.[19] Figure 9 shows the analysis of (34):

(34) *She had met him.*

19 In reality this is more indirect: the element in the SPR list of the embedded verb is identified with the first element of the ARG-ST list of the auxiliary. The second element on ARG-ST is the selected VP. The elements on ARG-ST are mapped to SPR and COMPS, resulting in the values provided in (33). We did not include this in (33) for reasons of readability. For further details see Müller and Ørsnes (2013a). This paper also contains a generalized description of auxiliaries that works for both English and German, which allows for frontings of partial verbal projections.

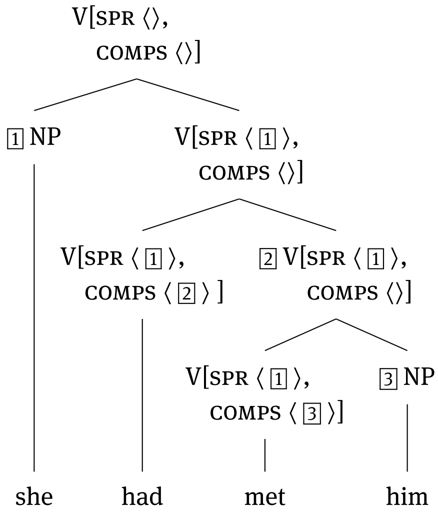


Figure 9: Analysis of sentence with auxiliary and subject raising.

The perfect participle *met* selects a subject via SPR ($\boxed{1}$) and an object via COMPS ($\boxed{3}$). It is combined with *him*, resulting in a VP. The auxiliary selects for a VP ($\boxed{2}$) and identifies its own element in SPR with the element in SPR of the VP ($\boxed{1}$). The result of the combination of auxiliary and VP is a VP that is still lacking a specifier. Since the specifier of the auxiliary was identified with the specifier of the VP headed by the perfect participle, the specifier of the complete VP is identical to the one of the VP headed by the perfect participle ($\boxed{1}$). This way *she* is both the subject of *had* and the subject of *met* although only the latter assigns a semantic role to it.

This kind of raising analysis also works for verbs like *seem* and verbs like *see* and *let* (also know as Exceptional Case Marking verbs). The same technique can be used for the analysis of complex predicates. For instance, for German and Persian it has been suggested that a predicate complex is formed. This can be modeled by not just raising the subject of the embedded predicate but by raising all arguments (Hinrichs and Nakazawa, 1994; Müller, 2010).

6.3 Case assignment

English has a relatively simple case system and until now we have ignored case in the example analysis. Pronouns do differ according to case and in order to rule out sequences like (35), one has to say something about case.

(35) * *Him likes he.*

Like other theories HPSG distinguishes between structural and lexical cases (Yip, Maling and Jackendoff, 1987, 222). Structural cases are those that change according to syntactic environments, lexical cases stay constant. We assume that the nominal arguments of one and two place verbs in English have structural case. The first two arguments of three-place verbs with nominal arguments have structural case and the third one has lexical case since it cannot be promoted to subject and get nominative in passive sentences. The Case Principle (Meurers, 1999b; Przepiórkowski, 1999) says for verbal environments that the first element with structural case in the ARG-ST list gets nominative and all other elements with structural case get accusative. This works well for our lexical items repeated here as (36) with the case information specified:

(36) lexical items (active):

	ARG-ST
a. meet	⟨ NP[*str*], NP[*str*] ⟩
b. turn	⟨ NP[*str*], PP[*to*] ⟩
c. introduce	⟨ NP[*str*], NP[*str*], PP[*to*] ⟩

The first element in the list is the subject. It has structural case and gets nominative. In cases in which the second element has structural case (see (36a) and (36c) but not (36b)), it gets accusative.

While case assignment by means of the Case Principle sounds straight forward, there are AcI verbs like *see* as in (37) that present a challenge for this version of the principle.

(37) *He saw her laugh.*

laugh is an intransitive verb with a subject argument. This argument is raised to the object of *see*. So it is simultaneously the subject of *laugh* and the object of *see*, which would result in a conflict if nothing special is said about such situations. Obviously, the higher verb should be given priority, so in order to avoid the conflict, the Case Principle has to be reformulated as follows: In verbal domains the first element of the ARG-ST list with structural case gets nominative provided it is not raised. In verbal domains, all other elements of the ARG-ST list with structural case get accusative provided they are not raised.[20] In order to be able to distinguish raised from non-raised elements a Boolean feature[21] RAISED is used. The elements of the ARG-ST list are not *synsem* objects but include *synsem* objects and have more internal structure:

(38) $\begin{bmatrix} \textit{arg} & \\ \text{ARG} & \textit{synsem} \\ \text{RAISED} & \textit{bool} \\ \text{REALIZED} & \textit{bool} \end{bmatrix}$

20 This is basically a reformulation of Yip, Maling & Jackendoff's Case Principle (1987) without overriding case values. They applied their case theory to Icelandic and it comes as no surprise that the monotonic reformulation works for Icelandic as well.

21 The possible (maximal specific) values of Boolean valued features are '+' and '–'.

When an argument is raised from a predicate, the value of RAISED is instantiated as '+'. All those elements that are not raised are marked as RAISED–. Case is assigned to those elements that are RAISED–.

As the following example from Webelhuth (1985, 210) shows, German allows for the fronting of non-finite verb phrases that contain a nominative.

(39) [*Zwei Männer erschossen*] *wurden während des Wochenendes.*
two men.NOM shot were.PL during the weekend
'Two men were shot during the weekend.'

The standard analysis of auxiliary-verb combinations in German assumes that auxiliaries form a verbal complex with the embedded verb (Hinrichs and Nakazawa, 1994). Auxiliaries attract all arguments of their embedded verbal complements and hence they can assign them case. The problem with examples like (39) is that *zwei Männer erschossen* 'two men shot' forms a phrase and the argument *zwei Männer* is not represented at the mother node and hence, when *wurden* 'were' is combined with *zwei Männer erschossen* 'two men shot', *wurden* 'were' cannot attract the argument of *erschossen* 'shot'. This problem was solved by assuming that elements that get saturated stay in the valence list but are marked as realized (i. e. they have a REALIZED value '+' rather than '–'). The realized elements are still around on the valence lists, which is the reason why they are called *spirits* (Meurers, 1999b). Figure 10 shows the analysis of (34) with spirits (marked with checked off boxes). With this slight change in representation the argument of *erschossen* 'shot' is still present at *zwei Männer erschossen* 'two

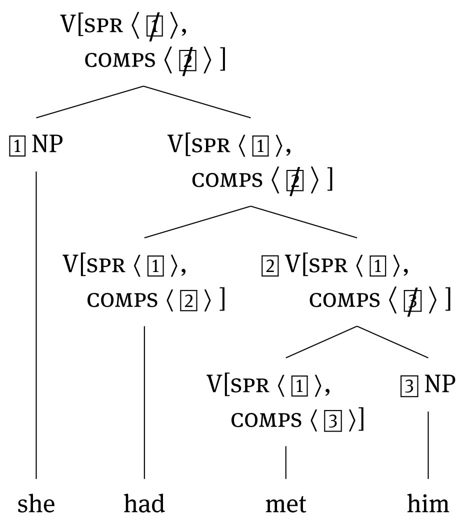


Figure 10: Analysis of sentence with auxiliary and subject raising.

men shot' and hence can be attracted by *wurden* 'were'. Since it is the first argument of *wurden* with structural case it is assigned nominative.

6.4 Agreement

A lot can be said about the many faces of agreement. A short overview article is not the right place to do this but the interested reader is referred to a great book by Wechsler and Zlatić (2003). Agreement in English is covered by Pollard and Sag (1994, Chapter 2). The verbs in our example sentence all are in the past tense so that no agreement is visible. But English has subject verb agreement. The subject is the first element of the argument structure list. A more general description of agreement is that the finite verb agrees with a nominative element. Since nominative is assigned to the first element of the ARG-ST list that has structural case (see Section 6.3), this element is the one that agrees with the finite verb. This characterization of verb-nominative agreement works for many languages including Icelandic, where we have quirky case subjects and objects in the nominative. These objects agree with the finite verb as is shown by examples like (40), which is taken from Zaenen, Maling and Thráinsson (1985, 451):

(40) *Hefur henni alltaf þótt Ólafur leibinlegur?*
has she.DAT always thought Olaf.NOM boring.NOM
'Has she always considered Olaf boring?'

Zaenen et al. (1985) show with several tests that *henni* 'she' should be regarded as the subject and *Ólafur* as the object. But agreement is with the nominative element and hence in sentences like (40) with the object. Here, it is assumed that quirky subjects have lexical case and hence an analysis assuming that the first NP with structural case (if there is any) agrees with the finite verb gets the facts right (Müller, 2018b).

Again, like with case assignment agreement relations can exist between finite verbs and nominatives that are embedded in a separate phrase (Höhle, 1997, 114):

(41) [*Die Hände gezittert*] *haben / * hat ihm diesmal nicht.*
the hands trembled have has him this.time not
'His hands did not tremble this time.'

Since the argument of *gezittert* 'trembled' is raised to the auxiliary *haben* 'have' as a spirit, it is accessible to *haben* and the agreement relation can be established. The same explanation works for the passive example in (39): since the nominative is an element of the ARG-ST of the auxiliary, the agreement relation can be established.

6.5 Semantics

Very little has been said about semantics so far since this book is about syntactic approaches. However, HPSG takes the integration of constraints on all linguistic levels seriously and therefore most HPSG analyses cover both syntactic and semantic aspects of the phenomena under consideration (Koenig and Richter, 2019).

HPSG started out with Situation Semantics (Barwise and Perry, 1983). Later, Minimal Recursion Semantics (Copestake, Flickinger, Pollard and Sag, 2005) was developed, which is assumed by many researchers nowadays. We cannot explain the inner workings of MRS but we can point out some of its merits: MRS is a so-called underspecified semantics framework. Scope relations are represented in an underspecified way. So for many situations one gets one semantic representation which stands for several readings.[22] So far the linking between syntax and semantics in lexical items has been explained, see for instance example (14). The semantic representation is contained under CONT and it consists of an index (basically an event variable or a variable for an individual) and a list of relations. The index usually plays a role in these relations. The semantic contribution of a phrase is the concatenation of the RELS lists of its daughters plus a special RELS list that is specified in phrasal schemata. For vanilla schemata like the Head-Complement Schema, the RELS list that is specified by the schema is the empty list. So the Head-Complement Schema is fully compositional in the sense that no extra information is added. However, phrasal schemata can contribute additional relations by this extra RELS list. It is a claim found in much of the Construction Grammar literature that certain phrasal configurations contribute their own meaning. This can be handled easily with HPSG's Semantics Principle, which states that the relations contributed by the mother node are the concatenation of the relations of the daughters plus the relations contributed by the phrasal schema (see for instance Copestake et al., 2005, Section 6.6).

The implemented grammar contains semantic constraints. So, the interested reader may inspect the analyzed example sentence[23] to see the semantic contributions of words and phrases. Click the top-most node in the tree and have a look at CONT, RELS, and HCONS. Click on boxes to inspect their content.

7 Conclusion

This paper introduced the framework of Head-Driven Phrase Structure Grammar (HPSG). We showed how HPSG captures morphology, constituent structure and relations between syntax and semantics by using feature value pairs, identity of values (structure sharing) and relational constraints. Issues of modularity and interfaces and psycholinguistic plausibility have been touched. Since the specification of features and values is purely declarative and no processing regime is associated with the linguistic knowledge it can be used in any direction, that is, it can be used for

22 See for instance Egg (1999) for the three readings of *Max opened all windows again.*

23 https://hpsg.hu-berlin.de/~stefan/Pub/current-approaches-hpsg.html parsing and generation. This kind of linguistic knowledge is compatible with incremental processing that processes information from all descriptive levels including world knowledge.

The paper analyzes one example sentence in detail. The phenomena discussed involved basic constituent structure, valence, linking, case assignment, agreement, raising, extraction (nonlocal dependencies) and relative clause formation. It has been shown that all relations can be modeled as local relations. In some cases in which transformations/movement are/is suggested in other frameworks, identifying information via structure sharing is used. For instance in raising, arguments can be the arguments of several heads. Nonlocal dependencies are modeled by passing information up until it is bound off by a filler.

References

Ajdukiewicz, Kazimierz. 1935. Die syntaktische Konnexität. *Studia Philosophica* 1. 1–27.

Alexopoulou, Theodora & Dimitra Kolliakou. 2002. On linkhood, topicalization and clitic left dislocation. *J. of Linguistics* 38(2). 193–245.

Alqurashi, Abdulrahman & Robert D. Borsley. 2013. Arabic relative clauses in HPSG. In Stefan Müller (ed.), *Proceedings of the 20th International Conference on Head-Driven Phrase Structure Grammar, Freie Universität Berlin*, 25–43. Stanford: CSLI Publications.

Bargmann, Sascha. 2015. Syntactically flexible VP-Idioms and the N-after-N Construction. *Poster presentation at the 5th General Meeting of PARSEME*, Iasi, 23-–24 September 2015.

Barwise, Jon & John Perry. 1983. *Situations and Attitudes*. Cambridge, MA/London, England: MIT Press.

Bender, Emily M. 2000. Syntactic variation and linguistic competence: The Case of AAVE copula absence. Stanford University, dissertation.

Bender, Emily M. & Dan Flickinger. 1999. Peripheral constructions and core phenomena: Agreement in tag questions. In Gert Webelhuth, Jean-Pierre Koenig & Andreas Kathol (eds.), *Lexical and Constructional Aspects of Linguistic Explanation* (Studies in Constraint-Based Lexicalism 1), 199–214. Stanford: CSLI Publications.

Bildhauer, Felix. 2008. Representing Information Structure in an HPSG Grammar of Spanish. Universität Bremen, dissertation.

Bildhauer, Felix. 2011. Mehrfache Vorfeldbesetzung und Informationsstruktur: Eine Bestandsaufnahme. *Deutsche Sprache* 39(4). 362–379.

Bildhauer, Felix & Philippa Cook. 2010. German multiple fronting and expected topic-hood. In Stefan Müller (ed.), *Proceedings of the 17th International Conference on Head-Driven Phrase Structure Grammar, Université Paris Diderot*, 68–79. Stanford: CSLI Publications.

Bird, Steven. 1995. *Computational Phonology: A Constraint-Based Approach* (Studies in Natural Language Processing). Cambridge, UK: Cambridge University Press.

Bird, Steven & Ewan Klein. 1994. Phonological analysis in typed feature systems. *Computational Linguistics* 20(3). 455–491.

Borsley, Robert D. 1987. Subjects and Complements in HPSG. Report No. CSLI-87-107. Stanford: Center for the Study of Language Information.

Borsley, Robert D. 1989. An HPSG approach to Welsh. *J. of Linguistics* 25. 333–354.

Borsley, Robert D. 1999. Mutation and constituent structure in Welsh. *Lingua* 109(4). 267–300.

Borsley, Robert D. 2009. On the superficiality of Welsh agreement. *Natural Language and Linguistic Theory* 27(2). 225–265.

Bouma, Gosse, Robert Malouf & Ivan A. Sag. 2001. Satisfying constraints on extraction and adjunction. *Natural Language and Linguistic Theory* 19(1). 1–65.

Branco, António. 2019. Binding theory. In Anne Abeillé, Robert D. Borsley, Jean-Pierre Koenig & Stefan Müller (eds.), *Head-Driven Phrase Structure Grammar: The Handbook* (Empirically Oriented Theoretical Morphology and Syntax), to appear. Berlin: Language Science Press.

Brants, Sabine, Stefanie Dipper, Peter Eisenberg, Silvia Hansen-Schirra, Esther König, Wolfgang Lezius, Christian Rohrer, George Smith & Hans Uszkoreit. 2004. TIGER: Linguistic interpretation of a German corpus. *Research on Language and Computation* 2(4). 597–620.

Bresnan, Joan. 1978. A realistic transformational grammar. In Morris Halle, Joan Bresnan & George A. Miller (eds.), *Linguistic Theory and Psychological Reality*, 1–59. Cambridge, MA: MIT Press.

Bresnan, Joan & Ronald M. Kaplan. 1982. Introduction: Grammars as mental representations of language. In Joan Bresnan (ed.), *The Mental Representation of Grammatical Relations* (MIT Press Series on Cognitive Theory and Mental Representation), xvii–lii. Cambridge, MA/London: MIT Press.

Butt, Miriam. 1995. *The Structure of Complex Predicates in Urdu* (Dissertations in Linguistics). Stanford: CSLI Publications.

Carpenter, Bob. 1992. *The Logic of Typed Feature Structures* (Tracts in Theoretical Computer Science). Cambridge: Cambridge University Press.

Chomsky, Noam. 1957. *Syntactic Structures* (Janua Linguarum / Series Minor 4). The Hague/Paris: Mouton.

Chomsky, Noam. 1981. *Lectures on Government and Binding*. Dordrecht: Foris.

Chomsky, Noam. 1995. *The Minimalist Program* (Current Studies in Linguistics 28). Cambridge, MA/London, England: MIT Press.

Chomsky, Noam & George A. Miller. 1963. Introduction to the formal analysis of natural languages. In R. Duncan Luce, Robert R. Bush & Eugene Galanter (eds.), *Handbook of Mathematical Psychology, vol. 2*, 269–321. New York: Wiley.

Copestake, Ann, Dan Flickinger, Carl Pollard & Ivan A. Sag. 2005. Minimal recursion semantics: An introduction. *Research on Language and Computation* 4(3). 281–332.

Croft, William. 2001. *Radical Construction Grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press.

Crysmann, Berthold & Olivier Bonami. 2016. Variable morphotactics in information-based morphology. *Journal of Linguistics* 52(2). 311–374.

De Kuthy, Kordula. 2002. *Discontinuous NPs in German* (Studies in Constraint-Based Lexicalism 14). Stanford: CSLI Publications.

Dowty, David R. 1978. Governed transformations as lexical rules in a Montague Grammar. *Linguistic Inquiry* 9(3). 393–426.

Dowty, David R. 1991. Thematic proto-roles and argument selection. *Language* 67(3). 547–619.

Egg, Markus. 1999. Derivation and resolution of ambiguities in *wieder*-sentences. In Paul J. E. Dekker (ed.), *Proceedings of the 12th Amsterdam Colloquium*, 109–114.

Engdahl, Elisabet & Enric Vallduví. 1996. Information packaging in HPSG. In Claire Grover & Enric Vallduví (eds.), *Edinburgh Working Papers in Cognitive Science* (Studies in HPSG 12), chap. 1, 1–32. Edinburgh: Centre for Cognitive Science, University of Edinburgh.

Flickinger, Dan, Stephan Oepen & Gisle Ytrestøl. 2010. Wikiwoods: syntacto-semantic annotation for english wikipedia. In Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, Mike Rosner & Daniel Tapias (eds.), *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, 1665–1671. Valletta, Malta: European Language Resources Association (ELRA).

Flickinger, Dan, Ann Copestake & Ivan A. Sag. 2000. HPSG analysis of English. In Wolfgang Wahlster (ed.), *Verbmobil: Foundations of Speech-to-Speech Translation* (Artificial Intelligence), 254–263. Berlin: Springer Verlag.

Flickinger, Dan, Carl Pollard & Thomas Wasow. 1985. Structure-sharing in lexical representation. In William C. Mann (ed.), *Proceedings of the Twenty-Third Annual Meeting of the Association for Computational Linguistics*, 262–267. Chicago, IL.

Fodor, Jerry A., Thomas G. Bever & Merrill F. Garrett. 1974. *The Psychology of Language: An Introduction to Psycholinguistics and Generative Grammar*. New York: McGraw-Hill Book Co.

Gazdar, Gerald. 1981. Unbounded dependencies and coordinate structure. *Linguistic Inquiry* 12. 155–184.

Gazdar, Gerald, Ewan Klein, Geoffrey K. Pullum & Ivan A. Sag. 1985. *Generalized Phrase Structure Grammar*. Cambridge, MA: Harvard University Press.

Ginzburg, Jonathan & Robin Cooper. 2004. Clarification, ellipsis, and the nature of contextual updates in dialogue. *Linguistics and philosophy* 27(3). 297–365.

Ginzburg, Jonathan & Ivan A. Sag. 2000. *Interrogative Investigations: The Form, Meaning, and Use of English Interrogatives* (CSLI Lecture Notes 123). Stanford: CSLI Publications.

Goldberg, Adele E. 1995. *Constructions: A Construction Grammar approach to argument structure* (Cognitive Theory of Language and Culture). Chicago/London: The University of Chicago Press.

Goldberg, Adele E. 2003. Words by default: The Persian Complex Predicate Construction. In Elaine J. Francis & Laura A. Michaelis (eds.), *Mismatch: Form-function Incongruity and the Architecture of Grammar* (CSLI Lecture Notes 163), 117–146. Stanford: CSLI Publications.

Halliday, Michael A. K. 1970. *A Course in Spoken English: Intonation*. Oxford University Press.

Hinrichs, Erhard W. & Tsuneko Nakazawa. 1994. Linearizing AUXs in German verbal complexes. In John Nerbonne, Klaus Netter & Carl Pollard (eds.), *German in Head-Driven Phrase Structure Grammar* (CSLI Lecture Notes 46), 11–38. Stanford: CSLI Publications.

Höhle, Tilman N. 1997. Vorangestellte Verben und Komplementierer sind eine natürliche Klasse. In Christa Dürscheid, Karl Heinz Ramers & Monika Schwarz (eds.), *Sprache im Fokus: Festschrift für Heinz Vater zum 65. Geburtstag*, 107–120. Republished as Höhle (2018b). Tübingen: Niemeyer.

Höhle, Tilman N. 1999. An architecture for phonology. In Robert D. Borsley & Adam Przepiórkowski (eds.), *Slavic in Head-Driven Phrase Structure Grammar*, 61–90. Republished as Höhle (2018a). Stanford: CSLI Publications.

Höhle, Tilman N. 2018a. An architecture for phonology. In Stefan Müller, Marga Reis & Frank Richter (eds.), *Beiträge zur Grammatik des Deutschen: Gesammelte Schriften von Tilman N. Höhle* (Classics in Linguistics 5). Originally published as Höhle (1999). Berlin: Language Science Press.

Höhle, Tilman N. 2018b. Vorangestellte Verben und Komplementierer sind eine natürliche Klasse. In Stefan Müller, Marga Reis & Frank Richter (eds.), *Beiträge zur Grammatik des Deutschen: Gesammelte Schriften von Tilman N. Höhle* (Classics in Linguistics 5). First published as Höhle (1997). Berlin: Language Science Press.

Jackendoff, Ray. 2008. Construction after Construction and its theoretical challenges. *Language* 84(1). 8–28.

Jacobson, Pauline. 1987. Review of Generalized Phrase Structure Grammar. *Linguistics and Philosophy* 10(3). 389–426.

Johnson, Mark. 1988. *Attribute-Value Logic and the Theory of Grammar* (CSLI Lecture Notes 14). Stanford: CSLI Publications.

Kathol, Andreas. 2001. Positional effects in a monostratal grammar of German. *J. of Linguistics* 37(1). 35–66.

Kay, Paul, Ivan A. Sag & Dan Flickinger. 2015. A Lexical Theory of Phrasal Idioms. Ms. CSLI Stanford.

Keenan, Edward L. & Bernard Comrie. 1977. Noun phrase accessibility and Universal Grammar. *Linguistic Inquiry* 8(1). 63–99.

King, Paul. 1994. An Expanded Logical Formalism for Head-Driven Phrase Structure Grammar. Arbeitspapiere des SFB 340 No. 59. Tübingen: Eberhard-Karls-Universität.

King, Paul. 1999. Towards truth in Head-Driven Phrase Structure Grammar. In Valia Kordoni (ed.), *Tübingen Studies in Head-Driven Phrase Structure Grammar* (Arbeitsberichte des SFB 340 No. 132), 301–352. Tübingen: Universität Tübingen.

Kiss, Tibor. 1995. *Infinite Komplementation: Neue Studien zum deutschen Verbum infinitum* (Linguistische Arbeiten 333). Tübingen: Niemeyer.

Klein, Ewan. 2000. A constraint-based approach to English prosodic constituents. In *38th Annual Meeting of the Association for Computational Linguistics: Proceedings of the Conference*, 217–224. Hong Kong: ACL.

Koenig, Jean-Pierre & Frank Richter. 2019. Semantics. In Stefan Müller, Anne Abeillé, Robert D. Borsley & Jean-Pierre Koenig (eds.), *Head-Driven Phrase Structure Grammar: The Handbook* (Empirically Oriented Theoretical Morphology and Syntax). To appear. Berlin: Language Science Press.

Kuhn, Jonas. 1996. An underspecified HPSG representation for information structure. In Jun-ichi Tsuji (ed.), *Proceedings of COLING-96. 16th International Conference on Computational Linguistics (COLING96). Copenhagen, Denmark, August 5–9, 1996*, 670–675. Copenhagen, Denmark: ACL.

Kuhn, Jonas. 2007. Interfaces in constraint-based theories of grammar. In Gillian Ramchand & Charles Reiss (eds.), *The Oxford Handbook of Linguistic Interfaces*, 613–650. Oxford: Oxford University Press.

Levine, Robert D. & Thomas E. Hukari. 2006. *The Unity of Unbounded Dependency Constructions* (CSLI Lecture Notes 166). Stanford: CSLI Publications.

Meurers, Walt Detmar. 1999a. German partial-VP fronting revisited. In Gert Webelhuth, Jean-Pierre Koenig & Andreas Kathol (eds.), *Lexical and Constructional Aspects of Linguistic Explanation* (Studies in Constraint-Based Lexicalism 1), 129–144. Stanford: CSLI Publications.

Meurers, Walt Detmar. 1999b. Raising spirits (and assigning them case). *Groninger Arbeiten zur Germanistischen Linguistik (GAGL)* 43. 173–226.

Meurers, Walt Detmar & Stefan Müller. 2009. Corpora and syntax. In Anke Lüdeling & Merja Kytö (eds.), *Corpus Linguistics: An International Handbook* (Handbücher zur Sprach- und Kommunikationswissenschaft 29), chap. 42, 920–933. Berlin: Mouton de Gruyter.

Meurers, Walt Detmar, Gerald Penn & Frank Richter. 2002. A web-based instructional platform for constraint-based grammar formalisms and parsing. In Dragomir Radev & Chris Brew (eds.), *Effective Tools and Methodologies for Teaching NLP and CL*, 18–25. Proceedings of the Workshop held at 40th Annual Meeting of the Association for Computational Linguistics. Philadelphia, PA: ACL.

Müller, Stefan. 1999a. An HPSG-analysis for free relative clauses in German. *Grammars* 2(1). 53–105.

Müller, Stefan. 1999b. *Deutsche Syntax deklarativ: Head-Driven Phrase Structure Grammar für das Deutsche* (Linguistische Arbeiten 394). Tübingen: Niemeyer.

Müller, Stefan. 2002. *Complex Predicates: Verbal Complexes, Resultative Constructions, and Particle Verbs in German* (Studies in Constraint-Based Lexicalism 13). Stanford: CSLI Publications.

Müller, Stefan. 2003. Solving the bracketing paradox: An analysis of the morphology of German particle verbs. *J. of Linguistics* 39(2). 275–325.

Müller, Stefan. 2007. Qualitative Korpusanalyse für die Grammatiktheorie: Introspektion vs. Korpus. In Gisela Zifonun & Werner Kallmeyer (eds.), *Sprachkorpora – Datenmengen und*

Erkenntnisfortschritt (Institut für Deutsche Sprache Jahrbuch 2006), 70–90. Berlin: Walter de Gruyter.

Müller, Stefan. 2009. On predication. In Stefan Müller (ed.), *Proceedings of the 16th International Conference on Head-Driven Phrase Structure Grammar, University of Göttingen, Germany*, 213–233. Stanford: CSLI Publications.

Müller, Stefan. 2010. Persian complex predicates and the limits of inheritance-based analyses. *Journal of Linguistics* 46(3). 601–655.

Müller, Stefan. 2012. On the Copula, Specificational Constructions and Type Shifting. Ms. Freie Universität Berlin.

Müller, Stefan. 2013a. *Head-Driven Phrase Structure Grammar: Eine Einführung*, 3rd edn. (Stauffenburg Einführungen 17). Tübingen: Stauffenburg Verlag.

Müller, Stefan. 2013b. Unifying everything: Some remarks on Simpler Syntax, Construction Grammar, Minimalism and HPSG. *Language* 89(4). 920–950.

Müller, Stefan. 2014a. Elliptical constructions, multiple frontings, and surface-based syntax. In Paola Monachesi, Gerhard Jäger, Gerald Penn & Shuly Wintner (eds.), *Proceedings of Formal Grammar 2004, Nancy*, 91–109. Stanford: CSLI Publications.

Müller, Stefan. 2014b. Kernigkeit: Anmerkungen zur Kern-Peripherie-Unterscheidung. In Antonio Machicao y Priemer, Andreas Nolda & Athina Sioupi (eds.), *Zwischen Kern und Peripherie* (studia grammatica 76), 25–39. Berlin: de Gruyter.

Müller, Stefan. 2015a. HPSG – A synopsis. In Tibor Kiss & Artemis Alexiadou (eds.), *Syntax – Theory and Analysis: An International Handbook*, 2nd edn. (Handbooks of Linguistics and Communication Science 42.2), 937–973. Berlin: Walter de Gruyter.

Müller, Stefan. 2015b. The CoreGram project: Theoretical linguistics, theory development and verification. *J. of Language Modelling* 3(1). 21–86.

Müller, Stefan. 2016. *Grammatical Theory: From Transformational Grammar to Constraint-Based Approaches* (Textbooks in Language Sciences 1). Berlin: Language Science Press.

Müller, Stefan. 2017. *German Sentence Structure: An Analysis with Special Consideration of So-Called Multiple Fronting* (Empirically Oriented Theoretical Morphology and Syntax). Revise and resubmit. Berlin: Language Science Press.

Müller, Stefan. 2018a. *A Lexicalist Account of Argument Structure: A Discussion of Template-Based Phrasal LFG Approaches and a Lexical HPSG Alternative* (Conceptual Foundations of Language Science 3). Berlin: Language Science Press.

Müller, Stefan. 2018b. Germanic Syntax. Ms. Humboldt-Universität zu Berlin, to be submitted to Language Science Press. Berlin. https://hpsg.hu-berlin.de/~stefan/Pub/germanic.html.

Müller, Stefan. 2018c. *Grammatical Theory: From Transformational Grammar to Constraint-Based Approaches*, 2nd edn. (Textbooks in Language Sciences 1). Berlin: Language Science Press.

Müller, Stefan. 2018d. *A Lexicalist Account of Argument Structure: Template-Based Phrasal LFG Approaches and a Lexical HPSG Alternative* (Conceptual Foundations of Language Science 2). Berlin: Language Science Press. https://doi.org/10.5281/zenodo.1441351.

Müller, Stefan. 2019. HPSG and Construction Grammar. In Stefan Müller, Anne Abeillé, Robert D. Borsley & Jean-Pierre Koenig (eds.), *Head-Driven Phrase Structure Grammar: The Handbook (Empirically Oriented Theoretical Morphology and Syntax)*. To appear. Berlin: Language Science Press.

Müller, Stefan & Walter Kasper. 2000. HPSG analysis of German. In Wolfgang Wahlster (ed.), *Verbmobil: Foundations of Speech-to-Speech Translation* (Artificial Intelligence), 238–253. Berlin: Springer Verlag.

Müller, Stefan & Bjarne Ørsnes. 2013a. Passive in Danish, English, and German. In Stefan Müller (ed.), *Proceedings of the 20th International Conference on Head-Driven Phrase Structure Grammar, Freie Universität Berlin*, 140–160. Stanford: CSLI Publications.

Müller, Stefan & Bjarne Ørsnes. 2013b. Towards an HPSG analysis of object shift in Danish. In Glyn Morrill & Mark-Jan Nederhof (eds.), *Formal Grammar: 17th and 18th International Conferences, FG 2012, Opole, Poland, August 2012, Revised Selected Papers, FG 2013, Düsseldorf, Germany, August 2013: Proceedings* (Lecture Notes in Computer Science 8036), 69–89. Berlin: Springer Verlag.

Müller, Stefan & Bjarne Ørsnes. 2015. *Danish in Head-Driven Phrase Structure Grammar* (Empirically Oriented Theoretical Morphology and Syntax), in preparation. Berlin: Language Science Press. https://hpsg.hu-berlin.de/~stefan/Pub/danish.html.

Oepen, Stephan & Dan Flickinger. 1998. Towards systematic grammar profiling: Test suite technology ten years after. *Journal of Computer Speech and Language* 12(4). (Special Issue on Evaluation), 411–436.

Oepen, Stephan, Erik Velldal, Jan Tore Lønning, Paul Meurer, Victoria Rosén & Dan Flickinger. 2007. Towards hybrid quality-oriented machine translation: On linguistics and probabilities in MT. In *Proceedings of 11th Conference on Theoretical and Methodological Issues in Machine Translation*, 144–153. Skövde, Sweden.

Orgun, Cemil Orhan. 1996. Sign-Based Morphology and Phonology. University of California, Berkeley, dissertation.

Paggio, Patrizia. 2005. Representing information structure in a formal grammar of Danish. In Takashi Washio, Akito Sakurai, Katsuto Nakajima, Hideaki Takeda, Satoshi Tojo & Makoto Yokoo (eds.), *New Frontiers in Artificial Intelligence: Joint JSAI 2005 Workshop Post-Proceedings* (Lecture Notes in Computer Science 4012), 93–102. Berlin: Springer Verlag.

Penn, Gerald. 2004. Balancing clarity and efficiency in typed feature logic through delaying. In Donia Scott (ed.), *Proceedings of the 42nd Meeting of the Association for Computational Linguistics (ACL'04), Main Volume*, 239–246. Barcelona, Spain.

Pollard, Carl 1984. Generalized Phrase Structure Grammars, Head Grammars, and Natural Language. Stanford University, dissertation.

Pollard, Carl 1996. On head non-movement. In Harry Bunt & Arthur van Horck (eds.), *Discontinuous Constituency* (Natural Language Processing 6), 279–305. Veröffentlichte Version eines Ms. von 1990. Berlin: Mouton de Gruyter.

Pollard, Carl 1999. Strong generative capacity in HPSG. In Gert Webelhuth, Jean-Pierre Koenig & Andreas Kathol (eds.), *Lexical and Constructional Aspects of Linguistic Explanation* (Studies in Constraint-Based Lexicalism 1), 281–298. Stanford: CSLI Publications.

Pollard, Carl & Andrew M. Moshier. 1990. Unifying partial descriptions of sets. In Philip P. Hanson (ed.), *Information, Language and Cognition* (Vancouver Studies in Cognitive Science 1), 285–322. Vancouver: University of British Columbia Press.

Pollard, Carl & Ivan A. Sag. 1987. *Information-Based Syntax and Semantics* (CSLI Lecture Notes 13). Stanford: CSLI Publications.

Pollard, Carl & Ivan A. Sag. 1992. Anaphors in English and the scope of Binding Theory. *Linguistic Inquiry* 23(2). 261–303.

Pollard, Carl & Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar* (Studies in Contemporary Linguistics). Chicago, IL & London: University of Chicago Press.

Przepiórkowski, Adam. 1999. On case assignment and "adjuncts as complements". In Gert Webelhuth, Jean-Pierre Koenig & Andreas Kathol (eds.), *Lexical and Constructional Aspects of Linguistic Explanation* (Studies in Constraint-Based Lexicalism 1), 231–245. Stanford: CSLI Publications.

Pullum, Geoffrey K. & Barbara C. Scholz. 2001. On the distinction between generative-enumerative and model-theoretic syntactic frameworks. In Philippe de Groote, Glyn Morrill & Christian Retor (eds.), *Logical Aspects of Computational Linguistics: 4th International Conference* (Lecture Notes in Computer Science 2099), 17–43. Berlin: Springer Verlag.

Reis, Marga. 1982. Zum Subjektbegriff im Deutschen. In Werner Abraham (ed.), *Satzglieder im Deutschen – Vorschläge zur syntaktischen, semantischen und pragmatischen Fundierung* (Studien zur deutschen Grammatik 15), 171–211. Tübingen: Gunter Narr Verlag.

Richards, Marc. 2015. Minimalism. In Tibor Kiss & Artemis Alexiadou (eds.), *Syntax – Theory and Analysis: An International Handbook* (Handbooks of Linguistics and Communication Science 42), 803–839. Berlin: Mouton de Gruyter.

Richter, Frank. 2004. A Mathematical Formalism for Linguistic Theories with an Application in Head-Driven Phrase Structure Grammar. Phil. dissertation (2000), Eberhard-Karls-Universität Tübingen,

Richter, Frank. 2007. Closer to the truth: A new model theory for HPSG. In James Rogers & Stephan Kepser (eds.), *Model-Theoretic Syntax at 10 – Proceedings of the ESSLLI 2007 MTS@10 Workshop, August 13–17*, 101–110. Dublin: Trinity College Dublin.

Richter, Frank & Manfred Sailer. 2009. Phraseological clauses as Constructions in HPSG. In Stefan Müller (ed.), *Proceedings of the 16th International Conference on Head-Driven Phrase Structure Grammar, University of Göttingen, Germany*, 297–317. Stanford: CSLI Publications.

Riehemann, Susanne. 1997. Idiomatic Constructions in HPSG. Paper presented at the 1997 HPSG conference, Ithaca.

Sag, Ivan A. 1997. English relative clause constructions. *Journal of Linguistics* 33(2). 431–484.

Sag, Ivan A. 2007. Remarks on locality. In Stefan Müller (ed.), *Proceedings of the 14th International Conference on Head-Driven Phrase Structure Grammar*, 394–414. Stanford: CSLI Publications.

Sag, Ivan A. 2010. English filler-gap constructions. *Language* 86(3). 486–545.

Sag, Ivan A. 2012. Sign-Based Construction Grammar: An informal synopsis. In Hans C. Boas & Ivan A. Sag (eds.), *Sign-based Construction Grammar* (CSLI Lecture Notes 193), 69–202. Stanford: CSLI Publications.

Sag, Ivan A., Rui Chaves, Anne Abeillé, Bruno Estigarribia, Frank Van Eynde, Dan Flickinger, Paul Kay, Laura A. Michaelis-Cummings, Stefan Müller, Geoffrey K. Pullum & Thomas Wasow. 2020. Lessons from the English auxiliary system. *J. of Linguistics* 56. Ahead of print (online). https://doi.org/ 10.1017/S002222671800052X.

Sag, Ivan A. & Janet Dean Fodor. 1994. Extraction without traces. In Raul Aranovich, William Byrne, Susanne Preuss & Martha Senturia (eds.), *Proceedings of the Thirteenth West Coast Conference on Formal Linguistics*, 365–384. Stanford: CSLI Publications/SLA.

Sag, Ivan A., Philip Hofmeister & Neal Snider. 2007. Processing complexity in subjacency violations: The Complex Noun Phrase Constraint. In Malcolm Elliott, James Kirby, Osamu Sawada, Eleni Staraki & Suwon Yoon (eds.), *Proceedings of the 43rd Annual Meeting of the Chicago Linguistic Society*, 215–229. Chicago, IL: Chicago Linguistic Society.

Sag, Ivan A. & Thomas Wasow. 2011. Performance-compatible competence grammar. In Robert D. Borsley & Kersti Börjars (eds.), *Non-Transformational Syntax: Formal and Explicit Models of Grammar: A Guide to Current Models*, 359–377. Oxford, UK/Cambridge, MA: Blackwell Publishing Ltd.

Sag, Ivan A., Thomas Wasow & Emily M. Bender. 2003. *Syntactic Theory: A Formal Introduction*, 2nd edn. (CSLI Lecture Notes 152). Stanford: CSLI Publications.

Schäfer, Roland & Felix Bildhauer. 2012. Building large corpora from the web using a new effcient tool chain. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Jan Odijk & Stelios Piperidis (eds.), *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, 486–493. Istanbul, Turkey: European Language Resources Association (ELRA).

Schlangen, David, Alex Lascarides & Ann Copestake. 2003. Resolving underspecification using discourse information. *Perspectives on Dialogue in the New Millennium* 114. 287.

Shieber, Stuart M. 1986. *An Introduction to Unification-Based Approaches to Grammar* (CSLI Lecture

Notes 4). Stanford: CSLI Publications.
Siegel, Melanie. 2000. HPSG analysis of Japanese. In Wolfgang Wahlster (ed.), *Verbmobil: Foundations of Speech-to-Speech Translation* (Artificial Intelligence), 264–279. Berlin: Springer Verlag.
Soehn, Jan-Philipp & Manfred Sailer. 2008. At first blush on tenterhooks: About selectional restrictions imposed by nonheads. In Gerhard Jäger, Paola Monachesi, Gerald Penn & Shuly Wintner (eds.), *Proceedings of Formal Grammar 2003, Vienna, Austria*, 149–161. Stanford: CSLI Publications.
Steedman, Mark J. & Jason Baldridge. 2006. Combinatory categorial grammar. In Keith Brown (ed.), *Encyclopedia of Language and Linguistics*, 2nd edn. 610–621. Oxford: Elsevier.
Tanenhaus, Michael K., Michael J. Spivey-Knowlton, Kathleen M. Eberhard & Julie C. Sedivy. 1995. Integration of visual and linguistic information in spoken language comprehension. *Science* 268(5217). 1632–1634.
Wahlster, Wolfgang (ed.), 2000. *Verbmobil: Foundations of Speech-to-Speech Translation* (Artificial Intelligence). Berlin: Springer Verlag.
Wasow, Thomas. 2019. Processing. In Anne Abeillé, Robert D. Borsley, Jean-Pierre Koenig & Stefan Müller (eds.), *Head-Driven Phrase Structure Grammar: The Handbook* (Empirically Oriented Theoretical Morphology and Syntax), to appear. Berlin: Language Science Press.
Webelhuth, Gert. 1985. German is configurational. *The Linguistic Review* 4(3). 203–246.
Wechsler, Stephen Mark & Larisa Zlatić. 2003. *The Many Faces of Agreement* (Stanford Monographs in Linguistics). Stanford: CSLI Publications.
Wechsler, Stephen, Jean-Pierre Koenig & Anthony Davis. 2019. Argument Structure and Linking. In Stefan Müller, Anne Abeillé, Robert D. Borsley & Jean-Pierre Koenig (eds.), *Head-Driven Phrase Structure Grammar: The Handbook* (Empirically Oriented Theoretical Morphology and Syntax). To appear. Berlin: Language Science Press.
Wilcock, Graham. 2005. Information structure and Minimal Recursion Semantics. In Antti Arppe, Lauri Carlson, Krister Lindén, Jussi Piitulainen, Mickael Suominen, Martti Vainio, Hanna Westerlund & Anssi Yli-Jyrä (eds.), *Inquiries into Words, Constraints and Contexts: Festschrift for Kimmo Koskenniemi on his 60th Birthday* (CSLI Studies in Computational Linguistics ONLINE), 268–277. Stanford: CSLI Publications.
Wittenberg, Eva, Ray Jackendoff, Gina Kuperberg, Martin Paczynski, Jesse Snedeker & Heike Wiese. 2014. The processing and representation of light verb constructions. In Asaf Bachrach, Isabelle Roy & Linnaea Stockall (eds.), *Structuring the Argument* (Language Faculty and Beyond 10), 61–80. Amsterdam: John Benjamins Publishing Co.
Wittenberg, Eva & Maria Mercedes Piñango. 2011. Processing light verb constructions. *The Mental Lexicon* 6(3). 393–413.
Yip, Moira, Joan Maling & Ray Jackendoff. 1987. Case in tiers. *Language* 63(2). 217–250.
Zaenen, Annie, Joan Maling & Höskuldur Thráinsson. 1985. Case and grammatical functions: The Icelandic passive. *Natural Language and Linguistic Theory* 3(4). 441–483.

Timothy Osborne

13 Dependency Grammar

Abstract: Dependency Grammar (DG) is a family of approaches to the study of syntax of natural languages that all take dependency as the principle organizing words into greater units of syntax (phrases and clauses). This contribution examines some of the definitional properties that distinguish syntax in terms of dependencies from syntax in terms of the constituencies of phrase structure grammar. It discusses the tools that dependency grammarians use to analyze sentence structure and it considers the goals that they pursue as well as the data they consider and produce in the process. It also presents the arguments in favor of doing syntax in terms of dependencies, these arguments being simplicity and accuracy.

1 Introduction

The term *dependency grammar* (DG) denotes a family of approaches to the syntax and grammar of natural languages that take *dependency*, as opposed to *phrase structure*, as the principle of organization grouping units of syntax (i. e. words) together to create larger units (phrases and clauses).[1] Dependency syntax has a long and venerable history, although its influence on mainstream syntactic theorizing in the last 60 years has been minor. Interesting in this regard is that Lucien Tesnière's (1894–1954) main work, *Éléments de syntaxe structurale*, appeared posthumously in 1959, just two years after the appearance of Noam Chomsky's first major work on syntax, *Syntactic Structures*, in 1957. Tesnière is the most prominent name associated with dependency syntax; he is considered the father of modern DGs.

The history of dependency syntax begins very early; it certainly predates phrase structure by hundreds if not thousands of years – that is, if claims about dependency being recognizable in the works of the ancient Sanskrit grammarian Panini are accurate. Dependency has been identified as the principle organizing syntactic units in the works of medieval Arabic grammarians (Versteegh, 1997, 45) and modistic medieval grammarians (e. g. Baum, 1976, 29–30; Covington, 1984; Percival, 1990). Worth noting in this regard, though, is that dependency and phrase structure have been present in related grammar traditions for many centuries, the former in the form of traditional case government, i. e. the verb governs the case of its nominal complements, and the latter in the form of the subject-predicate division of term logic from Aristotelian antiquity.

1 The main acronyms used in this contribution are listed here for easy reference: DG (dependency grammar), MTT (Meaning-Text Theory), PSG (phrase structure grammar), UD (Universal Dependencies).

Acknowledgement: Many thanks to Stefan Mueller for his feedback on this article.

https://doi.org/10.1515/9783110540253-013

This contribution presents and discusses the properties of dependency syntax (Section 2). It considers the conventions employed to show dependencies and compares them to the conventions used to show *phrase structure* (Section 3.1). The tools that dependency syntax uses to explore and analyze syntactic structures are considered, such as the ability of dependency analyses to abstract away from linear order (Section 3.2), the role of the syntactic functions (Section 3.3), and the notion of valency and the distinction between arguments and adjuncts (Section 3.4). It includes some history of DG and considers the goals of those who are using DG (Sections 4.1–4.2). The sorts of data that are used and produced by dependency grammarians are considered as well (Section 5). Two arguments in favor of dependency over phrase structure are also given, these arguments being simplicity and accuracy (Sections 6.1–6.2). A sample analysis of a complex sentence is then given (Section 7) before concluding.

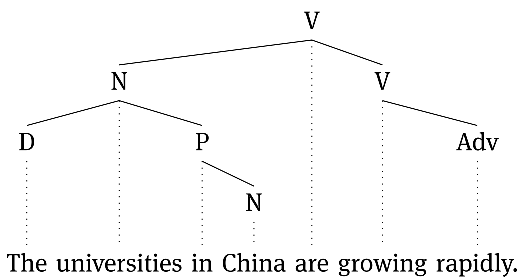
2 Properties

This section considers some of the main properties of dependency syntax. Comparisons of dependency and phrase structure like the one given next are common in the DG literature (e. g. Baumgärtner, 1970, Matthews, 1981, 71–95; Tarvainen, 1981, 11–13; Mel'čuk, 1988, 12–17; Schubert, 1987, 17–20; Jung, 1995, 15–27; Heringer, 1996, 27–29; Uzonyi, 2003; Hudson, 2010, 147–150).

Dependency is a strict *parent-child relation* between words. Given two co-occurring words whereby the one can appear by virtue of the appearance of the other, a dependency links the two. The dependency is directed; the one word, the *head* (or *governor* or *parent*), licenses the appearance of the other, the *dependent* (or *child*). The dependent *depends* on its head. Dependency stands in contrast to *phrase structure*, which is a relation between *siblings*; two (or more) units of syntax (words, phrases, or clauses) appear equi-level as siblings, forming a greater unit. These two relations, dependency and phrase structure, can be represented using tree diagrams. Some basic traits of dependency syntax are established next by comparing the tree structures of dependency with the corresponding tree structures of phrase structure.

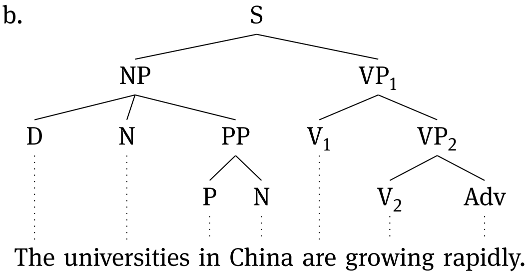
The next tree illustrates a dependency analysis of the given sentence:

(1) a. – A dependency analysis



The node labels identify the syntactic category of the words. The slanted edges indicate dependencies. The vertical lines are projection edges, each word projecting a node. The next tree is a phrase structure analysis of the same sentence:[2]

(1) b. – A phrase structure analysis



The part-whole relation of phrase structure is apparent in this tree insofar as each part (word or phrase) combines with another part (or two parts) to create a whole. When two (or three) parts combine, the relationship between them is symmetrical insofar as they appear equi-level in the structure.

The dependency analysis given as diagram (1a) is characterized by the following properties:

1. Verb centrality,
2. One-to-one mapping,
3. Strict headedness,
4. Syntactic structure in terms of trees, and
5. Simultaneous presence of both hierarchy and linearity.

Of these five properties, the phrase structure analysis given as diagram (1b) also adheres to the fourth and fifth properties. Concerning the fourth, the two diagrams have in common that they are rooted trees, that is, each node in the tree, excepting the root node, has one and only one parent node. Concerning the fifth property, both (1a) and (1b) organize the words hierarchically (think *dominance*) and linearly (think *precedence*) such that both encode and show the actual order of the words in the sentence.

The two trees differ regarding the first three properties, however. Concerning the first property, verb centrality in (1a) is visible in the fact that the finite verb *are* is positioned as the root of the entire tree. Verb centrality has been a main characteristic of

2 Note that "N" appears in the phrase structure tree over *China* instead of "NP". There is a tradition in this area that pre-dates Chomsky's works and that sees phrases as necessarily consisting of two or more words. This older tradition is assumed here for the phrase structure trees because it aids the comparison of dependency with phrase structure.

dependency syntax from the earliest works, and it stands in opposition to the binary subject-predicate division of most PSGs – this binary division is visible in (1b) in the manner in which the sentence S is first divided into an NP and a VP.

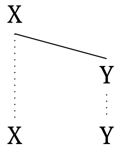
Verb centrality can be viewed as a consequence of the second property listed above (and of the third property). Given the necessity of one-to-one mapping (words to nodes), one of the words has to be positioned as the root of the entire structure. Intuitively, the verb was/is the most compelling choice for this role. Dependency is, namely, a strict one-to-one mapping of atomic units of syntax (words) to nodes in the structure (c. f. Mel'čuk, 1979, 96; Mel'čuk and Pertsov, 1987, 48, 57–58; Kahane, 1996, 45; Schubert, 1987, 78–86, 129; Engel, 1994, 25, 28; Hudson, 2003, 520;, 2007, 183; Carnie, 2010, 177; Osborne, 2005, 253, 2008, 1123), whereas phrase structure is a one-to-one-or-more mapping.[3] There are seven words in the sentence in (1) and there are seven nodes in the dependency tree (1a) (7 words and 7 nodes). In contrast, there are 12 nodes in the phrase structure tree (1b) (7 words but 12 nodes). Thus, dependency can often be distinguished from phrase structure by a simple counting of words and nodes. If the number of nodes in the sentence structure matches exactly the number of atomic units of syntax present, then one is likely dealing with a dependency analysis. If, in contrast, the number of nodes exceeds the number of atomic units, then at least some measure of phrase structure is present.

The fact that dependency structures are always completely headed, the third property, is another aspect of dependency that helps distinguish it from phrase structure. Examining the dependency tree (1a), one sees that excepting the root node V, each node present in the tree has one and only one parent node. This means that given any two words that are connected by a dependency, one of them is head over the other, or taking the opposite perspective, one of them is dependent on the other. Phrase structure analyses do not necessarily do the same. This is evident at the top of tree (1b), where the whole is viewed as an S (sentence) that is divided into an NP and a VP. The category status of S is distinct from that of NP and VP.

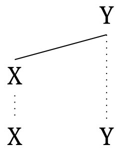
The point at issue is understood in terms of endo- and exocentrism, a distinction that goes back to Bloomfield (1933, 194–196). Dependency can hardly acknowledge exocentric structures, whereas phrase structure can if it so chooses. The following simple abstract trees illustrate the point:

3 A node is understood to mark a distinct grouping of words. If two or more vertices in a tree structure mark the same one grouping of words, then they together qualify as just one node. This qualification is necessary to preempt objections that can be raised about the claim here that one-to-one mapping (words to nodes) is a defining property of dependency syntax. Consider, for instance, that unary branching increases the number of nodes without increasing the number of distinct groupings of words.

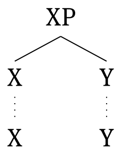
(2) a. – Dependency structure



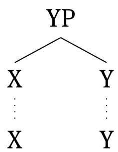
b. – Dependency structure



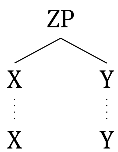
(3) a. – Endocentric phrase structure



b. – Endocentric phrase structure



c. – Exocentric phrase structure



The phrase structure given as (3c) is exocentric because the entirety, i. e. ZP, is neither a projection of X nor of Y. Dependency's inability to acknowledge such exocentric combinations means that it views all syntactic structures as necessarily headed.

Worth noting and emphasizing at this early stage of the discussion is that DGs are a varied bunch, meaning that the one or other DG might not adhere to one or more of the five properties listed and discussed in this section. A similar statement is of course true of PSGs, since they are also a quite varied bunch. The presentation of dependency syntax below attempts to accommodate the diversity among DGs by acknowledging and discussing some of the properties that allow one to distinguish between them.

3 Tools

The following subsections consider the tools of DG, whereby "tools" is understood in two ways: the first is in terms of the conventions used to indicate dependencies, and the second is in terms of the means DGs use to analyze syntactic structures, these means being more or less unique to DG. Three areas are considered in the latter regard: how DGs understand linear order, how they treat the syntactic relations, and how they understand and distinguish between arguments and adjuncts.

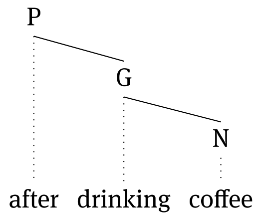
3.1 Conventions

DGs use a variety of visual devices to indicate the presence of dependencies. Three of these are listed next: 1) tree(-like) diagrams, 2) arced arrows, and 3) brackets. These three conventions are also suited for indicating the presence of phrase structure. The illustrations that follow hence give dependency analyses side by side with the corresponding phrase structure analyses. Comparison across the two ways of conceiving of syntactic organization promotes an understanding of the one as well as of the other.

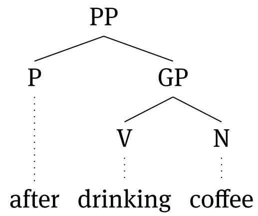
Tree diagrams like the ones produced in the previous section are often employed to show syntactic structure. Further such tree diagrams are given next:[4]

Dependency structures **Phrase structures**

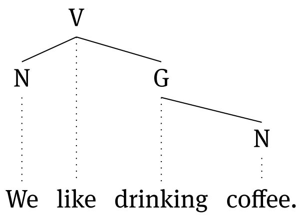
(4) a.



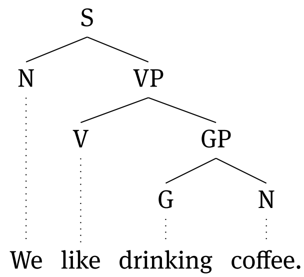
b.



(5) a.



b.

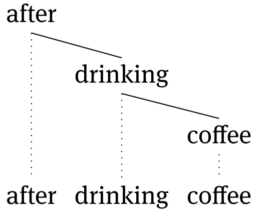


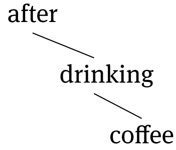
Dependency trees similar to the a-trees here are employed by many DGs (e. g. Hays, 1964; Kunze, 1975; Heringer, 1996; Groß, 1999; Eroms, 2000; etc.). Note, however, that there are a couple of variations on the convention given as the a-examples.

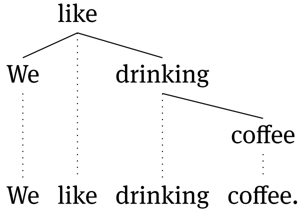
When using tree-diagrams to show dependencies, the words themselves are often used as the node labels, and the string of words below and the projection lines can then be omitted, e. g.

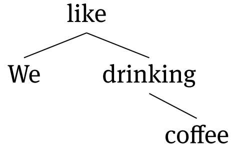
4 G = gerund, GP = gerund phrase.

Dependency structures

(6) a.


b.


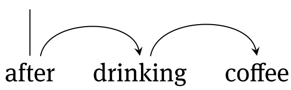
(7) a.


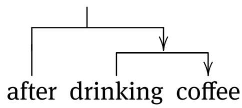
b.


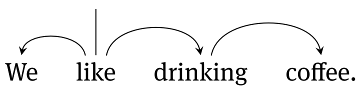
The convention employed in the a-trees is the one preferred in Thomas Groß' and my works (e. g. Osborne et al., 2011, 2012; Groß and Osborne, 2013; Osborne and Groß, 2012, 2016). The convention shown with the b-trees is minimal and completely transparent. Such trees are easy to produce in informal environments, for instance when giving syntactic analyses on the chalk board in the classroom.

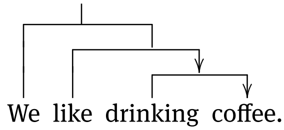
The use of arced arrows to illustrate the same dependencies and phrase structures is illustrated next:

Dependency structures **Phrase structures**

(8) a.


b.


(9) a.


b.


This convention for showing dependencies is preferred in the field of natural language processing (NLP), e. g. (Kübler et al., 2009). It is also the convention preferred in the Word Grammar framework (Hudson, 1984, 1990, 2007, 2010). Meaning-Text Theory (MTT) (Mel'čuk, 1979, 1988, 2009; Mel'čuk and Pertsov, 1987) also uses arced arrows extensively (in addition to tree diagrams). Concerning the phrase structure diagrams (8b) and (9b), they are unusual due to the appearance of the arrow heads; the addition of such arrow heads identifies heads.

Interestingly, brackets can also be used to show dependencies:

		Dependency structures		**Phrase structures**
(10)	a.	[after [drinking [coffee]]]	b.	[[after] [[drinking] [coffee]]]
(11)	a.	[[We] like [drinking [coffee]]].	b.	[[We] [[like] [[drinking] [coffee]]]].

The brackets are used consistently. Words appearing lower in the structure appear enclosed in more sets of brackets. An advantage that dependency has with respect to brackets is that it identifies heads, for heads and dependents are present in the a-examples. Phrase structure, in contrast, does not identify heads; the phrasal constituents in the b-examples are all shown as exocentric. In order to identify heads, phrase structure needs labels, e. g. *[$_{PP}$ [$_{P}$ after] [$_{VP}$ [$_{V}$ drinking] [$_{N}$ coffee]]]*.

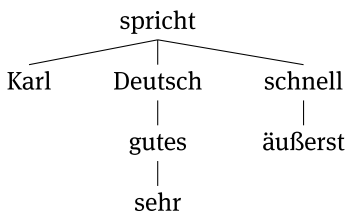
3.2 Linear order

The traditional DG stance toward linear order is that it is secondary to hierarchical order in the mind of a speaker (Tesnière, 1959/2015, Ch. 6). This aspect of early DGs is evident in the sentence diagrams, for instance in those of Franz Kern (1883) and Lucien Tesnière (1959). These diagrams typically show the organization of words with respect to hierarchy, but not with respect to linearity. In other words, the diagrams abstract away from actual word order. In fact, many accounts of the distinction between dependency and phrase structure emphasize that dependency itself is independent of actual word order, whereas linear order has been deemed by some as an inseparable trait of phrase structure (cf. Baumgärtner, 1970, 53; Korhonen, 1977, 31; Tarvainen, 1981, 13; Mel'čuk and Pertsov, 1987, 7; Schubert, 1987, 63; Jung, 1995, 16; Hudson, 2010, 170–172).

Two dependency trees illustrating this practice of abstracting away from actual word order are provided next:

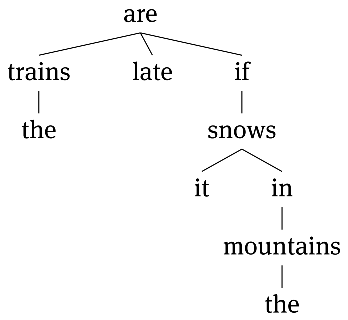
(12) a. Karl spricht sehr gutes Deutsch äußerst schnell.
K. speaks very good German extremely quickly

b.



(Tarvainen, 1981, 6)

(13) a. The trains are late if it snows in the mountains.

b. 

(Schubert, 1987, 95)

These trees give the hierarchical order of the words constituting the sentence at hand each time without also showing the actual word order. This fact is evident in the manner in which many of the words are positioned directly below their parent word. DGs that do this, of which there are many (e. g. Kern, 1883; Tesnière, 1959; Tarvainen, 1981; Schubert, 1987; Mel'čuk, 1988, 2009; Lobin, 1993; Jung, 1995; Engel, 1994), can be construed as multistratal in syntax because they assume that at some level of linguistic representation, the words are present and organized hierarchically but are not yet ordered with respect to linearity.

Numerous other DGs do not do this, however, that is, they do not separate linear order from hierarchical order. In not doing so, they are essentially, although they may not state so much explicitly, pursuing a monostratal approach to syntax, for they are granting hierarchical order and linear order equal status in the system, that is, the one cannot exist without the other (e. g. Hudson, 1984, 1990, 2007, 2010; Starosta, 1988; Heringer, 1996; Bröker, 1999, Groß, 1999; Osborne, 2005, 2008).

The ability of dependency syntax to easily abstract away from linear order is viewed by some as an advantage that dependency has over phrase structure. It is certainly part of the reason why dependency is deemed well-suited as the basis for modeling the syntax of languages that have freer word order than English.

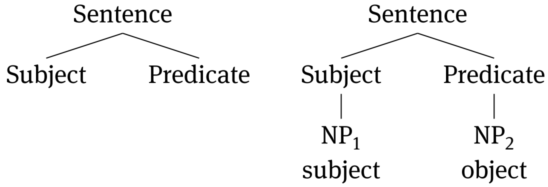
3.3 Syntactic relations

The *syntactic relations* (also called *grammatical relations* or *syntactic functions*) play a major role in many DGs. These relations are considered primitive, and their presence and importance can be construed as a consequence of the tendency to abstract away from linear order, as discussed in the previous section. Consider in this regard that PSG has traditionally defined the grammatical relations (subject, object, etc.) in terms of the constellation (Chomsky, 1965, 64). In grammars of English, the subject originally appears in a position external to the VP constituent, whereas the object is originally generated inside the VP. In contrast, many DGs take the subject and object relations to be given, i. e. primitives of the theory; they exist independently of the particular syntactic constellation in which they appear.

The varying approaches to the syntactic functions can be understood in terms of the question *What comes first?* In Chomskyan PSGs, the constellation comes first and then the syntactic functions are defined in terms of the constellation, as sketched next:

Constellation first **Syntactic relations second**

(14)

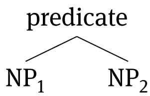


NP_1 is identified as the subject by virtue of the fact that it first appears in the constellation outside of the predicate, and NP_2 is identified as the object because it first appears in the constellation inside the predicate.

In contrast, many DGs see this order as reversed. The subject and object relations exist first, independently of the constellation in which they appear. The particular syntactic configuration that then occurs is determined in part by which syntactic functions are present. This order is schematized as follows:

Syntactic relations first **Constellation second**

(15) NP_1 (subject of predicate)
NP_2 (object of predicate)

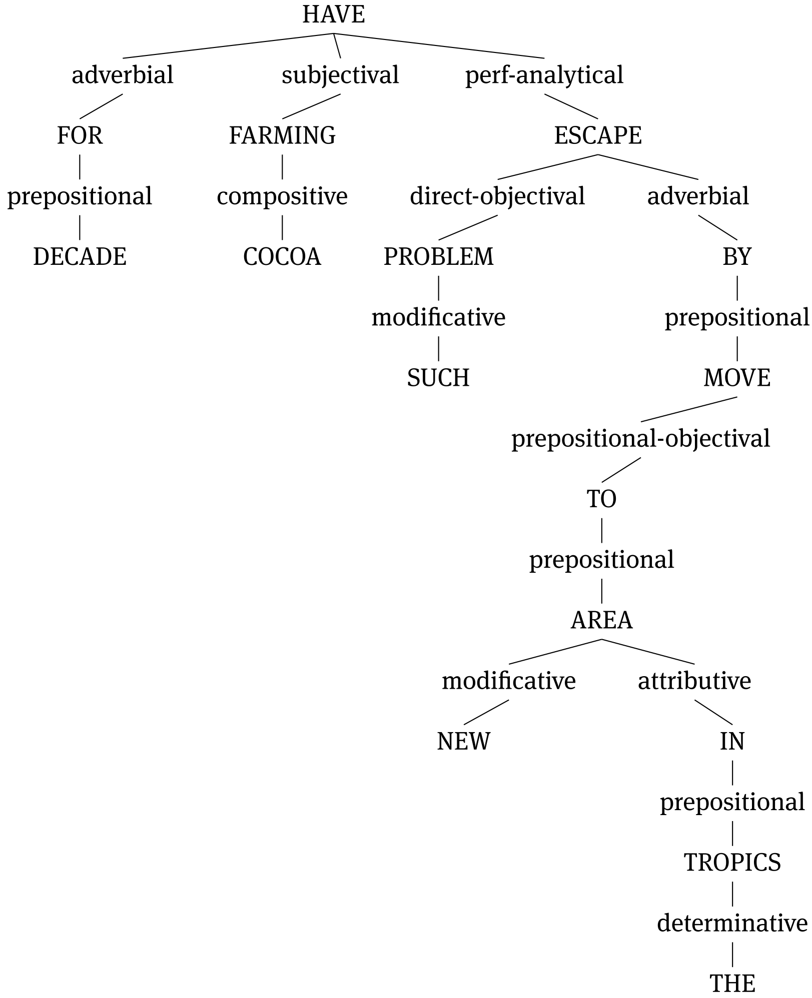


NP_1, since it is the subject, influences the shape of the constellation because it must precede the predicate, and NP_2, since it is the object, influences the shape of constellation insofar as it necessarily follows the predicate. Observe that the understanding of the syntactic relations suggested with (14) is based on the traditional subject-predicate division of Aristotelian term logic, which takes the predicate to be everything except the subject. In contrast, the understanding of the syntactic relations suggested in (15) views the predicate in the more modern sense of Fregean predicate calculus. The predicate is a semantic entity that opens slots for arguments, relating these arguments to each other.

Deciding which of these two approaches to the syntactic relations is more principled is a difficult, thorny issue, and one that is avoided here. What is of interest in the current context of DG, however, is simply that many DGs view the primitive nature of the syntactic relations as a defining characteristic of dependency syntax. They posit inventories of syntactic relations and assume that each and every dependency bears a syntactic relation. This is particularly true of the MTT framework, which emphasizes the importance of the syntactic relations (e. g. Mel'čuk and Pertsov, 1987, 61–72, Mel'čuk, 2009, 52–58). The next tree showing the syntactic functions is reproduced from Mel'čuk, (2009, 7):

(16) a. For decades, cocoa farming has escaped such problems by moving to new areas in the tropics.

b.



The labels on the dependency edges show that the MTT framework necessarily assumes that each and every dependency bears a syntactic function, whereby the inventory of syntactic functions is a primitive of the language at hand. A finite inventory of syntactic functions (in the dozens) is a characteristic trait of the syntax of the language.

DGs vary in the importance that they attach to the syntactic functions. For instance, Kern (1883) did not discuss the syntactic functions beyond acknowledging the importance of subjects. Tesnière's account of valency acknowledged 1st actants (subjects), 2nd actants (direct objects), and 3rd actants (indirect objects) as well as *circumstants* (adjuncts), but beyond these rather broad functions, he had little to say about the functions. In the field of NLP (see Section 4.2), however, the syntactic functions are particularly important, since by assigning each and every dependency a syntactic function, one can efficiently search treebanks for specific syntactic functions, and one is thus able to easily locate specific syntactic phenomena that one is interested in investigating.

3.4 Valency, subjects, other arguments, and adjuncts

One particular tool of analysis that DGs rely on heavily is the handling of subjects. Due to the fact that the verb is the root of sentence, the subject is reduced in status in a sense; it is treated as a dependent of the verb similar to the manner in which the object is a dependent of the verb. Tesnière developed this aspect of his approach to syntax in terms of valency. Valency is the notion that verbs and other valency carriers determine their syntactic environment; they open slots for actants the form of which they determine. The subject is one of these actants. The actants that a verb takes constitute together the valency of the verb. Valency has been a central subtheory within DG since Tesnière developed the concept in detail (1959/2015, Book D). Subordinating the subject actant to the verb is advantageous in a couple of respects. It allows dependency syntax to more easily accommodate subject-less sentences, such as imperatives and when subject pronouns are omitted in so-called pro-drop languages (cf. Järventausta, 2003, 787–789).

As noted above, Tesnière posited three actant types that verbs can take: *first actants*, *second actants*, and *third actants*. A few verbs take no actant at all, whereas most necessarily take a first actant. The extent to which verbs take a second or third actant varies greatly depending on the verb at hand. The next examples illustrate (just the most) basic types of verbal valency carriers:

	Avalent	
(17)	*Pluit*. Latin for 'It is raining.'	PLUERE [ ]
	Monovalent	
(18)	Jim stutters.	STUTTER [N]
	Bivalent (transitive)	
(19)	Jim studies political science.	STUDY [N, N]
	Trivalent (ditransitive)	
(20)	Cindy sent us a portrait of her.	SEND [N, N, N]

The valency of verbs and other valency carriers can be given in the manner shown on the right: the valency carrier is listed using small caps in its citation form and its actants are then enclosed in square brackets to its right (N = nominal). While Tesnière did not produce such valency frames, he did examine the traits of verb valency at length. He explored various mechanisms that alter the basic valency of verbs, e. g. causatives, reciprocals and reflexives, voice (passive), recessive, etc.

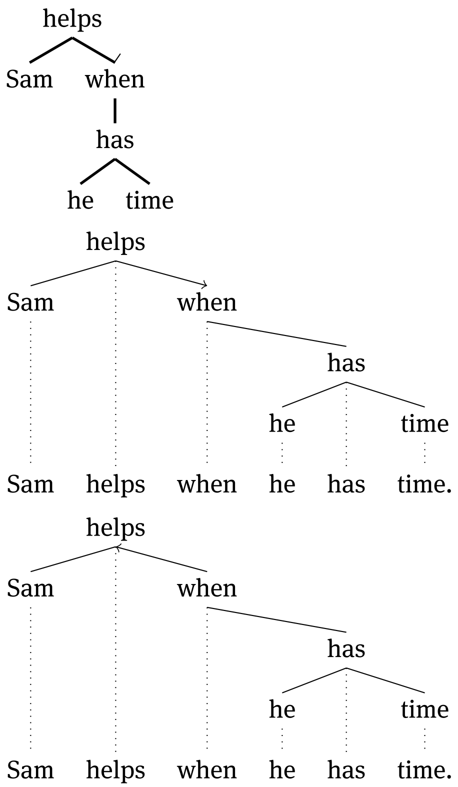
Much of the literature in Germany on valency theory starting in the 1970s was primarily concerned with distinguishing between arguments (Ger. *Ergänzungen*, Tesnière's actants) and adjuncts (Ger. *freie Angaben*, Tesnière's circumstants) – see the contributions on valency theory in Ágel et al. (2003, 2006). While most theories of syntax are like DG in that they view the argument vs. adjunct distinction as a primitive of the lexicon, the manner in which they identify the two in syntactic structures can vary significantly. X-bar theoretic approaches distinguish adjuncts from specifiers and complements by virtue of their position in the X-bar schema; adjuncts usually appear as sister and/or daughter constituents of bar-level projections of the head. In this manner, X-bar theoretic approaches identify adjuncts in sentence structures by virtue of where they appear in the X-bar schema. The same is certainly not true of dependency analyses, since the minimal sentence structures that dependency necessitates cannot grant adjuncts distinctive positions in the hierarchy of structure in relation to their heads.

Given this state of affairs, some DGs choose to indicate the presence of adjuncts in trees using a particular visual convention, e. g.

(21) a. – Engel (1994, 44)

b. – Eroms (2000, 85–6)

c. – Osborne (2012, 30)



The passage cited each time is where the convention for identifying adjuncts in trees is given – the sentence itself is not employed in the sources. The conventions Engel and Eroms use to identify *when he has time* as an adjunct are closely similar, whereas the convention I use has the indicator of adjunct status appearing on the opposite end of the dependency edge. Others use similar conventions for identifying adjuncts in tree structures (c. f. Tesnière, 1959/2015, 36; Baum, 1976, 79; Tarvainen, 1981, 61; Jung, 1995, 111–6; Uzonyi, 2003, 237).

The arrow pointing away from the adjunct towards its head in (21c) is an appropriate means for identifying adjuncts in trees, since it indicates that semantic selection is pointing in the opposite direction of what is normal, for the essence of most adjuncts is that they semantically select their head (or governor). Note that this convention of indicating adjuncts in dependency trees has been taken for granted above.

4 Goals

The next subsections consider the goals of those who have chosen to use dependency as the basis for syntactic analyses. Two major goals are addressed: the goal of establishing comprehensive and theoretically stringent frameworks for the analysis of natural language syntax and the goal of serving efficiently as a basis for the automated parsing of natural language texts.

4.1 Goal 1: Establishing comprehensive theories of syntax

Tesnière's ideas were well received above all in the two Germany's in the 1960s. By the 1970s, DG had become a known entity; it was recognized in some European countries as an alternative way to do syntax, i. e. alternative to the transformational syntax associated with Chomsky and his MIT school. The nature of DG at that time was, though, primarily focusing on the notion of valency as just discussed in the previous section. Worth noting in this regard, however, is that for Tesnière and for DGs more generally, valency theory was and is merely an important subtheory of the greater approach to syntax (cf. Järventausta, 2003, 783). Nevertheless, Tesnière's ideas received widespread acknowledgement in the two Germanys, so much so that to this day, courses and content on DG and valency theory are not uncommon at German universities, whereas they are seldom encountered at universities outside of Germany.

Numerous works of dependency syntax have been produced in the German speaking world (e. g. Baum, 1976; Tarvainen, 1981; Lobin, 1993; Engel, 1994; Jung, 1995; Heringer, 1996; Bröker, 1999; Groß, 1999; Eroms, 2000). The two-volume collection of essays on dependency and valency theory edited by Ágel et al. (2003, 2006), mentioned above, is worth taking note of in this regard. These two volumes contain many dozens of contributions in English and German by over a hundred authors. They stand as a primary resource for information about DG and valency theory. They address areas of natural language grammar and syntax such as language processing, pedagogical applications, psychological reality of dependencies, language change as exemplified by case studies, etc.

By the late 1970s, some prominent DG works outside of Germany were being written. For instance, the works of Igor Mel'čuk began to appear in English (Mel'čuk, 1979). Mel'čuk is a Russian/Canadian linguist who emigrated from Russia to Canada in the 1970s – he is at present an emeritus professor at the University of Montreal. His works and those of his collaborators in the MTT framework (e. g. Mel'čuk and Pertsov, 1987; Mel'čuk, 1988, 2009) have been influential in establishing dependency syntax as a sophisticated and theoretically stringent basis for describing and explaining natural language syntax. Mel'čuk's works remain central to the body of literature on dependency syntax.

Richard Hudson, currently an emeritus professor at the University College of London, began using dependency in the 1970s as the basis for his framework of syntax and grammar (Hudson, 1976), which is now known as Word Grammar (Hudson, 1984, 1990, 2007, 2010). Like Mel'čuk and his collaborators, Hudson and his collaborators have developed a sophisticated and theoretically stringent basis for describing and explaining natural language syntax. Hudson's journal articles in the 1980s on dependency syntax contributed greatly to the general awareness of dependency as a principled alternative (as opposed to phrase structure) for exploring the syntax of natural languages (e. g. Hudson, 1980a, 1980b, 1987).

The late Stanley Starosta (1939–2002) also played an important role in establishing dependency as a basis for doing theoretical syntax. Starosta was a professor at the University of Hawaii. The dependency-based framework he and his collaborators developed is known as *Lexicase Grammar*, or just *Lexicase* for short (Starosta, 1988, 2003). Starosta was investigating the Oceanic languages; he was fluent in Mandarin and studied and did fieldwork on numerous other languages that have not received as much attention from theoretical syntax as the European languages (Formosan languages, Japanese, Korean, Munda languages, Philippine languages, Sino-Tibetan languages, Thai, etc.). Starosta was a vehement advocate of dependency syntax, for example in his stance that VPs do not exist (Bender, 2002).[5]

The German schools (Baum, Tarvainen, Lobin, Engel, Jung, Heringer, Bröker, Groß, Eroms), Mel'čuk, Hudson, and Starosta are of course just a few of the prominent

5 The stance that DGs adopt concerning the (non)existence of a VP constituent is worth noting, since there has been some confusion in this area. While it is fair and accurate to say that DGs in general reject the existence of finite VP constituents, many DGs do acknowledge the existence of nonfinite VP constituents. This fact is evident in the tree diagrams produced in this contribution that grant nonfinite VPs the status of complete subtrees. See tree (1a), for instance, in which the nonfinite VP *growing rapidly* is a complete subtree and hence a constituent in the relevant sense.

names who have helped establish DG as a known entity in linguistics and dependency as a basis for comprehensive theories of natural language syntax. The discussion now turns to the use of dependency in NLP.

4.2 Goal 2: Serving as the basis for efficient automated parsing

The main motor that is driving interest in dependency syntax at present is its use for the automated parsing of sentences. The value and potential of dependency for serving as a basis for automated text parsing was recognized in the early 1960s by David Hays (1928–1995). Hays was a linguist and computer scientist working at the RAND Corporation in Santa Monica California. He was a pioneer of computational linguistics and one of the earliest advocates of dependency syntax in the United States. His article in *Language* in 1964, entitled *Dependency theory: A formalism and some observations* contributed significantly to the growing awareness of dependency as a formalism distinct from the phrase structure associated with Chomsky's Transformational Grammar, which was inspiring the linguistics world at that time. Hays authored the first textbook on computational linguistics in 1967, and he was involved in establishing the Association of Computational Linguistics (ACL) and served as its second president.

Due in part to Hays' efforts, dependency syntax has been a known entity in the field of NLP since the field's inception. Over the decades since Hays' most direct contributions, dependency has been employed and developed further by various computational linguists seeking a basis for the automated parsing of texts. Worth mentioning in this are the efforts of Schubert (1987) and Maxwell and Schubert (1989) and their collaborators to use DG as the basis for automated translation of texts across various languages (Bangla/Bengali, Danish, English, Esperanto, Finnish, French, German, Hungarian, Japanese, Polish). The discussion of dependency that Schubert (1987) provides is insightful and accessible; it serves well as an introductory discussion of the nature of dependency syntax.

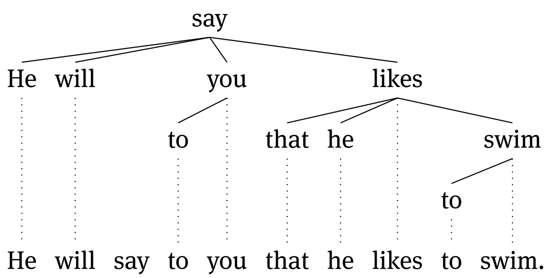
At present (2018), dependency syntax enjoys a prominent position in the field of NLP; it is intimately connected to the current state of the art in automated text parsing (cf. Maruyama, 1990; Eisner, 1996; Menzel and Schröder, 1998; McDonald, 2006; Kudo and Matsumoto, 2002; Yamada and Matsumoto, 2003; Nivre, 2008). Worth mentioning in particular are the efforts of the many dozens of computational linguists who are producing treebank corpora according to the same one DG annotation scheme. The project is known as *Universal Dependencies* (UD), and to date the greater UD project has produced treebank corpora of more than 60 languages (http://universaldependencies.org/).[6] The intent is to create openly accessible treebanks that are available for various uses, among these the typological study of natural language syntax. While

6 The UD initiative is a work in progress. All information concerning the initiative and its current state can be found at the web address given, including the relevant literature.

the UD project has helped generate much interest in dependency syntax in general, it is controversial because of its decision to subordinate most function words to the content words with which they cooccur (cf. Groß and Osborne, 2015 and Osborne and Maxwell, 2015). Most DGs over the decades have chosen to do the opposite; they subordinate most content words to the function words with which they cooccur.

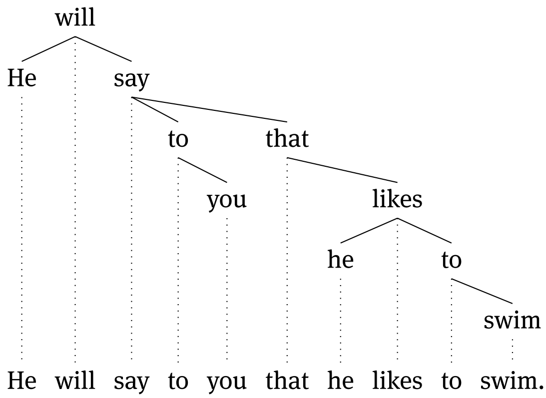
This aspect of UD project is worth considering more closely, since it has given rise recently to debate in DG circles about the best hierarchical analysis of simple sentences. The next two trees illustrate the competing analyses:

(22) a.



– UD analysis

b.



– Traditional analysis

UD's decision to subordinate function words to content words is shown in (22a): the modal auxiliary *will* is subordinated to the content verb *say*; the preposition *to* is subordinated to the referential pronoun *you*; the subordinator *that* is subordinated to the content verb *likes*; and the particle *to* introducing the infinitive *swim* is subordinated to *swim*. In contrast, the traditional analysis given as (22b) makes the opposite choices, subordinating the content words to the function words.

With the exception of tree (22a), the dependency hierarchies produced in this contribution adhere to the traditional assumptions illustrated with tree (22b), that is, most function words are taken as heads over the content words with which they cooccur. The reason the alternative analysis given as (22a) is mentioned here is due to the prominence of the UD project in general and further to the desire to demonstrate that DG is not a unified approach to syntax; it is, rather, a family of grammars that vary in significant ways.

5 Data

The data that DGs are interested in depend primarily on which of the goals just discussed is pursued. Dependency syntax actually has an advantage when it comes to locating and exploring specific phenomena of syntax. This advantage is due to the fact that most of the treebanks produced in the field of NLP are now dependency-based. Many of these treebanks are freely available and searchable (e. g. the Prague Dependency Treebank: https://ufal.mff.cuni.cz/pdt3.0, and the dozens of treebanks from the UD project mentioned in Section 4.2: http://universaldependencies.org/).

When the goal is to identify the quantitative laws governing natural language syntax (e. g. Liu, 2009, 2010), large amounts of data are of course needed. The treebank corpora being produced by the computational linguists are a necessary and tremendously valuable resource in this respect. For instance, if one is interested in determining the mean dependency distance – the average distance separating dependents from their heads – across distinct languages or languages families, the existence of the relevant treebanks makes this possible. Or if one is interested in determining the extent to which the syntactic structures of a given language are head-initial or head-final, the treebanks for that language make it possible.

When the goal is to develop a theoretically stringent and satisfying framework for the study of natural language syntax, introspection is the primary means linguists use. This is true for DG and PSG linguists alike. The dependency-based frameworks of the German schools (Baum, Tarvainen, Lobin, Engel, Jung, Heringer, Bröker, Groß, Eroms), Meaning-Text Theory, Word Grammar, and Lexicase mentioned in Section 4.1 above were all established before the mentioned treebanks became widely available for research purposes. Their authors hence had to rely on their own introspection to a large extent. Despite the availability of treebanks, introspection remains an important means of producing and analyzing data. Treebanks help reveal the extent to which syntactic structures and syntactic phenomena actually occur, but they do not help much when the goal is to gain insight into why certain sentence structures do not occur.

6 Evaluation

To evaluate the merits of doing syntax in terms of dependencies, one must consider and scrutinize the arguments produced by the DG community in favor of dependency. One major argument in favor of dependency is simplicity. Basic dependency analyses are so simple that children can learn them. This aspect of DG is established next by considering some of the history of the use of dependency for pedagogical goals. A second advantage that DG has is accuracy. Dependency is more efficient and accurate at modeling constituent structure.

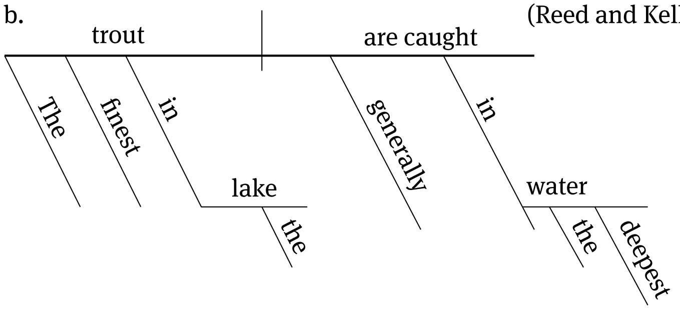
6.1 Simplicity

Many dependency grammarians see the simplicity of basic dependency analyses as an advantage that dependency has over phrase structure (see e. g. Engel, 1994, 23, 26; Hudson, 2007, 117). This simplicity is evident in the fact that dependency is dominant in the sentence diagrams that have been used to teach sentence grammar in schools over the past 150 years. Three examples of these diagrams are given in the following paragraphs.

The most well-known sentence diagrams used to teach sentence grammar in schools stem from the late 19th century; they are those of Alonzo Reed and Brainerd Kellogg. The Reed-Kellogg system of sentence diagramming primarily uses dependency to group words together, although a significant measure of phrase structure is also present in the Reed-Kellogg diagrams. An example of such a diagram is next:

(23) a. The finest trout in the lake are generally caught in the deepest water.

b. (Reed and Kellogg, 1909, 69)[7]



The manner in which *trout* and *are caught* are placed on the baseline and separated by the vertical divider is a manifestation of phrase structure, and so is the equi-level positioning of *are* and *caught*. The rest of the diagram, however, is in terms of dependency; the manner in which the individual words are connected to but dangle below the word(s) that they modify is dependency. The system of sentence diagramming developed by Reed and Kellogg has been taught in American schools since the late 1800s. The fact that Reed and Kellogg's understanding of sentence structure reached and continues to reach such prominence and that it uses dependency heavily to organize words into meaningful units of sentence structure suggests that dependency syntax is simple enough and appropriate for children to learn.

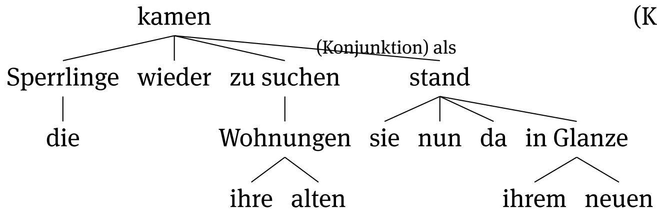
7 The example is taken from the 1909 edition of Reed and Kellogg's famous work *Graded Lessons in English*. The work first appeared in 1889. The publisher (Charles E. Merril Co.) writes at the start of the 1909 edition: "The orders for introduction received and filled during the year 1908 exceeded by more than 100,000 copies the introduction orders received during the preceding year." These numbers give an idea of just how widespread and influential the Reed-Kellogg sentence diagrams had become by the early 1900s.

Other systems of diagramming sentences were also appearing at approximately the same time, in the second half of the 19th century. The next example is from Franz Kern's work *Zur Methodik des deutschen Unterrichts* (1883). Kern was a pedagogue who was interested in reforming the manner in which sentence grammar was taught in Prussian schools. He advocated using sentence diagrams like the next one in the classroom:[8]

(24) a. Als sie nun in ihrem neuen Glanze dastand, kamen die Sperlinge
as it now in its new glory there.stood came the sparrows
wieder, ihre alten Wohnungen zu suchen.
again their old homes to seek

‘As it now stood there in its new glory, the sparrows came again to seek their old homes.’

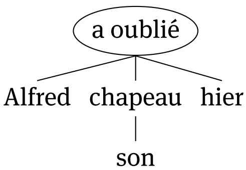
b. (Kern, 1883, 17)



This diagram is almost entirely based on dependency, a fact that is most evident in the status of the finite verb *kamen* as the root of the sentence – the binary subject-predicate division encountered in Reed and Kellogg's example above is hence absent from this analysis. It is possible, however, to interpret the diagram as also including a small measure of phrase structure, since Kern positioned preposition and noun together in a single node (*in Glanze*). Note as well that the subordinator *als* ‘as’ receives a unique status, since it appears as a label on a dependency edge connecting *stand* to *kamen*.

Tesnière was also interested in reforming the manner in which sentence grammar was taught in schools, but in French schools, since he was a Frenchman. The next example is a so-called *stemma*, the term Tesnière used to denote his tree-like diagrams. The stemma shows the hierarchy of words for the French sentence *Hier Alfred a oublié son chapeau* ‘Yesterday, Alfred forgot his hat’:

(25)



8 The sentence is from Lessing's fabel *The sparrows*; it is the second sentence in the fabel.

The bubble around *a oublié* ‘has forgotten’ marks what Tesnière called a *dissociated nucleus*. The two words *a* and *oublié* are together viewed as forming a single semantic unit, i. e. a nucleus. Tesnière viewed such diagrams as simple enough to be used as tools in the classroom. In Chapters 276–277 of his main oeuvre *Éléments de syntaxe structurale* (1959/2015), he documents a concerted effort to institute his stemmas and system of sentence diagramming in French schools.

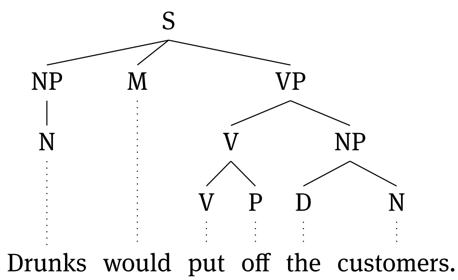
The example sentence diagrams just produced (from Reed and Kellogg, Kern, and Tesnière) were all deemed appropriate by their creators for illustrating sentence structure to children in schools. Such diagrams are, though, primarily dependency-based. In contrast, I am not aware of any sentence diagrams based more consistently on phrase structure ever being employed in the classroom to teach sentence grammar to children. The complexity of phrase structures oversteps what is practicle for teaching to children.

6.2 Accuracy

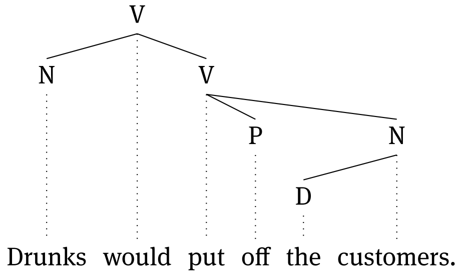
Proponents of phrase structure syntax can object at this point that in the absence of descriptive and explanatory accuracy, simplicity alone is of no value. The response that DG can produce is that dependency structures are in fact more accurate than phrase structures in a key respect, namely regarding basic tests for constituents. Most tests for constituents verify the existence of phrasal constituents only; they provide little support for the existence of subphrasal strings as constituents. This situation supports dependency over phrase structure because the subphrasal constituents of PSGs are not constituents in DGs to begin with.

Due to space limitations, this point can be sketched here only briefly. I have, however, presented and developed the message extensively in a number of places (Osborne, 2005, 254–8, 2006, 53–8, 2008, 1126–32, 2015, and especially, 2018). The point of contention is illustrated next using an example taken from Radford (1988), which is discussed at length in Osborne (2018, 14–18). Radford produces and motivates certain parts of the sentence structure given as (26a) using tests for constituents (e. g. movement, proform substitution, answer fragments). The corresponding dependency analysis of the same sentence is added as (26b):

(26) a. (Radford, 1988, 91)



b.



Most standard tests for constituents verify aspects of these two trees. For instance, they verify that the strings *Drunks*, *the customers*, and *put off the customers* are complete subtrees (= constituents),[9] e. g.

Topicalization

(27) a. (Inapplicable to *Drunks*)

b. ...and **the customers**, drunks would put off.

c. ...and **put off the customers**, drunks (certainly) would.

Proform substitution

(28) a. **They** would put off the customers. (*They* = *Drunks*)

b. Drunks would put **them** off. (*them* = *the customers*)

c. Drunks would **do so**. (*do so* = *put off the customers*)

Answer fragments

(29) a. Who would put off the customers? – **Drunks**.

b. Who would drunks put off? – **The customers**.

c. What would drunks do? – **Put off the customers**.

Topicalization is inapplicable to *Drunks* because *Drunks* is already in topic position at the start of the sentence. Based on these data, the two trees (26a) and (26b) are accurate because they both view these strings as complete subtrees, that is, as constituents.

Crucially, however, the phrase structure tree (26a) and the dependency tree (26b) differ concerning many of the other strings. They disagree with respect to the status of *would*, *put*, *customers*, and *put off*. The phrase structure tree views these strings as

9 The two terms *complete subtree* and *constituent* are synonymous in the current context. Both are defined over tree structures as *a node plus all the nodes that that node dominates*.

constituents, whereas the dependency tree does not.[10] The three tests for constituents support dependency in this regard:

Topicalization

(30) a. *…and **would** the customers put off the customers? (Unacceptable as a declarative statement)
b. *…and **put** drunks would off the customers.
c. *…and **customers** drunks would put off the.
d. *…and **put off** drunks would the customers.

Proform substitution

(31) a. *Drunks **do (so)** put off the customers. (*do* (*so*) = *would*)
b. *Drunks would **do (so)** the customers. (*do* (*so*) = *put*)
c. *Drunks would put off the **them**. (*them* = *customers*)
d. *Drunks would **do (so)** the customers. (*do* (*so*) = *put off*)

Answer fragments

(32) a. What about drunks putting off the customers? – ***Would**.
b. What would drunks do concerning the customers? – ***Put**.
c. ??The **who** would drunks put off? – **Customers**.
d. What would drunks do to the customers? – ***Put off**.

It should be apparent that the dependency tree (26b) is supported by these data, because in each of the four cases, the test string is not identified as a complete subtree. In other words, the tests are sensitive to the complete subtrees of dependency rather than to the complete subtrees of phrase structure.

One might object that two of the complete subtrees in (26a) and (26b) have been omitted from the discussion (so far), namely *the* and *off*. Identifying determiners and particles of phrasal verbs as constituents is difficult to do using tests for constituents due to the idiosyncratic traits of these elements. However, the ability to shift the particle *off* as illustrated in example (28) (*Drunks would put them off*) and the ability to omit the definite article *the* (*Drunks would put off customers*) support their status as constituents. More importantly, the two analyses agree about their status, since both (26a) and (26b) show *the* and *off* as complete subtrees. Hence there is no disagreement regarding these two strings.

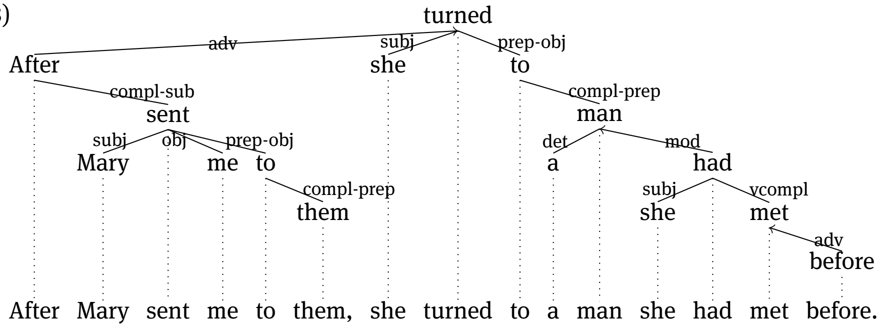
To summarize, most tests for constituents of the sort that are widely employed in linguistics, grammar, and syntax books actually support dependency over phrase structure, because they identify far less syntactic structure than phrase structure assumes.

10 They are not constituents in the dependency tree because they dominate other words, e. g. *customers* dominates *the*.

7 Sample analysis

The next diagram illustrates in one spot many of the aspects of DG discussed above:

(33)



This sentence is complex, containing three clauses, a matrix clause and two subordinate clauses. The key traits of dependency syntax are visible in the tree: verb centrality, one-to-one mapping (words to nodes), strict headedness (entirely endocentric). The diagram is a tree, which means that every node (except the root node) in the tree has one and only one parent node. It encodes and shows actual word order, and in this respect, it is unlike the tree diagrams produced by early DGs, which preferred to abstract away from actual word order.

The diagram includes the syntactic relations. Each and every dependency receives a label identifying the syntactic function of the entire subtree below it. The labels are abbreviated as follows:

Syntactic relations in (33)
adv = adverbial
compl-prep = complement of preposition
compl-sub = complement of subordinator
det = determiner
mod = modificative attribute
obj = object
prep-obj = prepositional object
subj = subject
vcompl = verbal complement

These particular functions have been chosen by me and are intended merely to be representative of the tendency of DGs to produce inventories of such functions. From the point of view of corpus linguistics, the presence of the functions would make the corpus containing such tree structures easily searchable for specific phenomena of syntax.

Tree (33) also contains the arrow edges that mark adjuncts. The adverbial clause *After Mary sent me to them*, the reduced relative clause *she had met before*, and the adverb *before* are thus identified as adjuncts. Their status as adjuncts is evident because they do not correspond to any of the actants in the valency frames of the main content verbs:

Valency frames of content verbs in (33)
MEET [N, N]
SEND [N, N, →]
TURN [N, P_{to}]

The arrow in the valency frame of *SEND* indicates an actant that gives the destination of movement; its form is somewhat flexible, e. g. *send it to you*, *send him out of the house*, *send them home*, *send it up*, *send it over*, etc. The preposition *to* is included in the valency frame of *TURN* because *turn to* is an idiosyncratic combination. Finally, note that the second N actant in the valency frame for *MEET* is not manifest in the tree due to the ability to omit non-subject relative pronouns from relative clauses in English.

8 Concluding comment

This contribution has attempted to provide some basic information about the nature of dependency grammar (DG), some of its history, some of the key properties that distinguish dependency from phrase structure, some of the tools DG uses to analyze syntactic structures, some of the goals it pursues and the data it uses and produces, and some of the advantages it has over phrase structure. Inevitably, the one or other dependency grammarian will feel shortchanged by the limited coverage given above, and further, some dependency grammarians likely disagree with key points in the account above. It must be emphasized in this regard that sentence structures in terms of dependencies is arguably the tradition that reaches back furthest in the history of syntactic theory, and that there has been and certainly still is great variation among DGs. Doing justice to the multifaceted approaches that view themselves as instantiations of DG is a difficult task indeed.

References

Ágel, Vilmos, Ludwig Eichinger, Hans-Werner Eroms, Peter Hellwig, Hans Jürgen Heringer & Henning Lobin (eds.). 2003. *Dependency and valency: An international handbook of contemporary research, vol. 1*, Berlin: Walter de Gruyter.

Ágel, Vilmos, Ludwig Eichinger, Hans-Werner Eroms, Peter Hellwig, Hans Jürgen Heringer & Henning Lobin (eds.). 2006. *Dependency and valency: An international handbook of contemporary research*. Vol. 25.2. Berlin: Walter de Gruyter.

Baum, Richard. 1976. *Dependenzgrammatik: Tesnière's Modell der Sprachbeschreibung in wissenschaftsgeschichtlicher und kritischer Sicht* (Beihefte zur Zeischrift für Romanische Philologie 151). Tübingen: Max Niemeyer Verlag.

Baumgärtner, Klaus. 1970. Konstituenz und Dependenz. In Hugo Steger (ed.), *Vorschläge für eine strukturale Grammatik des Deutschen*, 52–77. Darmstadt: Wissenschaftliche Buchgesellschaft.

Bender, Byron. 2002. In memoriam, Stanley Starosta 1939–2002. *Oceanic Linguistics* 41(2). 255–274.

Bloomfield, Leonard. 1933. *Language*. New York: Holt, Rinehart and Winston.

Bröker, Norbert. 1999. *Eine Dependenzgrammatik zur Kopplung heterogener Wissensquellen*. Tübingen: Niemeyer.

Carnie, Andrew. 2010. *Constituent structure*, 2nd edn. Oxford, UK: Oxford University Press.

Chomsky, Noam. 1957. *Syntactic structures*. The Hague: Mouton Publishers.

Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge: MIT Press.

Covington, Michael. 1984. *Syntactic theory in the High Middle Ages*. London: Cambridge University Press.

Eisner, Jason. 1996. An empirical comparison of probability models for dependency grammar. Technical Report IRCS-96-11, Institute for Research in Cognitive Science, University of Pennsylvania.

Engel, Ulrich. 1994. *Syntax der deutschen Gegenwartssprache*. 3rd ed. Berlin: Erich Schmidt Verlag.

Eroms, Hans-Werner. 2000. *Syntax der deutschen Sprache*. Berlin: Walther de Gruyter.

Groß, Thomas. 1999. *Theoretical foundations of dependency syntax*. München: Iudicium.

Groß, Thomas & Timothy Osborne. 2013. Katena und Konstruktion: Ein Vorschlag zu einer dependenziellen Konstruktionsgrammatik. *Zeitschrift für Sprachwissenschaft* 32(1). 41–73.

Groß, Thomas & Timothy Osborne. 2015. The dependency status of function words. *Proceedings of the Third International Conference on Dependency Linguistics*, August 27–30, 2013, Prague, Czech Republic, 111–120.

Hays, David. 1964. Dependency theory: A formalism and some observations. *Language* 40. 511–525.

Hays, David. 1967. *Introduction to computational linguistics*. London: Macdonald & Co. (publishers) ltd.

Heringer, Hans. 1996. *Deutsche Syntax: Dependentiell*. Tübingen: Stauffenburg.

Hudson, Richard. 1976. *Arguments for a non-transformational grammar*. Chicago: The University of Chicago Press.

Hudson, Richard. 1980a. Constituency and dependency. *Linguistics* 18(3/4). 179–198.

Hudson, Richard. 1980b. A second attack on constituency: A reply to Dahl. *Linguistics* 18. 489–504.

Hudson, Richard. 1984. *Word Grammar*. Oxford, UK: Basil Blackwell.

Hudson, Richard. 1987. Zwicky on heads. *Journal of Linguistics* 23(1). 109–132.

Hudson, Richard. 1990. *An English Word Grammar*. Oxford: Basil Blackwell.

Hudson, Richard. 2003. Word Grammar. In Vilmos Ágel, Ludwig Eichinger, Hans-Werner Eroms, Peter Hellwig, Hans Jürgen Heringer & Henning Lobin (eds.), *Dependency and valency: An international handbook of contemporary research*. Vol. 25.1, 508–526. Berlin: Walter de Gruyter.

Hudson, Richard. 2007. *Language networks: The new Word Grammar*. Oxford: Oxford University Press.

Hudson, Richard. 2010. *An introduction to Word Grammar*. Cambridge, UK: Cambridge University Press.

Järventausta, Marja. 2003. Das Subjektproblem in der Valenzforschung. In Vilmos Ágel, Ludwig Eichinger, Hans-Werner Eroms, Peter Hellwig, Hans Jürgen Heringer & Henning Lobin (eds.), *Dependency and valency: An international handbook of contemporary research*. Vol. 25.1, 781–794. Berlin: Walter de Gruyter.

Jung, Wha-Young. 1995. *Syntaktische Relationen im Rahmen der Dependenzgrammatik*. Hamburg: Helmut Buske Verlag.

Kahane, Sylvain. 1996. If HPSG were a dependency grammar.... *Actes de TALN*, Marseille, 22–24, May 1996, 45–49.

Kern, Franz. 1883. *Zur Methodik des deutschen Unterrichts*. Berlin: Nicolaische Verlags-Buchhandlung.

Korhonen, Jarmo. 1977. *Studien zu Dependenz, Valenz, und Satzmodell. Teil 1*. Bern: Peter Lang.

Kübler, Sandra, Ryan McDonald & Joakim Nivre. 2009. *Dependency parsing*. San Rafael, CA: Morgan & Claypool Publishers.

Kudo, Taku & Yuji Matsumoto. 2002. Japanese dependency analysis using cascaded chunking. *Proceedings of the 6th Workshop on Computational Language Learning (CoNLL)*, Taipei, Taiwan, 63–69.

Kunze, Jürgen. 1975. *Abhängigkeitsgrammatik* (Studia grammatika XII). Berlin: Akademie-Verlag.

Liu, Haitao. 2009. Probability distribution of dependencies based on Chinese Dependency Treebank. *Journal of Quantitative Linguistics* 16(3). 256–273.

Liu, Haitao. 2010. Dependency direction as a means of word-order typology: A method based on dependency treebanks. *Lingua* 120(6). 1567–1578.

Lobin, Henning. 1993. *Koordinationssyntax als prozedurales Phänomen* (Studien zur deutschen Sprache 46). Tübingen: Gunter Narr Verlag.

Maruyama, Hiroshi. 1990. Structural disambiguation with constraint propagation. *Proceedings of the 28th Annual Meeting of the Association for Computational Linguistics (ACL)*, Pittsburgh, PA, 31–38.

McDonald, Ryan. 2006. Discriminative Learning and Spanning Tree Algorithms for Dependency Parsing. Ph.D. thesis, University of Pennsylvania.

Matthews, Peter. 1981. *Syntax*. Cambridge, UK: Cambridge University Press.

Maxwell, Dan & Klaus Schubert (eds.). 1989. *Metataxis in Practice: Dependency Syntax for Multilingual Machine Translation*. Dordrecth: Foris Publications.

Mel'čuk, Igor. 1979. *Studies in dependency syntax*. Ed. by Paul T. Roberge. Ann Arbor: Karoma Publishers.

Mel'čuk, Igor. 1988. *Dependency syntax: Theory and practice*. Albany: State University of New York Press.

Mel'čuk, Igor. 2009. Dependency in natural language. In Alain Polguère & Igor Mel'čuk (eds.), *Dependency in linguistic description*, 1–110. Amsterdam: John Benjamins.

Mel'čuk, Igor & Nikolai Pertsov. 1987. *Surface syntax of English: A formal model with the Meaning-Text Framework*. Amsterdam: John Benjamins.

Menzel, Wolfgang & Ingo Schröder. 1998. Decision procedures for dependency parsing using graded constraints. In *Proceedings of the Workshop on Processing of Dependency-Based Grammars (ACL-COLING)*, Montreal, Canada, 78–87.

Nivre, Joakim. 2008. Algorithms for deterministic incremental dependency parsing. *Computational Linguistics* 34(4). 513–553.

Osborne, Timothy. 2005. Beyond the constituent: A DG analysis of chains. *Folia Linguistica* 39(3–4). 251–297.

Osborne, Timothy. 2006. Shared material and grammar: A dependency grammar theory of non-gapping coordination. *Zeitschrift für Sprachwissenschaft* 25. 39–93.

Osborne, Timothy. 2008. Major constituents and two dependency grammar constraints on sharing in coordination. *Linguistics* 46(6). 1109–1165.
Osborne, Timothy. 2012. Edge features, catenae, and dependency-based Minimalism. *Linguistic Analysis* 34(3–4). 321–366.
Osborne, Timothy. 2015. Tests for constituents: Dependency, constituency, and the status of function words. In *Proceedings of the Third International Conference on Dependency Linguistics*, Aug. 24–26, Uppsala, Sweden, 251–260.
Osborne, Timothy. 2018. Tests for constituents: What they really reveal about the nature of syntactic structure. *Language under Discussion* 5. 1–41.
Osborne, Timothy & Thomas Groß. 2012. Constructions are catenae: Construction Grammar meets Dependency Grammar. *Cognitive Linguistics* 23(1). 163–214.
Osborne, Timothy & Daniel Maxwell. 2015. A historical overview of the status of function words in dependency grammar. In *Proceedings of the Third International Conference on Dependency Linguistics*, Aug. 24–26, Uppsala, Sweden, 241–250.
Osborne, Timothy & Thomas Groß. 2016. The *do-so*-diagnostic: Against finite VPs and for flat non-finite VPs. *Folia Linguistica* 50(1). 97–135.
Osborne, Timothy, Michael Putnam & Thomas Groß. 2011. Bare phrase structure, label-less structures, and specifier-less syntax: Is Minimalism becoming a dependency grammar? *The Linguistic Review* 28. 315–364.
Osborne, Timothy, Michael Putnam & Thomas Groß. 2012. Catenae: Introducing a novel unit of syntactic analysis. *Syntax* 15(4). 354–396.
Percival, Keith. 1990. Reflections on the history of dependency notions in linguistics. *Historiographia Linguistica* 17. 29–47.
Radford, Andrew. 1988. *Transformational grammar: A first course*. Cambridge, UK: Cambridge University Press.
Reed, Alonzo & Brainerd Kellogg. 1909. *Graded lessons in English: An elementary English grammar*. New York: Charles E Merrill Co.
Schubert, Klaus. 1987. *Metataxis: Contrastive dependency syntax for machine translation*. Dordrecht: Foris Publications.
Starosta, Stanley. 1988. *The case for Lexicase: An outline of Lexicase grammatical theory*. London: Pinter Publishers.
Starosta, Stanley. 2003. *Lexicase grammar*. In Vilmos Ágel, Ludwig Eichinger, Hans-Werner Eroms, Peter Hellwig, Hans Jürgen Heringer & Henning Lobin (eds.), *Dependency and valency: An international handbook of contemporary research*. Vol. 25.1, 526–545. Berlin: Walter de Gruyter.
Tarvainen, Kalevi. 1981. *Einführung in die Dependenzgrammatik*. Tübingen: Max Niemeyer Verlag.
Tesnière, Lucien. 1959. *Éléments de syntaxe structurale*. Paris: Klincksieck.
Tesnière, Lucien. 2015. *Elements of structural syntax*. Translation of Tesnière 1966 (the 2nd edition of Tesnière 1959) by Timothy Osborne and Sylvain Kahane. Amsterdam: John Benjamins.
Uzonyi, Pál. 2003. Dependenzstruktur und Konstituenzstruktur. In Vilmos Ágel, Ludwig Eichinger, Hans-Werner Eroms, Peter Hellwig, Hans Jürgen Heringer & Henning Lobin (eds.), *Dependency and valency: An international handbook of contemporary research*. Vol. 25.1, 230–247. Berlin: Walter de Gruyter.
Versteegh, Kees. 1997. *Landmarks in linguistic thought III: The Arabic linguistic tradition*. London: Routledge.
Yamada, Hiroyasu & Yuji Matsumoto. 2003. Statistical dependency analysis with support vector machines. In *Proceedings of the 8th International Workshop on Parsing Technologies (IWPT)*, Nancy, France, 195–196.

Mark Steedman

14 Combinatory Categorial Grammar

Abstract: Combinatory Categorial Grammar (CCG) is a radically lexicalized theory of grammar in which all language-specific information, including the linear order of heads, arguments, and adjuncts, is specified in the lexicon, from which it is projected onto sentences by language-independent universal type-dependent combinatory rules of low "slightly non-context-free" expressive power, applying to strictly adjacent phonologically-realised categories. Syntactic and phonological derivation are isomorphic, and are synchronously coupled with semantic composition in a purely type-dependent rule-to-rule relation.

1 Overview

1.1 Goals

The central problem addressed by Combinatory Categorial Grammar (CCG) is the nature of the mapping between sound and meaning. The goal is to achieve an explanatory theory of natural language grammar that is immediately applicable to psychological and computational models of syntactic and semantic processing of spoken and written natural language, and of language acquisition by children.

1.2 Data

The data which are drawn upon in order to *define* CCG are facts generally agreed among linguists concerning long-range dependency, coordination, and prosodic structure, all of which give the appearance of displacement, or non-contiguity of elements that belong together semantically, such as governors (heads) and their complements.

The data to which the theory has been *applied* is much more various, and includes corpus data, both labeled and unlabeled, that is used to train parsers, and the various test-sets that are used to evaluate them, including corpora of child-directed utterance, and psycholinguistic data.

Acknowledgement: Thanks to the editors and reviewers of this volume for their helpful comments. Some of these ideas were developed through a class "Introduction to Combinatory Categorial Grammar" which was presented by the author at the 2017 LSA Summer Institute in Lexington KY. Thanks also to the participants for their input. The writing was supported by ERC Advanced Fellowship 742137 SEMANTAX.

https://doi.org/10.1515/9783110540253-014

1.3 Tools

Crosslinguistic similarities and differences are represented in CCG solely at the level of the lexicon, which specifies all language-specific properties including the linear order and semantic dependency of governors and dependent or complement constituents. The lexicon is projected onto the sentences of the language by "combinatory" rules—that is, by strictly string-adjacent operations, combining contiguous categories without the involvement of any form of "action at a distance", such as movement, copying, or deletion under identity.

The only representational levels in CCG are phonological and logical form. Syntactic derivation is not itself a level of representation, and is dispensible. All syntactic rules are type-dependent, rather than structure-dependent, and assemble logical and phonological form in lockstep with syntactic derivation. The hypothesis is that the degrees of freedom in the type-system of the lexicon and the combinatory rules are both necessary and sufficient for the analysis of the languages of the world. The categories are those of categorial grammar. The relations between categories are combinatory in the sense defined above, and are fully formalized.

The theory outlined in this chapter, and developed in slightly different forms and at greater length in earlier publications, has been applied to the syntactic and semantic analysis of coordination and unbounded dependency in a wide range of languages. It has also been widely applied computationally in practical natural language processing (NLP) applications, particularly those requiring that the syntax support semantic interpretation. There is a CCG-based computational account of acquisition and development, based on semantic bootstrapping of the language-specific lexicon (Abend et al., 2017). There is a hypothesis concerning the origins of the categories and combinatory rules in terms of their use for planning complex actions in human and prehuman cognition (Steedman, 2002, 2017). Neither is discussed at any length here. Wide coverage parsers for CCG have been developed.[1]

1.4 Sample analysis

The following sentence, selected by the editors for comparison across the various approaches in this volume, is quite long:

(1) After Mary introduced herself to the audience, she turned to a man that she had met before.

1 The interested reader can try out the "Easy CCG" parser (Lewis et al., 2016) by typing or pasting sentences such as (1) into the input box at http://4.easy-ccg.appspot.com/, bearing in mind that this is a probabilistic parser, with a lexicon and parsing model primarily trained on the Penn WSJ treebank.

Accordingly, its derivation is presented in Figure 1 in three steps, with all discussion of semantics and logical form deferred until the detailed discussion of the constructions involved.

First, the preposed adjunct *After Mary introduced herself to the audience* is derived syntactically as in Figure 1a. CCG derivations like this are written in the acceptance direction, with the words at the top and the "start symbol" (usually, *S*) at the bottom, but are otherwise equivalent to standard derivational phrase-structure trees. Slashes / and \ define the English transitive verb as looking for its first NP (object) argument to the right and its second (subject) argument to the left. Underlines indicate combination, and the directional arrows > and < indicate that the rule involved is forward (rightward) or backward (leftward) application. The ↑ notation indicates that the category in question such as $NP^{\uparrow}$ has a type-raised or cased category such as the *nominative* category $S/(S\backslash NP)$, abbreviated here for readability. Since the derivation shown is entirely applicative, type-raising has no effect here other than to reverse the directionality of the rule that combines verb and argument, so can be temporarily ignored. The binding of the reflexive anaphor "herself" is also lexicalized, via the logical form (not shown), whose details are discussed in Section 3.5.

The main clause involves an unbounded relativized dependency, and is more complicated syntactically, making crucial use of composition and type-raising, as in Figure 1b.

This derivation crucially involves composition rules, indexed $>\mathbf{B}$ and $>\mathbf{B}_{\times}$. Their operation, whose details are discussed in Section 4, crucially depends on the arguments being type-raised. In particular, the subject "she" of the relative clause must bear the *nominative* raised category for the derivation to go through, although for the purposes of this overview, we continue to abbreviate it as $NP^{\uparrow}$. (Thus, English is highly ambiguous as to case, unlike morphologically cased languages like Latin and Japanese.)

To complete the derivation, the sentential adjunct derived in 1a combines with the sentence derived in 1b by simple forward application, yielding a sentence, as in Figure 1c.

Although the assembly of logical form is not shown in this introductory analysis, its derivation is entirely compositional and homomorphic to the surface syntactic derivations shown. In particular, the logical form corresponding to the complex noun "man that she had met before" is under the analysis of relativization developed in Section 4 itself a property of type N with the appropriate logical form $\lambda n \lambda x.past\,(perfect\,(meet\,x\,pro_{she})) \wedge n\,x$ (cf. (25)). Such details are discussed at length in the body of the chapter.

Anaphoric relations, including the binding of the pronoun "she" in the main clause to "Mary" in the adjunct, and the binding of the past tense of the main clause "turned" in an "after' relation to the antecedent reference time of the adjunct "introduced", among others, are not treated in CCG as falling in the domain of sentence grammar proper.

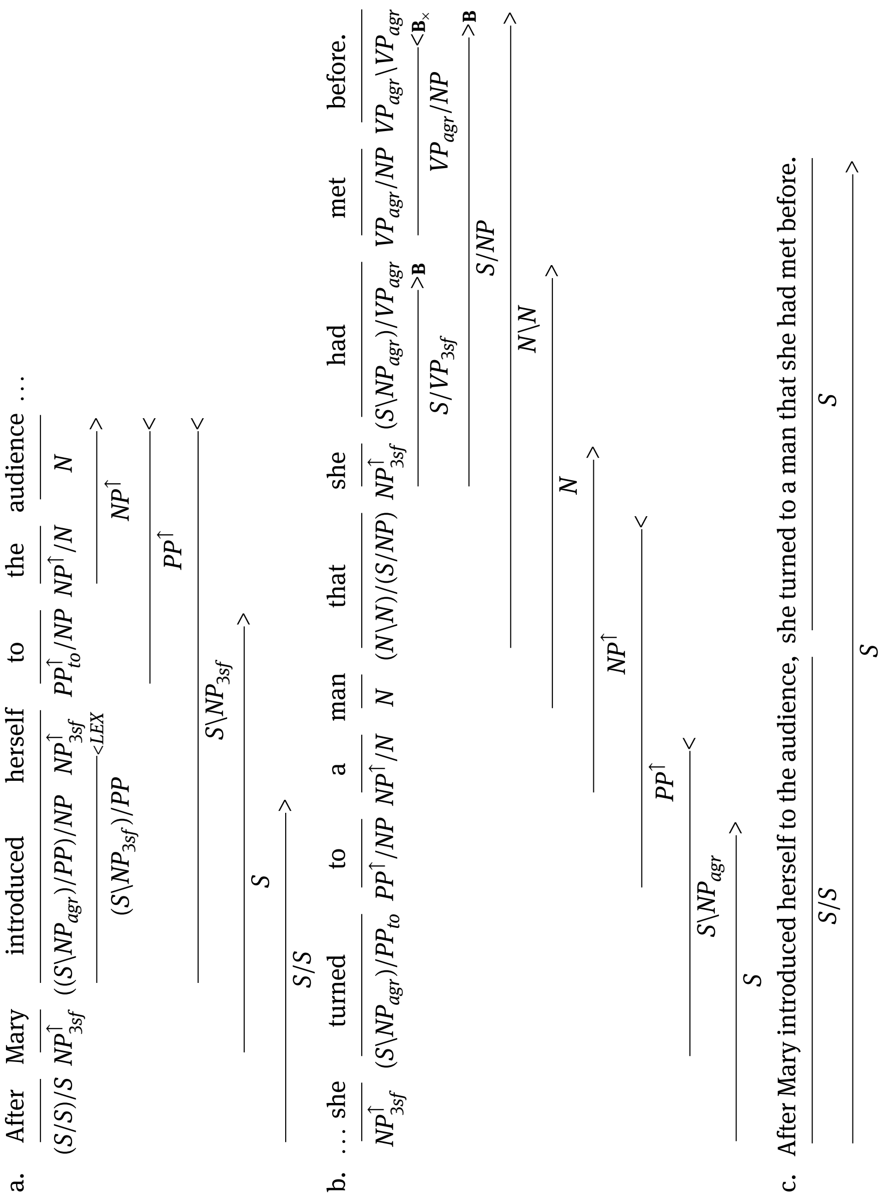


Figure 1: Sample analysis.

1.5 Evaluation criteria

The evaluation criteria for comparing CCG with alternative approaches are descriptive and explanatory adequacy, and applicability to practical computational natural language processing, including the building of logical form.

Descriptive adequacy is attained by capturing all the phenomena of a system. Explanatory adequacy is attained by capturing *only* those phenomena, and being unable to capture other comparable phenomena that are not exhibited by the system. A theory which can express phenomena that we believe can never occur is overly-expressive and less than explanatory (although such theories may be extremely useful in laying out the phenomena in ways that help us to find our way to more explanatory ones).[2]

Since in the case of the grammatical system we only have a fairly small sample of languages to work with, we don't have complete knowledge of the set of possible phenomena. It follows that any claim to explanatory adequacy in the theory of grammar is a hostage to fortune, and can be disproved by the discovery of languages that controvert its prediction of their non-existence.

Nevertheless, the available descriptive accounts make grammar seem relatively systematic. For example, we shall see in Section 7 that CCG predicts that two of the 24 possible ways of linearising the four elements corresponding to the English words comprising the noun-phrase "these five fat cats" are impossible, and will never be found in any natural language (cf. Greenberg, 1963; Cinque, 2005). A theory that is descriptively adequate in other respects, and also accurately predicts the same generalization concerning word orders in other constructions is empirically falsifiable, and therefore more explanatorily adequate than one that does not.[3]

2 Historical background to CCG

When syntactic theory as defined in Chomsky, 1965 (hereafter, *Aspects*) fragmented in the '70s and '80s, leading to the profusion of approaches assembled in the present volume, attempts to develop alternatives to *Aspects*-style transformational rules took two forms.

2 While initially acknowledging something like the above as a definition of explanatory adequacy, Chomsky (1965) proposes the provision of a theory of child language acquisition as a proxy for explanatory adequacy. However, once we have admitted that some amount of innate knowledge must be available for language acquisition to be possible at all, then *every* theory of language has a theory of acquisition if we assume that its key assumptions and constraints are by some evolutionary process innate. In order to avoid merely pushing the burden of explanation off onto the theory of evolution, an explanatory theory in the first sense is actually a prerequisite for an explanatory theory of child language acquisition.

3 For reasons that have nothing to do with the theory of grammar, all such variation is Zipfian in distribution, with the long tail of less common orders becoming double-exponentially rarer, so that it is in practice hard to know whether an unseen word order is truly impossible, or just so rare that we haven't seen it yet.

One group of constraint-based theories, usually expressed in unification-based formalisms, were as overly expressive as *Aspects* transformational grammars, but were easier to implement computationally (and therefore to automatically check for over- and under-generalization). They included Augmented Transition Network Grammar (ATNG. Woods, 1970), Functional Unification Grammar (FUG, Kay, 1984), Lexical Functional Grammar (LFG, Bresnan, 1982), Dependency/Word Grammar (DG/WG, Hays, 1964; Hudson, 1984), Head-driven Phrase Structure Grammar (HPSG, Pollard and Sag, 1987), Autolexical Grammar (ALG, Sadock, 1991), Role and Reference Grammar (R&RG, Van Valin, 1993), Sign-based Construction Grammar (SBCxG, Boas and Sag, 2012), Simpler Syntax (SS, Jackendoff, 1997, 2002; Culicover and Jackendoff, 2005), Type-Logical Grammar (TLG, Moortgat, 1988; Hepple, 1990; Dowty, 1993; Morrill, 1994; Jacobson, 1999), and some versions of Montague Grammar (MG, Montague, 1973; Bach, 1976; Cooper, 1983).

A second group sought for formalisms that were much less expressive in the first place, such as Generalized Phrase Structure Grammar (GPSG, Gazdar, 1981), Combinatory Categorial Grammar (CCG, Ades and Steedman, 1982; Szabolcsi, 1989), Head Grammar (HG, Pollard, 1984), and Tree Adjoining Grammar (TAG Joshi, 1988).

In the same period, the transformational theory itself evolved through the Extended Standard Theory (EST, Chomsky, 1972), the Revised Extended Standard Theory (REST, Chomsky and Lasnik, 1977), and Principles and Parameters/Government-Binding (P&P/GB, Chomsky, 1981), until the more radical reform of the current Minimalist Program (MP, Chomsky, 1995b,a, 2000), defined by the assumption that syntactic derivation is determined by its function of creating objects that are phonologically and semantically well-formed, an assumption that (at least in aspiration) makes it more akin to the latter approaches, and in particular, the present approach of Combinatory Categorial Grammar.

In particular, both Minimalism and CCG are committed to the view that syntactic derivation works by language-independent principles, from which it follows that all language-specific properties of constructions must derive from the lexicon of the language, not via language-specific rules or constraints.

In that sense, as Adger (2013) has pointed out, Chomskian Minimalism can be seen as a form of Categorial Grammar that adds Movement as the mechanism for handling discontinuous dependency, rather than the combinatory rules of CCG that are defined below. One of the purposes of the present chapter is to compare and contrast the movement theory of discontinuity with the alternative combinatory extension of Categorial Grammar proposed by CCG, as well as with the other alternative theories noted above.

In order to make these comparisons, the presentation of the formal specifics of CCG will be tied to the constructions in English and other languages that motivate their introduction. Our rules are empirically motivated, and it is only in the latter sections of the chapter that we turn to the question of why they take the form that they do.

The constructions in question fall into two groups. The first group comprises the “bounded” constructions like raising, control, passive, unaccusative, intransitivization, the binding of reflexive pronouns, etc., which concern relations or alternations in relations among the arguments of a single head, such as a verb, together with related matters like agreement and case. These are phenomena on which all theories more or less agree, differing only in the degree of lexicalization vs. syntacticalization that they assume, with CCG occupying the radically lexicalized end of the spectrum. These are dealt with fairly briefly in Section 3.[4]

The remaining sections of the paper concern a much more problematic range of constructions which we will loosely refer to as “unbounded”, which constitute a much more difficult problem for the theory of grammar, and for which CCG presents a radically different analysis from other theories. They include relativization and its allies such as topicalization and *wh*-question formation, together with its subspecies such as “pied-piping” and “parasitic” extractions, all of which have been attributed to unbounded movement, and are the subject of Section 4, which introduces all the remaining syntactic operations, and shows that the *wh*-constructions can be analysed without movement. Section 5 then goes on to show that various forms of coordination reduction, which have elsewhere been attributed to deletion under identity, copying, or parallelism, can be eliminated under exactly the same assumptions as movement.

Section 6 then briefly reviews the notion of constituency that is implicit in CCG, and notes that English intonation structure and its semantics reflect exactly the same notion of derivational constituency as the earlier constructions, without the stipulation of extra-syntactic features such as “edges” and non-syntactic processes such as “Focus Projection”.

Section 7 is more technical, and addresses the question of the degrees of freedom that have been exercised in achieving this account, and the explanatory adequacy of the theory that results. A brief conclusion then sums up.

3 Pure Categorial Grammar (CG)

The pure Combinatory Grammar of Ajdukiewicz and Bar-Hillel eschews language-specific syntactic production rules like (2) for English.

(2)
$$\begin{array}{lll} S & \rightarrow & NP\ VP \\ VP & \rightarrow & TV\ NP \\ TV & \rightarrow & \{proved,\ found,\ sees,\ \ldots\} \end{array}$$

4 Many of the constructions of central concern to Construction Grammarians arguably belong in this class of lexically-governed constructions.

Instead, the same language-specific syntactic information is lexicalized, via lexical entries like (3) for the English transitive verb:

(3) sees := $(S\backslash NP_{3s})/NP$

This "category" identifies the transitive verb as a function or governing category, specifying the type, directionality, and agreement of its NP arguments and the type of its result, S. Thus, it specifies "sees" as a transitive verb wanting an NP on the right (with unspecified agreement) as its first argument, to yield a function wanting an NP bearing third singular agreement on the left, to yield S.[5]

The lexical notation for Chomskean Minimalism is essentially categorial (Chomsky, 1995a, 2000; Stabler, 2011; Adger, 2013):

(4) sees :: { =D+case, =D V} ("yields V; selects two D (NPs); assigns case to the first")

(The above is Chomsky's notation, which omits directional alignment, like a categorial grammar with non-directional slashes $|\ X$, which loses CCG's transparency to language-specific linear order of governors and arguments. Stabler also discusses a Directional Minimalist Grammar (DMG) with =X and X= directionality, parallel to $/X$ and $\backslash X$.)

3.1 The categorial lexicon

When stated in full, categories also specify a semantics or logical form, as in (5a), as anatomized in (5b), in which the separator ":=" pairs a phonological/graphological form with a category, and the separator ":" pairs a syntactic type with a logical form:

(5) a. sees := $(S\backslash NP_{3s})/NP : \lambda x \lambda y.sees\, x\, y$

$$\text{b.}\quad \overbrace{\text{sees}}^{\text{phonological form}} \ := \ \overbrace{\overbrace{(S_{\underbrace{fin}_{\text{feature}}}\backslash NP_{3s})/NP}^{\text{syntactic type}} : \overbrace{\underbrace{\lambda x\,\lambda y.}_{\lambda\text{-binders}}\ \underbrace{sees\,x\,y}_{\text{predicate-argument structure}}}^{\text{logical form}}}^{\text{category}}$$

The predicate-argument structure component of logical form is assumed to be cross-linguistically universal, although elements like *sees* are of course a proxy for more complex structures involving tense, aspect, etc. Predicate-argument structure is therefore essentially equivalent to the lexical component of thematic structure in Minimalism, f-structure in LFG, ARG-ST in HPSG, the grammatical function tier of SS, and dependency structures in DG/WG, although unlike some of these formalisms, CCG

5 3s agreement is of course specified by -s morpho-phonology by a process discussed later in this section, and in Section 7. We assume a standard mechanism of simple non-recursive feature-value unification.

does not include explicit rôle-labels, or define these structures as linearly ordered or aligned.

The use of an explicit representational level of linguistic form, distinct from syntactic derivation, is a point of difference from the TLG tradition of Lambek in Categorial Grammar. While TLG often presents interpretations as λ-terms, they are included purely for ease of reading, and are proclaimed to be dispensible. The semantics itself is defined by "direct surface composition" on the syntactic derivation itself (Jacobson, 1999—see Bozşahin, 2012, 87–106 for discussion).

While direct surface compositionality is technically possible for CCG, there is a good reason to include a representational level of logical form. Since the only plausible account of child language acquisition is that it is semantically bootstrapped from a prior representation of meaning, and since that meaning representation must be independent of the surface syntax of any specific language, it must be a language-independent logical form. It is therefore syntactic derivation, rather than LF, that is dispensible as a level of representation.

To take a slightly more complex lexical verb, the following is the category for a subject control verb for a sentence such as *He promises her to leave*, again anatomized as (b):

(6) a. promises $:= ((S\backslash NP_{3s})/VP_{to})/NP : \lambda x\lambda p\lambda y.promises\,(p\,y)\,x\,y$

b. $$\overbrace{\text{promises}}^{\text{phonological form}} := \overbrace{\overbrace{((S_{\underbrace{fin}_{\text{feature}}}\backslash NP_{3s})/VP_{to})/NP}^{\text{syntactic type}} : \overbrace{\underbrace{\lambda x\,\lambda p\,\lambda y.}_{\lambda\text{-binders}}\;\underbrace{promises\,(p\,y)\,x\,y}_{\text{predicate-argument structure}}}^{\text{logical form}}}^{\text{category}}$$

The control relation between the subject and the infinitival VP complement is entirely captured at the level of the logical form, via the variables y and p. Here, the minimalist notation would be slightly different, treating the infinitival complement as a "small" clause, and treating the relation to the surface syntactic subject as mediated by movement or an anaphoric element. To that extent, CCG can be seen as lexicalizing the A-movement analysis *statically*, via the use of bound variables at the level of logical form. Because such lexicalization is by definition limited to relations between co-arguments of the verb, via the logical form, it necessarily obeys minimality conditions variously identified as "subjacency", "relativized minimality", "the Minimal Link Condition", etc., as do GPSG/HPSG/LFG also.

In CCG, all bounded constructions, such as passive, reflexivization, raising, and control, are lexically governed in a similar way, so that all instances of so-called A-movement are specified statically in the lexicon via the logical form of the governor.[6]

6 It follows that the phenomenon of so-called Backward Control, in which an explicit subject in an infinitival complement clause appears to bind an implicit argument in the matrix clause, as has been proposed for Tsez by Polinsky and Potsdam (2000), cannot be handled as control in CCG.

As in any theory of grammar, rather than listing every single lexical category in its own right, we may want to capture "parametric" generalizations across the lexicon for any given language via "lexical redundancy rules" Jackendoff (1975), such as that not only "sees", but every transitive tensed main verb has an SVO syntactic type, or that some identifiable class of the same transitive verbs including "eat" but not "find" map systematically onto a corresponding class of intransitivized verbs. Some, like intransitivization itself, will be morphologically unmarked, others like passivization, marked. Such regularities will allow the language learner to infer the existence of other members of such paradigms when they first encounter a novel verb, and possibly to learn more rapidly. They also have the advantage of allowing the grammar to be represented more compactly, although the fact that all of these generalizations are liable to admit of exceptions or irregularities shows that this is not the only consideration, and it may well be the case that such paradigms are compiled out into multiple lexical entries in their own right, as soon as encountered. All such solutions are formally equivalent, and the present chapter remains agnostic as to which should be preferred.

The present chapter maintains an open mind about exactly how those lexical generalizations should be captured. Give or take the notational idiosyncrasies applying to the lexicon, and the use in addition in some of the other theories discussed in this volume of language-specific syntactic rules, all theories are pretty much equivalent in this respect, and can be applied to the categorial lexicon. Accordingly, we will pass them over in this chapter, to concentrate on the syntactic component, which is more distinctive to CCG.

3.2 Syntactic rules I: Pure application

Verb categories like (5) and (6) combine most simply by the following rules, which are universal and language-independent:

(7) a. Forward Application:

$$X/_{\star}Y \quad Y \quad \Rightarrow \quad X \qquad (>)$$

b. Backward Application:

$$Y \quad X\backslash_{\star}Y \quad \Rightarrow \quad X \qquad (<)$$

X and Y can be any syntactic CCG type, and may include simple feature-value pairs, such as agreement. $\Rightarrow$ reads as "the things on the left combine to yield the thing on the right", and is entirely analogous to the reverse of the rewrite arrow of the PS rules

---

Cormack and Smith (2002) show that the construction in Tsez is limited to just two verbs, meaning "begin" and "continue", and that the supposedly controlling infinitival complement subject cannot be referential, as with a universally quantified NP. Both restrictions are reminiscent of *there* insertion in English, which is limited to raising verbs and non-referential subjects, so it seems reasonable to assume that so-called backward control arises from some form of lexicalized equivalent of expletive insertion at the level of lexical logical form.

in (2). $>$ and $<$ are annotations in derivations indicating the application of the rule in question. The slash-type $\star$ on the functor categories in (8) means that any functor category of the form X/Y or $X\backslash Y$ can combine by these rules.

Such rules, like all rules in CCG, correspond to the Minimalist operation of (external) **merge**, with the effect of "canceling" Y term, as if they were rules of fractional multiplication. They are analogous to Minimalist "feature-checking" between argument and governor or head (Adger, 2003, 90–96).

Such rules are rules of semantic merger, as well as syntactic. Thus we can extend them as follows, with ":" again acting as an uninterpreted separator between syntactic type and semantic interpretation or logical form:

(8) a. Forward Application:

$$X/_{\star}Y : f \quad Y : a \quad \Rightarrow \quad X : f\,a \qquad (>)$$

b. Backward Application:

$$Y : a \quad X\backslash_{\star}Y : f \quad \Rightarrow \quad X : f\,a \qquad (<)$$

Such rules are both syntactically and semantically rules of functional application, or combination of functions with their arguments, as in the following derivation, in which syntactic derivation and the composition of logical form are synchronous and homomorphic.

(9)
$$\begin{array}{ccc}
\text{Harry} & \text{sees} & \text{Sally} \\
\overline{NP_{3s}} & \overline{(S\backslash NP_{3s})/NP} & \overline{NP} \\
: harry & : \lambda x\lambda y.sees\,x\,y & : sally \\
 & \multicolumn{2}{c}{\overline{\hspace{10em}}{>}} \\
 & \multicolumn{2}{c}{S\backslash NP_{3s}} \\
 & \multicolumn{2}{c}{: \lambda y.sees\,sally\,y} \\
\multicolumn{3}{c}{\overline{\hspace{16em}}{<}} \\
\multicolumn{3}{c}{S} \\
\multicolumn{3}{c}{: sees\,sally\,harry}
\end{array}$$

(The absence of any explicit slash type on the categories in the derivation (9) means that those categories can combine by any rule, including some discussed below that are more restricted than the application rules.)

Since categories like (6) achieve the effect of "A-movement" via λ-binding at the level of logical form, the application rules are all that is needed to capture the phenomenon of control, as in the following derivation:

(10)
$$\begin{array}{ccccc}
\text{Harry} & \text{promises} & \text{Sally} & \text{to} & \text{leave} \\
\overline{NP_{3sm}} & \overline{((S\backslash NP_{3s})/VP_{to})/NP} & \overline{NP_{3sf}} & \overline{VP_{to}/VP} & \overline{VP} \\
: harry & : \lambda x\lambda p\lambda y.promises\,(p\,y)\,x\,y & : sally & : \lambda p.p & : \lambda y.leave\,y \\
 & \multicolumn{2}{c}{\overline{\hspace{14em}}{>}} & \multicolumn{2}{c}{\overline{\hspace{8em}}{>}} \\
 & \multicolumn{2}{c}{S\backslash NP_{3s}} & \multicolumn{2}{c}{VP_{to}} \\
 & \multicolumn{2}{c}{: \lambda p\lambda y.promises\,(p\,y)\,sally\,y} & \multicolumn{2}{c}{: \lambda y.leave\,y} \\
 & \multicolumn{4}{c}{\overline{\hspace{24em}}{>}} \\
 & \multicolumn{4}{c}{S\backslash NP : \lambda y.promises\,(leave\,y)\,sally\,y} \\
\multicolumn{4}{c}{\overline{\hspace{22em}}{<}} & \\
\multicolumn{4}{c}{S : promises\,(leave\,harry)\,sally\,harry} &
\end{array}$$

Such purely applicative derivations will correctly form "chains" of raising and control relations in examples like the following, which are left as exercises:

(11) a. Harry promises Sally to persuade Alice to leave.
b. Harry seems to promise Sally to leave.
c. Harry wants to try to begin to write a play.

3.3 The Combinatory Projection Principle

The application rules in (8) constitute directionally specified forms of the simplest "external" form of the Chomskian Minimalist operation "Merge". They conform to a simple generalization which governs all rules in CCG:

(12) *The Combinatory Projection Principle (CPP)*
Syntactic combinatory rules are *binary, linearly-ordered, type-dependent rules*, applying to string-adjacent categories, whose linear order is consistent with their directional types, and *project unchanged the type and directionality of any argument in the inputs that also appears in the result*.

This principle is defined more formally in Steedman, 2000 in terms of three more fundamental principles of *Adjacency* or Contiguity, Directional *Inheritance*, and Directional *Consistency*, and forbids rules like the following, which combine forward functions backward (a), combine inner arguments before outer (b), or switch directionality between input and output (c):

(13) a. $Y{:}a \quad X/Y{:}f \quad \not\Rightarrow \quad X{:}f\,a$
b. $(X/Y)/Z{:}f \quad Y{:}a \quad \not\Rightarrow \quad X/Z{:}f\,a$
c. $(X/Y)/Z{:}f \quad Z{:}a \quad \not\Rightarrow \quad X\backslash Y{:}f\,a$

All bounded constructions—that is, those defining relations over the arguments of a single head, such as raising, control, passive, unaccusatives, causatives, etc.—are defined morpholexically in CCG. Where there are systematic relations or alternations between subcategorizations by the same head, as in *the door opened*, *the door was opened*, *Harry opened the door*, *the door opened itself*, etc., these may be mediated by morphological markers, or by lexical rules, or by autonomous lexical entries, or by some mixture of the above, all of which may admit of phonological regularities and exceptions, as in many other theories mentioned above, such as LFG, HPSG, SS, etc. While such choices may be extremely important to efficient representation of the grammar for purposes of processing or acquisition, they are all formally equivalent, and will not be distinguished here.

3.4 Case and morpholexical type-raising

Case is assumed to be a universal primitive of grammar (cf. Vergnaud, 1977/2006). That is to say that all noun-phrases (NP) like "Harry" are (polymorphically) *type-raised* in the morpholexicon. Type raising is so-called because it assigns to predicate-argument-structural arguments the category of a higher-order function over predicates that take NPs like “Harry” as an argument. For example, in place of (9), we have

(14)
$$
\begin{array}{ccc}
\text{Harry} & \text{sees} & \text{Sally} \\
\hline
S/(S\backslash NP_{3s}) & (S\backslash NP_{3s})/NP & (S\backslash NP_{3s})\backslash((S\backslash NP_{3s})/NP) \\
:\lambda p.p\ harry & :\lambda x\lambda y.sees\ x\ y & :\lambda p.p\ sally \\
\cline{2-3}
& \multicolumn{2}{c}{S\backslash NP_{3s} \quad <} \\
& \multicolumn{2}{c}{:\lambda y.sees\ sally\ y} \\
\hline
\multicolumn{3}{c}{S \quad >} \\
\multicolumn{3}{c}{:sees\ sally\ harry}
\end{array}
$$

The effect of type-raising is to swap the roles of function and argument between subject and predicate, so that the forward application rule (8a) applies, rather than the backward rule (8b), and vice versa. crucially, the logical form that results is identical to that in (9).

Type-raising is the job of case morphemes like the nominative suffixes *-ga* in Japanese and *-us* in Latin, as in Figure 2a, in which the \\ double-slash indicates a morphemic function that can only apply inside the lexicon. In contrast, English NPs are underspecified as to case, as in Figure 2b and c (the latter involves first-person subject pro-drop, represented in the logical form as anaphorically bound one_{1s}).

Thus, even in English, type-raising is an operation of the lexical component of the grammar, not a syntactic rule.[7]

From now on we will usually abbreviate English underspecified type-raised NP etc. as $NP^{\uparrow}$ etc., with the meaning “whatever type-raised NP category is required for the derivation”. Determiners will accordingly be written as $NP^{\uparrow}/N$.

Type-raising (Case) makes arguments into function categories that are more like adjuncts or specifiers than like complements (Adger, 2013). Adger uses such type-raising to avoid problematic “roll-up” derivations under a minimalist approach. This use seems parallel to the use in CCG of lexicalized type-raising to capture pied-piping relatives and *in situ wh* in examples like (32) below.

3.5 Reflexive anaphora

We assume for present purposes that reflexive pronouns are clitic, like French *se*. The boundedness of reflexivization then arises from the fact that cliticization is a morpholexical process, despite the fact that in both languages the clitic in question is written as a separate word.

7 Of course, the processor might choose for reasons of efficiency to leave case under-specified, and apply type-raising dynamically in context, under the control of a parsing “oracle” such as a statistical parsing model. (This is in fact how all CCG parsers for English work.) However, that does not imply that it is a rule of syntax.

a. Balb +us mur +um ædificat

B. .NOM wall .ACC build.PRES.3SM

$$N_{3sm} \quad (S/(S\backslash NP_{nom,3sm}))\backslash\!\backslash N_{3sm} \quad N \quad ((S\backslash NP_{nom,agr})/((S\backslash NP_{nom,agr})\backslash NP_{acc}))\backslash\!\backslash N \quad (S_{pres}\backslash NP_{nom,3s})\backslash NP_{acc}$$

$$\underline{\qquad}^{<LEX}\quad S/(S\backslash NP_{nom,3sm}) \qquad \underline{\qquad}^{<LEX}\quad (S\backslash NP_{nom,agr})/((S\backslash NP_{nom,agr})\backslash NP_{acc})$$

$$\underline{\qquad}^{>}\quad S\backslash NP_{nom,3s}$$

$$\underline{\qquad}^{>}\quad S$$

"Balbus is building a wall."

b. Harry sees Sally

$$S/(S\backslash NP) : \lambda p.p\ harry \quad (S\backslash NP)/NP : sees \quad (S\backslash NP)\backslash((S\backslash NP)/NP) : \lambda p.p\ sally$$

$$\underline{\qquad}^{<}\quad S\backslash NP : sees\ sally$$

$$\underline{\qquad}^{>}\quad S : sees\ sally\ harry$$

c. missed the Saturday dance

$$S/NP : \lambda x.missed\ x\ one_{1s} \quad S\backslash(S/NP) : \lambda p.p\ saturdaydance$$

$$\underline{\qquad}^{<}\quad S : missed\ saturdaydance\ one_{1s}$$

d. man who I think likes Lester

$$N_{3s} : man \quad (N_{agr}\backslash_{\diamond}N_{agr})/_{\diamond}(S/NP_{+WH,agr})\ \lambda p\lambda n\lambda x.p\,x \wedge n\,x \quad NP^{\uparrow}_{1s}\ \lambda p.p\ me \quad ((S\backslash NP)/NP_{+WH,agr})/(S\backslash NP_{agr})\ \lambda p\lambda x\lambda y.think\,(p\,x)\,y \quad (S\backslash NP_{3s})/NP\ \lambda x\lambda y.likes\,x\,y \quad NP^{\uparrow}\ lester$$

$$\underline{\qquad}^{>\mathbf{B}^2}\quad (S/NP_{+WH,agr})/(S\backslash NP_{agr}) : \lambda p\lambda x.think\,(p\,x)\,me \qquad \underline{\qquad}^{<}\quad S\backslash NP_{agr} : \lambda y.likes\ lester\ y$$

$$\underline{\qquad}^{>}\quad S/NP_{+WH,3s} : \lambda x.think\,(likes\ lester\ x)\,me$$

$$\underline{\qquad}^{>}\quad N_{3s}\backslash N_{3s} : \lambda n\lambda x.think\,(likes\ lester\ x)\,me \wedge n\,x$$

$$\underline{\qquad}^{>}\quad N_{3s} : \lambda x.think\,(likes\ lester\ x)\,me \wedge man\,x$$

Figure 2

We have the following categories for clitic "himself", in which the morphological slash \\ restricts its application to lexical verbs:

(15) himself := $(S\backslash NP_{3sm})\backslash\backslash((S\backslash NP_{3sm})/NP) : \lambda p\lambda y.p\,(self\ y)\,y$
$((S\backslash NP_{3sm})/PP)\backslash\backslash(((S\backslash NP_{3sm})/PP)/NP) : \lambda p\lambda w\lambda y.p\,w\,(self\ y)\,y$
etc.

Syntactically, these categories are accusative instances of type-raised cased $NP^{\uparrow}$. However, their semantics is not **T**.[8]

The derivation for a simple reflexive transitive clause is the following, where *self harry* evaluates to *harry*:

(16)
$$\begin{array}{lll}
\text{Harry} & \text{sees} & \text{himself.}\\
\cline{1-1}\cline{2-3}
NP^{\uparrow}_{3sm} & (S\backslash NP_{3s})/NP & (S\backslash NP_{3sm})\backslash\backslash((S\backslash NP_{agr})/NP)\\
:\lambda p.p\ harry & :\lambda x\lambda y.sees\ x\ y & :\lambda p\lambda y.p\ (self\ y)\ y\\
\cline{2-3}
 & \multicolumn{2}{c}{S\backslash NP_{3sm} : \lambda y.sees\ (self\ y)\ y \quad {<}LEX}\\
\cline{1-3}
\multicolumn{3}{c}{S : sees\ (self\ harry)\ harry \quad {>}}
\end{array}$$

For reflexive ditransitives of the kind we saw in (1), we have the following:

(17)
$$\begin{array}{llll}
\text{Mary} & \text{introduced} & \text{herself} & \text{to the audience.}\\
\cline{1-1}\cline{2-3}\cline{4-4}
NP^{\uparrow}_{3sf} & ((S\backslash NP_{agr})/PP_{to})/NP & ((S\backslash NP_{3sf})/PP)\backslash\backslash(((S\backslash NP_{3sf})/PP)/NP) & PP^{\uparrow}\\
:\lambda p.p\ mary & :\lambda x\lambda w\lambda y.introduced\ w\ x\ y & :\lambda p\lambda w\lambda y.p\ w\ (self\ y)\ y & :\lambda p.p\ audience\\
\cline{2-3}
 & \multicolumn{2}{c}{(S\backslash NP_{3sf})/NP : \lambda w\lambda y.introduced\ w\ (self\ y)\ y \quad {<}LEX} & \\
\cline{2-4}
 & \multicolumn{3}{c}{S\backslash NP_{3sf} : \lambda y.introduced\ audience\ (self\ y)\ y \quad {<}}\\
\cline{1-4}
\multicolumn{4}{c}{S : introduced\ audience\ (self\ mary)\ mary \quad {>}}
\end{array}$$

It seems reasonable to assume that *Harry talks to himself* is also a true *se*-type reflexive arising from lexicalization of "talks to", as in the following derivation:[9]

(18)
$$\begin{array}{lll}
\text{Harry} & \text{talks to} & \text{himself.}\\
\cline{1-1}\cline{2-3}
NP^{\uparrow}_{3sm} & (S\backslash NP_{3s})/NP & ((S\backslash NP_{3sm}))\backslash\backslash((S\backslash NP_{agr})/NP)\\
:\lambda p.p\ harry & :\lambda x\lambda y.talks\ (to\ x)\ y & :\lambda p\lambda y.p\ (self\ y)\ y\\
\cline{2-3}
 & \multicolumn{2}{c}{S\backslash NP_{3sm} : \lambda y.talks\ (to\ (self\ y))\ y \quad {<}LEX}\\
\cline{1-3}
\multicolumn{3}{c}{S : talks\ (to\ (self\ harry))\ harry \quad {>}}
\end{array}$$

Example (19a) can be analysed similarly to (17). However, the reflexives in (19b–e) cannot reasonably be analysed as clitic in the same way, and must be "exempt" or logophoric pronouns, of a kind to be discussed below:

(19) a. Harry showed himself a movie.
b. Harry showed a movie to himself.

8 The analysis is similar to that of Szabolcsi (1989), which is also lexicalized.

9 This possibility may be related to the cross-linguistically unusual possibility in English of "preposition-stranding" *wh*-extraction.

c. Harry talks to and about himself.
d. Harry talks to only himself.
e. Harry sees and admires himself.

The following further "subject reflexive" instance of the type-raised reflexive for the non-existent "*heself", (20) is excluded for English because it is not a possible English cased category, since $(S\backslash NP)/NP_{3sm}$ is not an English transitive verb category:[10]

(20) *heself := $(S/NP)/\!/((S\backslash NP)/NP_{3sm}) : \lambda p\lambda x.p\,x\,(self\ x)$

The CCG identification of a languages case-system with type-raising over its verbal categories therefore predicts the "anaphor agreement effect" of Rizzi (1990), rather than requiring it as a stipulative constraint, thereby capturing Condition A of Chomsky (1981).

The presence in English of "logophoric" reflexives that are homophonous to the reflexive, but are non-clause bound, like pronouns, is a source of confusion. Such forms are exempt from the binding conditions, and refer to the individual whose viewpoint the text presents (Jackendoff, 1972; Higgins, 1973; Zribi-Hertz, 1989; Pollard and Sag, 1992), as in:

(21) a. The fact that there is a picture of himself$_i$ hanging in the post office is believed by Mary to be disturbing Tom$_i$.
b. A fear of himself$_i$ is John$_i$'s greatest problem.
c. John saw a picture of himself.

We will assume following Pollard and Sag that cases attributed to "reconstruction" like the following in fact arise from the involvement of exempt logophoric pronouns of this kind, rather than from true reflexives.

(22) a. Which pictures of himself$_i$ did Harry$_i$ see?
b. Alice wonders which pictures of himself$_i$ Harry$_i$ saw.
c. Alice wonders who$_i$ saw which pictures of himself$_i$.

4 The unbounded *wh*-constructions

Relativization and all unbounded dependencies crucially involve type-raising and the syntactic combinatory rules of function composition, considered next.

10 $(S\backslash NP)/NP_{3sm}$ is in fact the category of an *ergative* transitive verb with absolutive agreement. Woolford (1999, 267–8) shows that ergative languages with absolutive agreement such as Inuit and Nez Perce disallow the equivalent of "He$_{erg}$ sees himself$_{abs}$", which would require raising over the English transitive type, while ergative languages without absolutive agreement, such, as Enga, do allow them.

4.1 Syntactic rules II: Composition

We will need the following rules of composition, which constitute all and only those allowed by the CPP (12) for these (first-order) categorial types:

(23) a. Forward Composition:

$$X/_{\diamond}Y : f \quad Y/Z : g \quad \Rightarrow \quad X/Z : \lambda z.f\,(g\,z) \qquad (>\mathbf{B}\ )$$

b. Backward Composition:

$$Y\backslash Z : g \quad X\backslash_{\diamond}Y : f \quad \Rightarrow \quad X\backslash Z : \lambda z.f\,(g\,z) \qquad (<\mathbf{B}\ )$$

c. Forward Crossing Composition:

$$X/_{\times}Y : f \quad Y\backslash Z : g \quad \Rightarrow \quad X\backslash Z : \lambda z.f\,(g\,z) \qquad (>\mathbf{B}_{\times})$$

d. Backward Crossing Composition:

$$Y/Z : g \quad X\backslash_{\times}Y : f \quad \Rightarrow \quad X/Z : \lambda z.f\,(g\,z) \qquad (<\mathbf{B}_{\times})$$

Like the application rules (8), these rules have the effect of "canceling" Y, as if this were fractional multiplication. The types $\diamond$ and $\times$ on the slashes in these rules mean that only categories whose own slash type is compatible under a type-hierarchy of slash types (Baldridge, 2002) can combine by these rules. The simplified convention used in the present paper is that only categories with $\diamond$ slashes or unrestricted slashes can combine by $\diamond$ rules, and $\diamond$ categories cannot combine by the crossing $\times$ rules. Similarly, categories with a $\times$ or unrestricted slash-type can combine by $\times$ rules, but $\times$ categories cannot combine by $\diamond$ rules.[11]

The rules in (23) obey the CPP (12), including the Principle of Adjacency or contiguity. The absence of slash-typing on the secondary function $Y|Z$ means that it can apply to any type, but the CPP requires that that type will be passed to the result $X|Z$. Thus, as with the application rules (8), in Minimalist terms, the composition rules (23) constitute additional cases of External Merge, except for allowing the equivalent of some "feature-checking" that is not allowed under Minimalist merger (Adger, 2003, 93–94). However, they thereby achieve the same result as Minimalist *Internal* Merge, or Move.

To see this in the context of the relative clause, we will assume the following category for the English relative pronoun:

(24) that := $(N\backslash_{\diamond}N)/_{\diamond}(S/NP)$

We can then derive a relativized noun modifier "that she had met" of type $N\backslash_{\diamond}N$ (from the introductory example (1) and Figure 1b, slight simplified), as follows, using the first of the composition rules (23a) to form a constituent of type $S\backslash NP$ adjacent to the relative pronoun by adjacent merger of the elements of the residue of relativization.

11 However, another less restrictive convention is possible, in which these two slash types are explicitly conjunctive, written $\diamond*$ and $\times*$, while $\diamond$ and $\times$ types can *only* compose.

(25) (The man)

$$\begin{array}{llll}
\text{that} & \text{she} & \text{had} & \text{met} \\
\hline
(N\backslash_{\diamond}N)/_{\diamond}(S/NP) & S/(S\backslash NP) & (S\backslash NP)/VP_{pptcpl} & VP_{pptcpl}/NP \\
: \lambda p\lambda n\lambda x.p\,x \wedge n\,x & : \lambda p.p\,pro_{she} & : \lambda p\lambda y.past\,(perfect\,(p\,y)) & : \lambda x\lambda y.meet\,x\,y \\
\cline{2-3}
 & \multicolumn{2}{c}{S/VP \quad {>}\mathbf{B}} & \\
 & \multicolumn{2}{c}{: \lambda p.past\,(perfect\,(p\,pro_{she}))} & \\
\cline{2-4}
 & \multicolumn{3}{c}{S/NP \quad {>}\mathbf{B}} \\
 & \multicolumn{3}{c}{: \lambda x.past\,(perfect(meet\,x\,pro_{she}))} \\
\cline{1-4}
\multicolumn{4}{c}{N\backslash_{\diamond}N \quad {>}} \\
\multicolumn{4}{c}{: \lambda n\lambda x.past\,(perfect(meet\,x\,pro_{she})) \wedge n\,x}
\end{array}$$

4.1.1 Unbounded relativization

The same combinatory rule (23a) can apply recursively, to build a constituent of the same type S/NP by multiple adjacent mergers, to which the relative pronoun can apply as before to yield a noun modifier (semantics omitted):

(26) (The man)

$$\begin{array}{llllll}
\text{that} & \text{she} & \text{says} & \text{that} & \text{she} & \text{met} \\
\hline
(N\backslash_{\diamond}N)/_{\diamond}(S/NP) & NP^{\uparrow} & (S\backslash NP)/_{\diamond}S' & S'/_{\diamond}S & NP^{\uparrow} & (S\backslash NP)/NP \\
\cline{2-3} \cline{5-6}
 & \multicolumn{2}{c}{S/_{\diamond}S' \quad {>}\mathbf{B}} & & \multicolumn{2}{c}{S/NP \quad {>}\mathbf{B}} \\
\cline{4-6}
 & & & \multicolumn{3}{c}{S'/NP \quad {>}\mathbf{B}} \\
\cline{2-6}
 & \multicolumn{5}{c}{S/NP \quad {>}\mathbf{B}} \\
\cline{1-6}
\multicolumn{6}{c}{N\backslash_{\diamond}N \quad {>}}
\end{array}$$

4.1.2 Embedded subject extraction

However, the $\diamond$ slash-types of the complement of *think* correctly forbids extraction of the subject of a *that* complement:

(27) *men$_N$ that [she says that]$_{S/_{\diamond}S}$ [met her]$_{S\backslash NP}$

This is not a stipulation that could be otherwise: if a language like English allowed *says* or *says that* to compose with *sees her* by crossed composition it would immediately allow such non-English orders as the following with the meaning "she says the men met her":

(28) *She the men says met her.

Such considerations lead us to expect that a general prohibition against extraction of complement subjects should be a specific characteristic of rigidly SVO languages and constructions, in which the directionality of S and O arguments differ, but not of SOV and VSO, in which they are the same, a generalization that appears to be correct (Pesetsky, 2017).

It also follows that in order to overcome the general prohibition on subject extraction out of bare complements for the class of English verbs that allow them, we need to give them an extra lexical category, like the following:

(29) think, believe, ... $:= ((S\backslash NP)/NP_{+WH,agr})/(S\backslash NP_{agr}) : \lambda p \lambda x \lambda y.think(px)y$

This category combines with a tensed predicate to yield something requiring an NP marked as $+WH$ which can only reduce with a relativized category, as in the derivation of "[a] man who I think likes Lester" in Figure 2d.[12]

4.1.3 Parasitic extraction

For completeness, we note in passing that a further class of rule including the one here indexed $<\mathbf{S}_{\times}$, constituting a "duplicating" generalization of composition is needed to capture "parasitic gapping" cases of relativization, again building a constituent of type S/NP by a succession of adjacent mergers:

(30) (The articles)

$$
\begin{array}{ccccc}
\text{that} & \text{Harry} & \text{rejected} & \text{without} & \text{reading} \\
\hline
(N\backslash_{\diamond}N)/_{\diamond}(S/NP) & NP^{\uparrow} & (S\backslash NP)/NP & ((S\backslash NP)\backslash(S\backslash NP))/VP & VP/NP \\
 & & & \multicolumn{2}{c}{\overline{\quad ((S\backslash NP)\backslash(S\backslash NP))/NP \quad}^{>\mathbf{B}}} \\
 & & \multicolumn{3}{c}{\overline{\quad (S\backslash NP)/NP \quad}^{<\mathbf{S}_{\times}}} \\
 & \multicolumn{4}{c}{\overline{\quad S/NP \quad}^{>\mathbf{B}}} \\
\multicolumn{5}{c}{\overline{\quad N\backslash_{\diamond}N \quad}^{>}}
\end{array}
$$

We pass over the details of the **S** rules here, referring to Szabolcsi (1983, 1989) and Steedman (1987, 1996), noting not only that they are constrained by the Combinatory Projection Principle (12), but also that they exploit all degrees of freedom allowed under that principle, which correctly allows parasitism to be supported by complement subject extraction, forcing the choice of *rightward* subcategorization for the extracting NP in the subject-extracting category (29).

(31) A man that Harry will [tell you is a crook]$_{VP/NP_{+WH}}$ [while pretending to admire]$_{(VP\backslash VP)/NP}$

4.1.4 Pied-piping extraction

The phenomenon of "pied-piping" in relativizations like the following is captured by giving relative pronouns like *which* (but not *that*) the further category shown in the fol-

12 The reason for the extracting NP being a *rightward* argument will become clear below, when we consider the "Across-the-Board" constraint on extraction under coordination.

lowing example, the details of whose derivation we pass over (Steedman, 1987; Morrill, 1994; Steedman, 2012):

(32) [[Reports]$_N$ [[[the height of the lettering on the covers of]$_{(S\backslash(S/NP))/NP}$ [which]$_{((N\backslash N)/(S/NP))\backslash(((S\backslash(S/NP))/NP))}$ [the government prescribes]$_{S/NP}$

4.2 Crossing dependencies

The inclusion of crossing composition rules, together with the following generalization of the composition rules to "higher order" cases with second-order secondary functors of the form $(Y/Z)/W$ allows the set of possible non-terminal categories to grow unboundedly, showing CCG grammars to be trans-context free.

(33) a. Forward 2nd-order Composition:

$$X/_{\diamond}Y : f \quad (Y/Z)/W : g \quad \Rightarrow \quad (X/Z)/W : \lambda w\lambda z.f\,(g\,w\,z) \qquad (>\mathbf{B}^2)$$

b. Backward 2nd-order Composition:

$$(Y\backslash Z)\backslash W : g \quad X\backslash_{\diamond}Y : f \quad \Rightarrow \quad (X\backslash Z)\backslash W : \lambda w\lambda z.f\,(g\,w\,z) \qquad (<\mathbf{B}^2)$$

c. Forward Crossing 2nd-order Composition:

$$X/_{\times}Y : f \quad (Y\backslash Z)\backslash W : g \quad \Rightarrow \quad (X\backslash Z)\backslash W : \lambda w\lambda z.f\,(g\,w\,z) \qquad (>\mathbf{B}^2_{\times})$$

d. Backward Crossing 2nd-order Composition:

$$(Y/Z)/W : g \quad X\backslash_{\times}Y : f \quad \Rightarrow \quad (X/Z)/W : \lambda w\lambda z.f\,(g\,w\,z) \qquad (<\mathbf{B}^2_{\times})$$

(Again, the effect of these rules is to "cancel" Y.) We assume following *SP* that these rules are the only higher-order composition rules. (In particular, there are no such rules with mixed directionality in the secondary function.)

This feature of the theory allows elegant capture of a Germanic control construction that allows unboundedly many arguments to cross dependencies with their governing verbs, as in Figure 3a,b. This was a phenomenon which allowed the first formal proof that natural languages were not even weakly context-free (Huybregts, 1984; Shieber, 1985, data for Zurich German from Shieber).

Some alternative orders to those in Figure 3a,b including the following are correctly also allowed in CCG (Shieber, 1985):

(34) a. Das mer em Hans hälfed es huus aastriiche.

b. Das mer d'chind lönd em Hans hälfe es huus aastriiche.

5 Coordination

5.1 Right node raising

Because the category S/NP of the domain of relativization does not distinguish the status of the argument $/NP$ as extracted or lexical, we immediately predict the possibility a.

das	mer	em Hans	es huus	hälfed	aastriiche
that	we–NOM	Hans–DAT	the house–ACC	helped	paint
	$NP^{\uparrow}_{nom}$	$NP^{\uparrow}_{dat}$	$NP^{\uparrow}_{acc}$	$((S_{+SUB}\backslash NP_{nom})\backslash NP_{dat})/_{\times}VP$	$VP\backslash NP_{acc}$

$$\xrightarrow{>\mathbf{B}_{\times}} ((S_{+SUB}\backslash NP_{nom})\backslash NP_{dat})\backslash NP_{acc}$$

$$\xrightarrow{>} (S_{+SUB}\backslash NP_{nom})\backslash NP_{dat}$$

$$\xrightarrow{>} S_{+SUB}\backslash NP_{nom}$$

$$\xrightarrow{>} S_{+SUB}$$

"that we helped Hans paint the house"

b.

das	mer	d'chind	em Hans	es huus	lönd	hälfe	aastriiche
that	we–NOM	the children–ACC	Hans–DAT	the house–ACC	let	help	paint
	$NP^{\uparrow}_{nom}$	$NP^{\uparrow}_{acc}$	$NP^{\uparrow}_{dat}$	$NP^{\uparrow}_{acc}$	$((S_{+SUB}\backslash NP_{nom})\backslash NP_{acc})/_{\times}VP$	$(VP\backslash NP_{dat})/_{\times}VP$	$VP\backslash NP_{acc}$

$$\xrightarrow{>\mathbf{B}^{2}_{\times}} (((S_{+SUB}\backslash NP_{nom})\backslash NP_{acc})\backslash NP_{dat})/_{\times}VP$$

$$\xrightarrow{>\mathbf{B}_{\times}} (((S_{+SUB}\backslash NP_{nom})\backslash NP_{acc})\backslash NP_{dat})\backslash NP_{acc}$$

$$\xrightarrow{>} ((S_{+SUB}\backslash NP_{nom})\backslash NP_{acc})\backslash NP_{dat}$$

$$\xrightarrow{>} (S_{+SUB}\backslash NP_{nom})\backslash NP_{acc}$$

$$\xrightarrow{>} S_{+SUB}\backslash NP_{nom}$$

$$\xrightarrow{>} S_{+SUB}$$

"that we let the children help Hans paint the house"

Figure 3: Crossing dependencies in Zurich German.

of unbounded *rightward* movement, as in the following derivation:

$$
\begin{array}{lcccc}
(35) & [\text{Harry sees}] & \text{and} & [\text{Fred says he likes}] & \text{Sally}\\
\cline{2-5}
 & S/NP \;{>}\mathbf{B} & (X\backslash_{\star}X)/_{\star}X & S/NP \;{>}\mathbf{B} & NP^{\uparrow}\\
\cline{3-4}
 & & \multicolumn{2}{c}{(S/NP)\backslash(S/NP) \;{>}} & \\
\cline{2-4}
 & \multicolumn{3}{c}{(S/NP) \;{<}} & \\
\cline{2-5}
 & \multicolumn{4}{c}{S \;{<}}
\end{array}
$$

The category $(X\backslash_{\star}X)/_{\star}X \;:\; \lambda p\lambda q.p \sqcap q$ of the conjunction is restricted by $\star$ slash-typing to only combining by the application rules (8) rather than the more restricted composition rules (23). This condition imposes Ross' 1967 "Across the Board" (ATB) constraint (1967) on both rightward and leftward extraction out of coordinate structures including the "same case" condition, as in the following examples.

(36) a. the woman that [sees Harry]$_{S/NP}$ and [likes Fred]$_{S\backslash NP}$
b. the woman that [Harry sees]$_{S/NP}$ and [Fred likes]$_{S/NP}$
c. *the woman that[Harry sees]$_{S/NP}$ and [Fred likes her]$_{S}$
d. *the woman that[Harry sees her]$_{S}$ and [Fred likes]$_{S/NP}$
e. *the woman that[Harry sees]$_{S/NP}$ and [likes Fred]$_{S\backslash NP}$

The restriction on the conjunction category prevents application to S in one conjunct and composition into S on the other. This restriction should be seen as a consequence of the semantics, which is defined as Partee and Rooth's transitive closure $p \sqcap q$ over function types (1983), and must therefore apply to p and q of the same type.

Constituents including extracted complement subjects can coordinate with those containing extracted objects:

(37) a man that we had invited, and believed would come

Under the CCG account of coordination proposed here, this fact again forces the assumption that they have the same category—in this case VP_{pstp}/NP—with the subject-extracted ones differing only in being restricted to *wh* "antecedent government" via the subject-extracting category (29).

The same restriction means that right-node raising corresponding to (37) is blocked, in a rare exception to the symmetry of right- and left-extraction:

(38) *We had invited, and believed would come, the man who broke the bank at Monte Carlo.

5.2 Argument/Adjunct cluster coordination

Less obviously, the assumption that all arguments are mopholexically *cased*, or type-raised, including accusatives and datives as well as nominatives predicts the following "Argument/Adjunct Cluster" coordination (Dowty, 1988):

(39)

$$
\begin{array}{cccccc}
\text{Give} & \text{Harry} & \text{books} & \text{and} & \text{Sally} & \text{records}\\
\hline
(VP/NP)/NP & (VP/NP)\backslash((VP/NP)/NP) & VP\backslash(VP/NP) & (X\backslash X)/X & (VP/NP)\backslash((VP/NP)/NP) & VP\backslash(VP/NP)\\
give & :\lambda p.p\ harry & :\lambda p.p\ books & :\lambda q\lambda p.p\sqcap q & :\lambda p.p\ sally & :\lambda p.p\ records\\
\cline{2-3}\cline{5-6}
 & \multicolumn{2}{c}{{}^{<\mathbf{B}}\ VP\backslash((VP/NP)/NP)} & & \multicolumn{2}{c}{{}^{<\mathbf{B}}\ VP\backslash((VP/NP)/NP)}\\
 & \multicolumn{2}{c}{:\lambda p.p\ harry\ books} & & \multicolumn{2}{c}{:\lambda p.p\ sally\ records}\\
\cline{4-6}
 & & & \multicolumn{3}{c}{{}^{>}\ (VP\backslash((VP/NP)/NP))\backslash(VP\backslash((VP/NP)/NP))}\\
 & & & \multicolumn{3}{c}{:\lambda p\lambda r.r\ harry\ books\wedge p\ r}\\
\cline{2-6}
 & \multicolumn{5}{c}{{}^{<}\ VP\backslash((VP/NP)/NP)}\\
 & \multicolumn{5}{c}{:\lambda r.r\ harry\ books\wedge r\ sally\ records}\\
\hline
\multicolumn{6}{c}{{}^{<}\ VP : give\ harry\ books\wedge give\ sally\ records}
\end{array}
$$

The argument cluster coordination construction (39) is an example of a universal tendency for “deletion under coordination” to respect basic word order: in all languages, if arguments are on the left of the verb then argument clusters coordinate on the left, if arguments are to the right of the verb then argument clusters coordinate to the right of the verb, while SVO languages pattern with verb-initial (Ross, 1970):

(40)			
	SVO:	*SO and SVO	SVO and SO
	VSO:	*SO and VSO	VSO and SO
	SOV:	SO and SOV	*SOV and SO

For example, all contiguous substrings of Shieber’s Zurich German examples in Figure 3a,b are correctly predicted to coordinate with sequences of the same type (Steedman, 1985). (However, there is more to say concerning the precise mechanism that allows verb-medial gapping in the SVO case, and why it patterns with VSO—see Steedman, 1990, 2000.)

Such cluster coordinations were the motivation for Pesetsky’s (1995) postulation of a level of “cascade structure” as the domain of binding and coordination, distinct from “layered structure”, the domain of movement. In CCG, layered and cascade structure correspond to the same single level of derivation structure.

6 On intonation structure and the notion “surface constituent” in CCG

It is important to be clear at this point that CCG categories like S/NP are not equivalent to SLASH notations in GPSG/HPSG. In particular, in CCG, a category of the form X/NP does not denote a constituent of type X including a trace or gap of type NP.

On the contrary, S/NP is in CCG a constituent type in its own right, free to either combine with a preposed or *in situ* relativized element, or to combine with a full NP—in particular, one that is right node-raised across-the-board, as in (35).[13]

13 The accounts of right node-raising in Gazdar (1981) and Gazdar et al. (1985) both require a separate metarule to license RNR.

As a consequence, CCG generalizes the notion of constituency beyond the traditional notion to include *any sequence that is typable using combinatory rules*, including *Harry sees*, of type S/NP, and even *Harry books*, which we saw in (39) can be typed as $VP\backslash((VP/NP)/NP)$. As a consequence, CCG also necessarily allows (many) alternative derivations or constituent structures for canonical sentences. For example, as well as the earlier standard derivation (14), it allows the following:

(41)
$$\dfrac{\dfrac{\dfrac{\text{Harry}}{NP^{\uparrow} : \lambda p.p\ harry} \quad \dfrac{\text{sees}}{(S\backslash NP)/NP : sees}}{S/NP : \lambda x.sees\ x\ harry}{>}\mathbf{B} \quad \dfrac{\text{Sally}}{NP^{\uparrow} : \lambda p.p\ sally}}{S : sees\ sally\ harry}{<}$$

For longer sentences of length n there will be a number of alternative analysis up to the nth number in the more than exponentially rapidly-growing Catalan series, all yielding the same logical form.

This proliferation of constituent structures is sometimes referred to as "spurious ambiguity". However, it should not be regarded as a weakness in the combinatory theory of grammar. Many languages have freer word order than English, and do not support any clear notion of surface constituency. Even for English, there is no clear consensus on whether the surface structure of the ditransitive VP or the subject auxilliary-inverted question is flat and ternary-branching, left-branching, or right-branching (Adger, 2003, 122–131, and cf. Barss and Lasnik, 1986; Larson, 1988; Pesetsky, 1995; Jackendoff, 1990; Larson, 1990, *passim*). Nor is structural ambiguity a problem for performance or processing. It is just a fact of life (many other constructions, such as noun-noun compounding, yield Catalan-serial numbers of analyses). Parsers are good at dealing with other sources of ambiguity by the use of an "oracle" such as a statistical model, and they can do the same with this one. (It is worth remembering that ambiguity is endemic in all natural languages, and that none of them shows the slightest sign of evolving in the direction of reducing their overall level of ambiguity—Labov, 1994, 40–42, chs. 19,20; Croft, 2000, 68, 102–4; Newmeyer, 2003, 694; *passim*.)

CCG's unorthodox notion of constituency is transparently reflected in prosodic structure. Thus the following intonation contours appropriately mark the two alternative derivations for the transitive clause appropriate to the two context setting questions Q:

(42) Q: I know BARRY sees ALICE. But who sees SALLY?

A: (HARRY) (sees SALLY) .

(HARRY)	(sees SALLY) .
H*	L+H* LH%
$NP^{\uparrow}$	$S\backslash NP$

(43) Q: I know BARRY sees ALICE. But who does HARRY see?

A: (HARRY sees) (SALLY) .

L+H* LH % H* LL %

$\underline{\qquad\qquad\qquad}$ $\underline{\qquad\quad}$

S/NP $NP^{\uparrow}$

The notation for intonation-phrasal tunes is from Pierrehumbert and Hirschberg (1990). Here, L+H* LH % marks topic or Theme, H* LL % marks comment or Rheme. Exchanging the A(nswer)s to the Q(uestions) is highly unacceptable. Steedman (2014) develops a theory of intonation structure and its meaning using the Alternative-Semantic framework of Rooth (2016).

As in the Match Theory of Intonation Structure of Selkirk (2011) and its earlier Edge-marking incarnation (1990), Theme/Rheme marking is projected onto phrasal constituents directly, by syntactic derivation alone, bounded by combination of the phrase with an edge-based boundary tone. However, no independent extra-syntactic mechanism of "Focus Projection" is needed to achieve the semantics of "broad focus". Nor are any violable constraints needed to explain departures of intonation structure from syntactic derivational structure, for there are no such departures. In CCG, surface syntactic structure is simply identical to phonological form.

Thus, the domain of the prosodic phrase ϕ is the same as the domain of *wh*-movement, a state which is aspired to in Minimalist Contiguity Theory (Richards, 2016, Ch.3; Richards, 2016, 9). Prosodic structure is thereby defined as part of "Narrow Syntax" in the sense of Chomsky (2001).[14]

7 Explanatory adequacy

CCG's combination of type-raising and composition subject to the CPP (12) yields a permuting and rebracketing calculus closely tuned to the needs of and constraints on natural grammar.

CCG thereby reduces Minimalisms' MOVE/INTERNAL MERGE and COPY/DELETE, together with intonational phrasing, to contiguous EXTERNAL MERGE.

Constraints on dependency projection, such as the **that-t* and across-the-board conditions, arise from the nature of the lexicon and combinatorics of CCG rather than from additional constraints on syntactic derivation. Part of this explanatory force arises from the low expressive power of the combinatory rules under the CPP, to which we now turn.

14 However, there are some important differences. For Richards, the possibility of *in situ wh*-elements depends on everything between COMP and *wh* forming a single prosodic phrase. Otherwise, *wh*-movement is forced. In the present terms, *wh*-movement *also* is only possible if everything can be composed and thereby also become a contiguous intonational phrase.

7.1 Expressive power of CCG

CCG and TAG are provably weakly equivalent to Linear Indexed Grammar (LIG) (Vijay-Shanker and Weir, 1994, Kuhlmann et al., 2015).[15] Both are therefore "Mildly Context Sensitive" under the definition of Joshi (1988) and Joshi et al. (1991), which (among other properties), in the latter case calls for non-permutation-completeness.

In particular, of the $n!$ possible permutations on n functional heads, CCG only allow a proportion defined by the nth Large Schröder Number $S(n)$. For example, for a "cartographic" right-branching spine of only 8 functional heads, nearly 80 % of the 8! permutations are excluded.

This property was first noted by Williams (2003, 125) for his categorial system CAT. Williams (2003, 209) wrongly claimed that the inclusion of type-raising in CCG would allow all orders. However, Williams failed to note that, as we have seen, type-raising in CCG is morpho-lexical and defined over lexical functions over the original types, rather than a free syntactic operation. Type-raising changes the set of types involved, and therefore changes the "Basic Order of Merger", defined by purely applicative derivation. Nevertheless, any fixed set of types, including raised types, is as a consequence non-permutation-complete.

Specifically, for a set of four categories of the form $A|B$, $B|C$, $C|D$, D, determining a basic order of merger $\{1, 2, 3, 4\}$, 22 out of the 24 possible permutations are allowed. The two that CCG excludes are the following:

(44) a. 3 1 4 2
 b. 2 4 1 3

An example of a construction of this form is the nounphrase construction investigated by Cinque (2005) and Abels and Neeleman (2012), and mentioned in Section 1.5, for which the categories in English are the following:

(45) 1: $\text{these}_{NP^{\uparrow}/N_{num}}$ 2: $\text{five}_{N_{num}/N}$ 3: $\text{fat}_{N/N}$ 4: cats_{N}

The prediction is that no language will require or allow orders corresponding to either of the following glosses:[16]

(46) a. *fat these cats five
 b. *five cats these fat

These two orders are indeed not listed among the fourteen orders that Cinque identifies as attested for the languages of the world, nor are they included among the nine-

15 Weak equivalence means that they admit the same stringsets, though not via the same derivations. Kuhlmann et al. (2015) show that the specific slash-typing version of CCG presented in this chapter is actually slightly less expressive than TAG.

16 In testing such predictions, Cinque (2005) points out that it is important to be sure in particular that adjectives like "fat" are functioning as such, rather than being extraposed.

teen orders that Nchare (2012) identifies as occurring in the free word-order language Shupamem.

If we renumber the original set 1, 2, 3, 4 as X, 1, 2, 3, then (44b) is the *1-3-X-2 constraint on movement observed by Svenonius (2007) for adjuncts, an observation which led Svenonius to complex stipulations of strong features and null functional heads to limit movement in "roll-up" derivations such as pied-piping in Germanic. Thus, it seems likely that the orders in (44) are indeed universally excluded. In CCG this restriction is of the kind identified in *Aspects* as a Formal Universal, stemming without stipulation from the theory of grammar itself.

This property of non-permutation-completeness puts CCG at a different level of the extended language hierarchy of "abstract families of languages" (Ginsburg and Greibach, 1967) than standard Minimalist theories. Michaelis (1998, 2001) showed that Minimalist Grammars (MG) under the definition of Stabler (2011) and including the "Shortest Move" or Minimal Link Condition (MLC) on movement, are weakly equivalent to Linear Context-Free Rewriting Systems (LCFRS), or, equivalently, to Multiple Context-Free Grammars (MCFG). At the time, it was conjectured that LCFRS/MCFG were Mildly Context Sensitive (MCS), from which it seemed to follow that MG+MLC was also MCS. However, it has subsequently been shown by Salvati (2011/2015) that the artificial language MIX, consisting of all permutations on the string $a^n b^n c^n$ is an MCFL.

Under the (informal) definition of Joshi et al. (1991), mild context-sensitivity explicitly excludes permutation-completeness. So, since MCFL includes MIX, a permutation-complete language by definition, MCFG is not MCS under that definition, and so neither is MG. On the other hand, under the formal definition of MCS given in Kallmeyer, 2010, 23–24, the MCS languages include MCFL. Salvati suggests that the languages characterized by the well-nested subset of MCFG, MCFG_{WN} might formally correspond to the set of MCS languages, and shows that TAG is weakly in MCFG_{WN}, while Kanazawa and Salvati (2012) show that MIX is *not* in MCFL_{WN}. However, the MCFGs to which Stabler's MG corresponds are known to be non-well-nested (Boston et al., 2010), (although Kanazawa et al., 2011 show that the addition of a further Specifier Island Contraint to MG restricts them to a subset of MCFG_{WN}).

By contrast, Kanazawa and Salvati (2012) also show that MIX is not a Tree Adjoining Language (TAL), and therefore not a Combinatory Categorial Language (CCL), so TAG and CCG remain mildly context sensitive in a stronger sense, without constraints, under all definitions.

In this rather confusing state of affairs, it therefore seems helpful to distinguish the latter, not merely as mildly context sensitive, but more narrowly as "Slightly Non-Context-Free" (SNCF).

7.2 Child language acquisition

The restriction of the CCG lexicon and combinatory rules to strict type-transparency between constituents and their logical forms means that it supports a practically computable model of child language acquisition via "semantic bootstrapping" (Pinker, 1979), or, more properly, learning given access to contextually accessible universal logical form.

The problem of child language acquisition then reduces to the problem of learning (a) a lexicon, and (b) a parsing model, for all rules consistent with the (noisy) language-specific data and the (ambiguously) contextually available meaning representations, of which those sound-meaning pairs belonging to the actual target grammar will be vastly more frequent than the spurious ones (Abend et al., 2017).

Interestingly, the learner of Abend et al. gives a superficial appearance of learning parameters. (For example, in the later stages of learning English, the probability mass assigned to the SVO category for an unseen transitive verb will be near 1, and all other categories will be near 0.) However, there is no learned parametric prior directly associating this semantic type with this category. Instead, the information is implicit in the probabilities assigned to the various instantiated syntactic rules in the grammar as a whole that are used in calculating the prior probabilities of alternative derivations.

8 Conclusion: Towards a combinatory minimalism

CCG is a theory that embodies in a very direct form the Minimalist insight that syntactic derivation is determined by need to create objects that are phonologically and semantically well-formed, and nothing else, reducing unbounded movement to contiguous composition and case to morpholexical type-raising.

Other bounded types of movement, including "raising" and "control" relations; "head movement" (Roberts, 2001), "scrambling" (Ross, 1967), and "sideward movement" (Nunes, 2001) are defined statically, in the lexicon, at the level of logical form, from which they are projected by syntactic derivation.

In view of recent invocations among Minimalists of both prosodic contiguity and type-raising, there seems to be a possibility of extending the Minimalist Program to cover the full range of movement, coordination, and prosodic structural phenomena under the following equation, subject to the combinatory projection principle (CPP), (12):

Minimalism = Categorial Grammar + Case + Composition

References

Abels, Klaus & Ad Neeleman. 2012. Linear asymmetries and the LCA. *Syntax* 15. 25–74.

Abend, Omri, Tom Kwiatkowski, Nathaniel Smith, Sharon Goldwater & Mark Steedman. 2017. Bootstrapping language acquisition. *Cognition* 164. 116–143.

Ades, Anthony & Mark Steedman. 1982. On the order of words. *Linguistics and Philosophy* 4. 517–558.

Adger, David. 2003. *Core Syntax: A Minimalist Approach*. Oxford: Oxford University Press.
Adger, David. 2013. *A Syntax of Substance*. MIT Press.
Ajdukiewicz, Kazimierz. 1935. Die syntaktische Konnexität. In Storrs McCall (ed.), *Polish Logic 1920–1939*, 207–231. Oxford: Oxford University Press. Trans. from *Studia Philosophica* 1. 1–27.
Bach, Emmon. 1976. An extension of classical Transformational Grammar. *Problems in Linguistic Metatheory: Proceedings of the 1976 Conference at Michigan State University*, 183–224. Lansing, MI: Michigan State University.
Baldridge, Jason. 2002. Lexically Specified Derivational Control in Combinatory Categorial Grammar. Ph.D. thesis, University of Edinburgh.
Bar-Hillel, Yehoshua. 1953. A quasi-arithmetical notation for syntactic description. *Language* 29. 47–58.
Barss, Andrew & Howard Lasnik. 1986. A note on anaphora and double objects. *Linguistic Inquiry* 17. 347–354.
Boas, Hans & Ivan A. Sag. 2012. *Sign-based Construction Grammar*. Stanford, CA: CSLI Publications.
Boston, Marisa Ferrara, John Hale & Marco Kuhlmann. 2010. Dependency structures derived from Minimalist Grammars. *Proceedings of the Conference on the Mathematics of Language*, 1–12. Springer.
Bozşahin, Cem. 2012. *Combinatory Linguistics*. Berlin: de Gruyter.
Bresnan, Joan (ed.). 1982. *The Mental Representation of Grammatical Relations*. Cambridge, MA: MIT Press.
Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
Chomsky, Noam. 1972. *Studies on Semantics in Generative Grammar*. Mouton: The Hague.
Chomsky, Noam. 1981. *Lectures on Government and Binding*. Dordrecht: Foris.
Chomsky, Noam. 1995a. Bare phrase structure. In Gert Webelhuth (ed.), *Government and Binding Theory and the Minimalist Program*, 383–439. Oxford: Blackwell.
Chomsky, Noam. 1995b. *The Minimalist Program*. Cambridge, MA: MIT Press.
Chomsky, Noam. 2000. Minimalist inquiries: The framework. In Roger Martin, David Michaels & Juan Uriagereka (eds.), *Step by Step: Essays in Minimalist Syntax in Honor of Howard Lasnik*, 89–155. Cambridge, MA: MIT Press.
Chomsky, Noam. 2001. Derivation by phase. In Michael Kenstowicz (ed.), *Ken Hale: a Life in Language*, 1–52. Cambridge, MA: MIT Press.
Chomsky, Noam & Howard Lasnik. 1977. Filters and control. *Linguistic Inquiry* 8. 425–504.
Cinque, Guglielmo. 2005. Deriving Greenberg's universal 20 and its exceptions. *Linguistic Inquiry* 36. 315–332.
Cooper, Robin. 1983. *Quantification and Syntactic Theory*. Dordrecht: Reidel.
Cormack, Annabel & Neil Smith. 2002. Compositionality, copy theory, and control. *Working Papers in Linguistics* 14. 355–373.
Croft, William. 2000. *Explaining Language Change: An Evolutionary Approach*. Longman.
Culicover, Peter W. & Ray Jackendoff. 2005. *Simpler Syntax*. Oxford: Oxford University Press.
Dowty, David R. 1988. Type-raising, functional composition, and nonconstituent coordination. In Richard Oehrle, Bach Emmon & Deirdre Wheeler (eds.), *Categorial Grammars and Natural Language Structures*, 153–198. Dordrecht: Reidel.
Dowty, David R. 1993. variable-free' syntax, variable-binding syntax, the natural deduction Lambek calculus, and the crossover constraint. *Proceedings of the Eleventh West Coast Conference on Formal Linguistics, 1992*, 161–176., Stanford, CA: SLA.
Gazdar, Gerald. 1981. Unbounded dependencies and coordinate structure. *Linguistic Inquiry* 12. 155–184.
Gazdar, Gerald, Ewan Klein, Geoffrey Pullum & Ivan A. Sag. 1985. *Generalized Phrase Structure Grammar*. Oxford: Blackwell.

Ginsburg, Seymour & Sheila Greibach. 1967. Abstract families of languages. *Proceedings of Eighth Annual Symposium on Switching and Automata Theory, 1967, SWAT 1967*, 128–139. IEEE.

Greenberg, Joseph. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In Joseph Greenberg (ed.), *Universals of Language*, 73–113. Cambridge, MA: MIT Press.

Hays, David. 1964. Dependency theory: a formalism and some observations. *Language* 40. 511–525.

Hepple, Mark. 1990. The Grammar and Processing of Order and Dependency: A Categorial Approach. Ph.D. thesis, University of Edinburgh.

Higgins, F. Roger. 1973. The Pseudocleft Construction in English. Ph.D. thesis, MIT, Cambridge, MA.

Hudson, Richard. 1984. *Word Grammar*. Oxford: Blackwell.

Huybregts, Riny. 1984. The weak inadequacy of context-free phrase-structure grammars. In Ger de Haan, Mieke Trommelen & Wim Zonneveld (eds.), *Van Periferie naar Kern*, 81–99. Dordrecht: Foris.

Jackendoff, Ray. 1972. *Semantic Interpretation in Generative Grammar*. Cambridge, MA: MIT Press.

Jackendoff, Ray. 1975. Morphological and semantic regularities in the lexicon. *Language* 50. 639–671.

Jackendoff, Ray. 1990. On Larson's treatment of the double object construction. *Linguistic inquiry* 21. 427–456.

Jackendoff, Ray. 1997. *The Architecture of the Language Faculty* (Linguistic Inquiry Monograph 28). Cambridge, MA: MIT Press.

Jackendoff, Ray. 2002. *Foundations of Language*. Oxford: Oxford University Press.

Jacobson, Pauline. 1999. Towards a variable-free semantics. *Linguistics and Philosophy* 22. 117–184.

Joshi, Aravind, K. 1988. Tree-adjoining grammars. In David R. Dowty, Lauri Karttunen & Arnold Zwicky (eds.), *Natural Language Parsing*, 206–250. Cambridge: Cambridge University Press.

Joshi, Aravind, K., K. Vijay-Shanker & David Weir. 1991. The convergence of mildly context-sensitive formalisms. In Peter Sells, Stuart Shieber & Thomas Wasow (eds.), *Processing of Linguistic Structure*, 31–81. Cambridge, MA: MIT Press.

Kallmeyer, Laura. 2010. *Parsing Beyond Context-Free Grammars*. Springer.

Kanazawa, Makoto, Jens Michaelis, Sylvain Salvati & Ryo Yoshinaka. 2011. Well-nestedness properly subsumes strict derivational Minimalism. *International Conference on Logical Aspects of Computational Linguistics*, 112–128. Springer.

Kanazawa, Makoto & Sylvain Salvati. 2012. MIX is not a Tree-Adjoining Language. *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 666–674. Jeju Island, Korea: ACL.

Kay, Martin. 1984. Functional unification grammar: A framework for machine translation. *Proceedings of the 22nd Annual Conference of the Association for Computational Linguistics, Stanford, CA*, 75–78. ACL Anthology.

Kuhlmann, Marco, Alexander Koller & Giorgio Satta. 2015. Lexicalization and generative power in CCG. *Computational Linguistics* 41. 187–219.

Labov, William. 1994. *Principles of Linguistic Change. Vol. 1: Internal Features*. Oxford: Blackwell.

Lambek, Joachim. 1958. The mathematics of sentence structure. *American Mathematical Monthly* 65. 154–170.

Larson, Richard. 1988. On the double object construction. *Linguistic Inquiry* 19. 335–392.

Larson, Richard. 1990. Double objects revisited: Reply to Jackendoff. *Linguistic Inquiry* 21. 589–632.

Lewis, Mike, Kenton Lee & Luke Zettlemoyer. 2016. LSTM CCG parsing. *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*, 221–231. San Diego CA: ACL.

Michaelis, Jens. 1998. Derivational minimalism is mildly context-sensitive. In Michael Moortgat (ed.), *Logical Aspects of Computational Linguistics (LACL'98)* (Lecture Notes in Artificial Intelligence 2014), 179–198. Berlin: Springer.

Michaelis, Jens. 2001. Transforming linear context-free rewriting systems into minimalist grammars. In Phillipe de Groote, Glyn Morrill & Christian Retoré (eds.), *Logical Aspects of Computational Linguistics (LACL'01)*, (Lecture Notes in Computer Science 2099), 228–244. Berlin: Springer.

Montague, Richard. 1973. The proper treatment of quantification in ordinary English. In Jaakko Hintikka, Julius Moravcsik & Patrick Suppes (eds.), *Approaches to Natural Language: Proceedings of the 1970 Stanford Workshop on Grammar and Semantics*, 221–242. Dordrecht: Reidel, reprinted as Thomason, 1974, 247–279.

Moortgat, Michael. 1988. Categorial Investigations. Ph.D. thesis, Universiteit van Amsterdam. published by Foris, Dordrecht, 1989.

Morrill, Glyn. 1994. *Type-Logical Grammar*. Dordrecht: Kluwer.

Nchare, Abdoulaye Laziz. 2012. The Grammar of Shupamem. Ph.D. thesis, New York University.

Newmeyer, Frederick J. 2003. Grammar is grammar and usage is usage. *Language* 79. 682–707.

Nunes, Jairo. 2001. Sideward movement. *Linguistic Inquiry* 32. 303–344.

Partee, Barbara & Mats Rooth. 1983. Generalised conjunction and type ambiguity. In Rainer Bäuerle, Christoph Schwarze & Arnim von Stechow (eds.), *Meaning, Use, and Interpretation of Language*, 361–383. Berlin: de Gruyter.

Pesetsky, David. 1995. *Zero Syntax*. Cambridge, MA: MIT Press.

Pesetsky, David. 2017. Complementizer-*trace* effects. In Martin Everaert & Henk van Riemsdijk (eds.), *Blackwell Companion to Syntax*. New York: Wiley.

Pierrehumbert, Janet & Julia Hirschberg. 1990. The meaning of intonational contours in the interpretation of discourse. In Philip Cohen, Jerry Morgan & Martha Pollack (eds.), *Intentions in Communication*, 271–312. Cambridge, MA: MIT Press.

Pinker, Steven. 1979. Formal models of language learning. *Cognition* 7. 217–283.

Polinsky, Maria & Eric Potsdam. 2000. Backward control in Tsez. *19th West Coast Conference on Formal Linguistics*, 357–370.

Pollard, Carl. 1984. Generalized Context Free Grammars, Head Grammars, and Natural Languages. Ph.D. thesis, Stanford University.

Pollard, Carl & Ivan A. Sag. 1987. *Information-Based Syntax and Semantics, Vol. 1*, Stanford, CA: CSLI Publications.

Pollard, Carl & Ivan A. Sag. 1992. Anaphors in English and the scope of binding theory. *Linguistic Inquiry* 23. 261–303.

Richards, Norvin. 2016. *Contiguity Theory* (Linguistic Inquiry Monograph 73). Cambridge, MA: MIT Press.

Rizzi, Luigi. 1990. On the anaphor-agreement effect. *Rivista di Linguistica* 2. 27–42.

Roberts, Ian. 2001. Head movement. In Mark Baltin & Chris Collins (eds.), *Handbook of Contemporary Syntactic Theory*, 113–147. Oxford: Blackwell.

Rooth, Mats. 2016. Alternative semantics. In Caroline Féry & Shinichiro Ishihara (eds.), *The Oxford Handbook of Information Structure*, 19–40. Oxford: Oxford University Press.

Ross, John Robert. 1967. Constraints on Variables in Syntax. Ph.D. thesis, MIT, published as Ross, 1986.

Ross, John Robert. 1970. Gapping and the order of constituents. In Manfred Bierwisch & Karl Heidolph (eds.), *Progress in Linguistics*, 249–259. The Hague: Mouton.

Ross, John Robert. 1986. *Infinite Syntax!*. Norton, NJ: Ablex.

Sadock, Jerrold. 1991. *Autolexical Syntax: A Theory of Parallel Grammatical Representations*. University of Chicago Press.

Salvati, Sylvain. 2011/2015. Mix is a 2-mcfl and the word problem in $\mathbb{Z}^2$ is captured by the io and the oi hierarchies. *Journal of Computer and System Sciences* 81. 1252–1277, circulated in 2011.

Selkirk, Elisabeth. 1990. On the nature of prosodic constituency. *Papers in Laboratory Phonology* 1. 179–200.

Selkirk, Elisabeth. 2011. The syntax-phonology interface. In John Goldsmith, Jason Riggle & Alan Yu (eds.), *The Handbook of Phonological Theory*, 2nd edn., 435–483, Blackwell.

Shieber, Stuart. 1985. Evidence against the context-freeness of natural language. *Linguistics and Philosophy* 8. 333–343.

Stabler, Edward. 2011. Computational perspectives on minimalism. In Cedric Boeckx (ed.), *Oxford Handbook of Linguistic Minimalism*, 617–641. Oxford University Press.

Steedman, Mark. 1985. Dependency and coordination in the grammar of Dutch and English. *Language* 61. 523–568.

Steedman, Mark. 1987. Combinatory grammars and parasitic gaps. *Natural Language and Linguistic Theory* 5. 403–439.

Steedman, Mark. 1990. Gapping as constituent coordination. *Linguistics and Philosophy* 13. 207–263.

Steedman, Mark. 1996. *Surface Structure and Interpretation* (Linguistic Inquiry Monograph 30). Cambridge, MA: MIT Press.

Steedman, Mark. 2000. *The Syntactic Process*. Cambridge, MA: MIT Press.

Steedman, Mark. 2002. Plans, affordances, and combinatory grammar. *Linguistics and Philosophy* 25. 723–753.

Steedman, Mark. 2012. *Taking Scope: The Natural Semantics of Quantifiers*. Cambridge, MA: MIT Press.

Steedman, Mark. 2014. The surface-compositional semantics of English intonation. *Language* 90. 2–57.

Steedman, Mark. 2017. The emergence of language. *Mind and Language* 32. 579–590.

Svenonius, Peter. 2007. 1…3–2. In Gillian Ramchand & Charles Reiss (eds.), *The Oxford Handbook of Linguistic Interfaces*, 239–287. Oxford University Press.

Szabolcsi, Anna. 1983. *ECP in categorial grammar*. Ms., Max Planck Institute, Nijmegen.

Szabolcsi, Anna. 1989. Bound variables in syntax: Are there any? In Renate Bartsch, Johan van Benthem & Peter van Emde Boas (eds.), *Semantics and Contextual Expression*, 295–318. Dordrecht: Foris.

Thomason, Richmond (ed.). 1974. *Formal Philosophy: Papers of Richard Montague*. New Haven, CT: Yale University Press.

Van Valin, Robert. 1993. A synopsis of role and reference grammar. In Robert Van Valin (ed.), *Advances in Role and Reference Grammar*, 1–164. Amsterdam: John Benjamins.

Vergnaud, Jean Roger. 1977/2006. Letter to Noam Chomsky and Howard Lasnik on “Filters and Control, ” April 17, 1977. In Robert Freidin & Howard Lasnik (eds.), *Syntax: Critical Concepts in Linguistics*, 21–34. Oxford: Blackwell.

Vijay-Shanker, K. & David Weir. 1994. The equivalence of four extensions of context-free grammar. *Mathematical Systems Theory* 27. 511–546.

Williams, Edwin. 2003. *Representation Theory*. MIT Press.

Woods, William. 1970. Transition network grammars for natural language analysis. *Communications of the Association for Computing Machinery* 18. 264–274.

Woolford, Ellen. 1999. More on the anaphor agreement effect. *Linguistic Inquiry* 30. 257–287.

Zribi-Hertz, Anne. 1989. Anaphor binding and narrative point of view: English reflexive pronouns in sentence and discourse. *Language* 65. 695–727.

Part II: **Metatheoretical foundations**

Philip Carr

15 Syntactic knowledge and intersubjectivity

Abstract: I argue against Chomsky's claim that humans are born with innate syntactic knowledge. The autonomous study of syntactic knowledge is the study of social conventions, as conceived of by Itkonen (1978, this volume), and can be viewed as objective knowledge, in Popper's (1972) sense. Social conventions are not the object of naturalistic enquiry, since the object of such enquiry, by definition, contains no conventions. Autonomous generative syntax is not naturalistic science. However, the internalisations of those social conventions falls within naturalistic enquiry, conceived of as the study of the mind as part of the physical world, as does the study of the neural mechanisms underpinning the acquisition of syntactic knowledge.

1 Introduction: The epistemological and ontological bases of generative grammar

What is the significance of metatheoretical reflections for analytic/descriptive work in generative linguistics? I argue that one's metatheoretical assumptions underpin the kinds of descriptive work undertaken by linguists, both generative and non-generative. For instance, the work done by Smith (1973) on phonological acquisition is based on Chomskyan assumptions concerning philosophical positions such as Chomskyan naturalism, innateness, scientific realism and autonomism (see below on all of these). Work done by Vihman (1996 and elsewhere) on the acquisition of phonology is *not* based on autonomism or Chomskyan innateness, so the nature of the descriptive research conducted by these two scholars is distinct. For instance, Smith, given his Chomskyan assumptions, argues that the child has adult-like phonological representations, whereas Vihman argues that there are transitional phonological representations in the child's mind which are not adult-like. The child's deviations from adult-like pronunciations are taken to be performance errors by Smith, but not by Vihman, who takes them to reflect to structure of the child's evolving phonological system.

These distinct metatheoretical approaches have implications for generative and non-generative approaches to language change, both phonological and syntactic. A long-standing view in generative linguistics (Lightfoot, 1979) is that language change is rooted in language acquisition, a view opposed by Vihman, and by sociolinguists, who hold that the temporal locus of language change is in adolescence. For Vihman, the child will pass through transitional grammars, but will come to have a mentally constituted grammar which is adult-like.

In Carr (1990), I presented two metatheoretical arguments concerning generative linguistics: (a) that generative linguistics could be taken to be scientific in nature,

https://doi.org/10.1515/9783110540253-015

given Popper's falsificationist philosophy of science, and (b) that generative linguistics should be construed, *pace* Chomsky, in a non-internalistic manner, under which linguistic knowledge is to be conceived of as objective knowledge, in Popper's (1972) sense, whereby objective knowledge is intersubjective, rather than mind-internal, in nature, as it is for Chomsky. In adopting Popper's falsificationism, I did not assume that falsifying evidence came in the form of certainly true observation statements. This is because Popper himself argued that there are no such things: this was part of critique of logical positivism, and is at the heart of Popperian fallibilism: our observations are always fallible, as are our theories, since, as Popper points out, observation is theory-dependent, and our theories are fallible.

In that work, I rejected Itkonen's (1978, 1983, this volume) hermeneutic conception of linguistic knowledge as knowledge of socially-established norms (conventions). I now accept that conception.

I also defended, as Itkonen does, the role of intuitive grammaticality judgements in linguistics, a position which remains controversial today, particularly with the advent of much corpus-based work on language.

Here, I continue to defend the use of grammaticality judgements in generative linguistics, but I now adopt a conception of intersubjectivity based on the work of Tomasello, 2001).

I also now argue that generative linguistics is not, *pace* Chomsky (2002), natural science. I take natural science to subsume scientific theories of the natural world, principally physics, chemistry and biology. The case of biology is important here, since Chomsky adopts a naturalistic interpretation of generative linguistics which takes the object of inquiry to be biological in nature.

While generative linguistics adopts hypotheses which are falsifiable, with the evidence coming from intuitive grammaticality judgements, there is, I now argue, no way of choosing a single paradigm in generative linguistics which can be said to be 'true'. A crucial distinction between generativist linguistics and the natural sciences is this: in the natural sciences (subsuming biology), while there are competing paradigms over time, natural scientists, such as biologists and physicists, typically come to alight on a single paradigm. Examples are gravitational theory, the heliocentric theory in astronomy, electro-magnetic theory and, in biochemistry, the double-helix theory of DNA. I refer to this as *convergence*: theories in the natural sciences, subsuming biological theories, may converge on a single model. No such acceptance of a single theory can be had in generative grammar: thus the plethora of grammatical theories, reflected in this volume. The reason for this is that, as Itkonen argues, generative linguistics studies knowledge of social conventions, which are not, I argue, biological in nature. In making this argument, I adopt a distinction between nature and social conventions, which I seek to defend here, against the views of sociobiologists. Evidence that theories in the 'special sciences' (Fodor, 1981), such as economics, do not allow for convergence is the fact that no convergence has ever been attained in economic theory: there are many economic theories, and economists are well-known to be in continual disagreement.

I argue that Sampson (2005), who adopts Popper's philosophy of science as a means of arguing that generative linguistics is not empirical science, is mistaken in assuming that corpus linguistics constitutes empirical science: corpus linguistics depends, as Itkonen (this volume) says, on intuitive judgements, which Sampson rejects as a source of evidence in what he takes to be 'empirical' linguistics.

In relation to this, I argue that Sampson's conception of child language acquisition is flawed: it is based on the idea, which can be encountered in the history of generative linguistics, in the work of Fodor (1983) and that of Gopnik (2000), that the child is 'a little scientist', formulating falsifiable hypotheses about the ambient language, and testing them against the empirical evidence. As an alternative to this, I adopt Tomasello's (2001) view that the child is *not* formulating hypotheses, but is undergoing socialization. Crucial to this process is the emergence of entrance into social conventions, among which are the rules of the ambient language(s).

Section 2 contains discussion of Chomskyan naturalism. In section 3, I consider an alternative version of naturalism, as adopted by Burton-Roberts. I then proceed, in section 4, to the idea of syntactic knowledge as knowledge of social conventions, focusing on the work of Itkonen and Tomasello. In section 5, I discuss the acquisition of syntactic knowledge as knowledge of social conventions, and in section 6, I present a re-interpretation of Popper's World 2, based on Burton-Roberts' notion 'I-physical'.

2 Generative syntax and epistemology: natural science, Chomsky's naturalism and knowledge without grounds

I begin with Chomsky's naturalism and his philosophy of science. Chomsky believes that generative linguistics is science. More precisely, he believes that it is a branch of the natural sciences, specifically biology. Chomsky argues that the object of linguistic enquiry is biological in nature: he postulates an innate language module which is shared across the members of our species. It is said to be an expression of the genes, part of our genetic endowment. Since each individual member of our species is said to possess this postulated innate endowment, linguistic knowledge in general, and syntactic knowledge in particular, is conceived of as I-language, whereby 'I' stands for both individual and internal.

Chomsky adopts scientific realism, a position adopted in the philosophy of science, in opposition to instrumentalism and positivism, whereby it is claimed that there are aspects of the physical world which are not, *in principle*, observable, but are nonetheless real. I have always adhered to this position: it is central to Carr (1990). Examples are gravitational force and electro-magnetic forces. Chomsky's point is that, while one can observe the effects of these, one cannot, in principle, observe those forces. It is not that we currently lack the instruments to observe such forces: they will never be observed, since they are unobservable. One can observe apples falling from trees, which is evidence for gravitational force, but one cannot observe that force, which is real. We can observe meter readings of electrical forces, but that does not constitute the observation of the underlying forces. Thus Chomsky's point that the study of physics is not the science of meter readings.

This has been central to Chomsky's point about idealization in science, and thus, for Chomsky, in generative linguistics. The distinction between competence and performance is rooted in the doctrine of scientific realism: if generative linguistics is natural science, then scientific investigation in generative linguistics follows the practice of idealization. An example in physics is the principle of the conservation of energy: if one discounts the effect of friction, one can argue that a trolley running down a slope conserves energy. Friction creates heat energy, so that trollies will lose energy in the form of heat. But if it weren't for friction, there would be no energy loss. Similarly, if it weren't for air resistance, a kilo of feathers dropped from the same height as a kilo of lead would hit the ground at the same time. Chomsky argues that this kind of idealization, the factoring out of phenomena which conceal the underlying reality, is standard practice in the natural sciences.

Chomsky extended this to the study of mentally-constituted grammars: if we factor out phenomena such as the age, social class and sex of the native speakers of a language, we can arrive at their shared underlying competence, which is unobservable in principle, since the mind is, in principle, unobservable, but is nonetheless real. Thus the concept of the ideal speaker-hearer in a homogeneous speech community. It was never a coherent counter-argument that no such speakers exist, and those who made this counter-argument failed to understand Chomsky's point. Chomsky was never claiming that such speakers existed: he was postulating the competence of native speakers, with factors such as age, sex, social class and context of utterance factored out.

A consequence of Chomsky's radical internalism is his claim that syntactic knowledge is 'knowledge without grounds', the view that syntactic knowledge is not grounded in the experience of the mind-external world, or in more general aspects of cognition, such as the capacity to form inductive and analogical generalisations. This follows from the innateness hypothesis: if syntactic knowledge is innate, then, by definition, it is not internalized from the environment. Furthermore, since the putative language module is species-specific and specifically linguistic, any cognitive capacity which is not restricted to language is, by definition, not part of the innate language module. Thus the exclusion of the cognitive capacities for the forming of inductive and analogical generalisations, which are domain-general, not domain-specific.

Chomsky's (1986) claim that syntactic knowledge is knowledge without grounds reflects his opposition to Empiricism: linguistic knowledge is not, for Chomsky, derived (or, more weakly, is not *largely* derived) from experience of the mind-external world. Rather, humans are said to be born with innate linguistic knowledge, in the form of an innate, species-specific language module, which Chomsky used to refer to as Universal Grammar (UG), and which is now referred to as the human language faculty, also known as the Faculty of Language (FoL), which is innate, radically mind-internal and individual. By 'radically mind-internal', I mean not internal*ised* from the environment, despite Chomsky's inappropriate (given his assumptions) use of that very term. *Weak* internalism, unlike *radical* internalism, allows that linguistic knowledge may be internalised from the environment.

By 'individual', Chomsky means that the FoL is a biological property of each individual member of our species: Chomsky rejects any public/social/cultural conception of language, and thus the idea that linguistic knowledge is knowledge of social conventions. He claims that the innermost part of the FoL generates linguistic expressions which have phonological and semantic properties which allow them to be fed into the two interface performance systems: Phonetic Form (PF) and Logical Form (LF). Chomsky has consistently upheld the centrality of syntax thesis, whereby phonology and semantics fall outside of core linguistic knowledge. This was previously expressed as the notion that semantics *interprets* syntactic structure, and is now expressed as the idea that there is a semantic interpretation interface (LF). Both of these modules are currently described as performance modules, which act as interfaces in relation to the 'innermost cognitive system' (ref), namely the syntax, conceived of as sets of computations over linguistic representations (see below on different conceptions of 'representation').

The principal argument in favour of a radically internal innate language module is the argument from the Poverty of the Stimulus (POS). Chomsky claims that the stimulus (the utterances heard by the child) is impoverished, given the hesitations, unfinished utterances of sentences and switches from one structure to another which are typical of ordinary everyday speech. His argument is that such impoverished input is insufficient to support language acquisition, and that the child must therefore have help, in the form of the innate language module, which contains innate knowledge, such as knowledge of recursion.

Sampson (2005) has shown, using spoken data from the British National Corpus, that most everyday speech consists of utterances of perfectly well-formed sentences: the stimulus is not impoverished to any great extent.

I would go further than Sampson, and suggest that even fragmented speech contains well-formedness. Consider this kind of phenomenon in everyday speech:

> 'What I mean is, well, the thing is, it's just that, well, you know, mummy is having another baby.'

This kind of input is not entirely ill-formed: each of the fragments in this fragmented speech contains structure which conforms to the grammatical rules of Standard English. Although 'What I mean is' is incomplete, its internal structure is well-formed.

The utterance 'What I mean is that mummy is having another baby' is the utterance of a well-formed sentence, so 'What I mean is' is a well-formed fragment, from which the regularities of the syntax of Standard English can be internalised.

Those who support the POS have made much of the sequencing of auxiliaries in Standard English. The syntactic constraint on that sequencing is this: modal, perfective, progressive, passive, so that 'This place might have been being built for ten years now' is well-formed. The supporters of the POS argue that the chances of hearing an utterance of a sentence with all four types of auxiliary are vanishingly small, so the input alone cannot allow for the acquisition of the constraint.

I have two responses to this claim.

Firstly, the child does not need to be exposed to utterances with all four auxiliaries in order to extract the generalization. The child hears utterances like 'This place might have taken ages to build', and many parallel structures with a modal followed by perfective 'have', then the child can internalize the fact that modals must precede perfective 'have'. The child also hears sequences such as 'The mouse has been eating the cheese', and can therefore internalise the generalization that the perfective auxiliary precedes the progressive auxiliary. And with utterances such as 'Granny is being kept in hospital', the child can internalize the generalization that the progressive auxiliary precedes the passive auxiliary. In order the extract the sequencing generalization from the input, the child does not need to be exposed to utterances of sentences with all four types of auxiliary.

Secondly, why would an innate language module provide the child with knowledge of a constraint that is specific to English? By its very nature, the putative language module contains knowledge of universal properties of human language, not knowledge of syntactic constraints in specific languages.

In stark contrast to the claim that UG could play a role in the acquisition of auxiliary sequencing in English, the idea that knowledge of recursion could reside in an innate language module is much more plausible. If recursion is a *robustly* universal property of human language, that is, a property of *every* human language, then that would appear to be a fact about the nature of human language which could be innately endowed.

Sampson's (1997, 66–67) response to this innateness claim is entirely coherent, I suggest. Sampson argues that a child can learn, from the input, a simple rule of English whereby a preposition such as *in* can form a unit with a noun phrase (such as *the kitchen*, having previously learned that definite articles can form units with nouns) to form a prepositional phrase, such as *in the kitchen*, and can also learn that nouns can be post-modified by prepositional phrases, thus resulting in the knowledge that noun phrases can be post-modified by prepositional phrases, so that the child can come to understand and utter structures such as *the knife in the kitchen*. Knowledge of recursion can thus be learned, piecemeal, and need not be innate.

Another attempt to counter Chomsky's claim about recursion is offered by Everett et al. (2005), who claims to have found a language (Pirahã) which lacks recursion.

The response of the Chomskyans has been varied. One response is to say that Everett is mistaken about Pirahã. The other is to say that, even if there *are* languages which lack recursion, this does not undermine Chomsky's claim.

I argue that the latter response cannot be sustained: if there is an innate language module, then its core feature (arguably, the ONLY universal feature left to Chomsky) *must* be attested in *all* human languages.

The former response to Everett is sustainable, however. I argue that Everett's claim is based on his mode of glossing his Pirahã examples. Consider the following:

That neighbour I know is lying.

(Here, 'I know' is a restrictive relative clause lacking a relativiser, but nonetheless constitutes a clause within a clause; thus the sentence exhibits recursion.)

Let us consider the following thought experiment. The Roman invasion of the British Isles was successful. The Viking invasion and the Norman invasion never happened. English died out, except in a tiny community living in a remote woodland area, cut off from all contact with the rest of the population of the British Isles, which spoke a Latin-derived language. This community is discovered in the late 20th century by a linguist (Dan Everett) who transcribes some of their structures like this:

That neighbour
I know
is lying.

The field linguist claims that this community speak a little-known language which its speakers call English, which contains no recursion. This would be a trick of transcription, and that, I suggest, is what Everett has done.

If it is a fact that recursion is a *robustly* universal property of human language, present in every human language, does that fact support the claim that there is an innate language module, unique to human beings, and specific to language, thus species-specific and specifically linguistic? Perhaps not.

Consider two claims made by Sampson (see above). Firstly, hierarchical structure is a property of complex systems, whether linguistic or not. If this is correct (I argue that it is), then recursion is not specific to human language, and thus cannot be a part of a species-specific, specifically linguistic innate module. Secondly, recursion is *learned* by the child. As Sampson says, the English-acquiring child learns nouns and verbs, and then comes to learn articles, so that, after having acquired nouns such as 'knife' and 'table', the child masters noun phrases such as 'the knife' and 'the table'. With mastery of prepositional phrases, the child has constructions such as 'on the table'. With further syntactic development, the child can master structures such as 'the knife on the table', which is recursive, containing a noun phrase within a noun phrase. There is no need to postulate innate knowledge of recursion.

It is worth pointing out that Pinker's 1995 version of naturalism, widely taken to be a popular exposition of Chomskyan naturalism, is not consistent with Chomsky's views. Pinker argues that there is a language *instinct*. But this is at variance with Chomsky's views. Chomsky has consistently insisted that his conception of linguistic knowledge is not knowing how to do something (procedural knowledge and memory: Rose 1993), and is not knowing *that* (declarative and semantic knowledge and memory: Rose 1993). It is important to bear in mind that Chomsky's position is not only anti-behaviourist, but anti-behaviour*al*. Linguistic knowledge, for Chomsky, is not knowing how to engage in behavioural acts, and the innate language module does not constitute innate disposition to behave in certain ways. Instincts are, by definition, innate capacities to engage in certain kinds of behavior. For instance, members of spider species which are born with the capacity to spin webs (and thus do not have to *learn* how to do so) are born with an innate capacity to engage in that kind of behavior. This is an innate capacity, but it is entirely distinct in kind from Chomskyan innate linguistic knowledge. Innate capacities to behave in certain ways do not constitute Chomskyan knowledge, and Chomskyan innate knowledge is not an instinct to behave.

It might be suggested that language acquisition cannot be viewed as a philosophical issue, but must be viewed as an empirical matter to be investigated by, among others, psycholinguistics. That view, I suggest, is mistaken. Chomsky's notion of an innate language module is a philosophical notion: it falls within the domain of epistemology, the study of the nature of knowledge, in this case linguistic knowledge. Chomsky has claimed for decades that he is working with the philosophical tradition of Descartes. Whether he is justified in claiming that he is a Cartesian, the fact remains that his central claim is a philosophical one. The same is true of the work of the philosopher Jerry A. Fodor, who has made it clear he works in the field of the philosophy of cognitive science. Thus the subtitle of Fodor (1981).

3 Burton-Roberts' naturalism

Burton-Roberts (2011, henceforth B-R) argues that it is incoherent to postulate, as Chomsky does, mentally constituted linguistic representations which have both conceptual and phonological properties, and that, at the heart of Chomsky's double interface (LF and PF) assumption, there is inconsistency: Chomsky, B-R argues, has inherited a Saussurean assumption (the linguistic sign, with both phonological and semantic properties) which is not, B-R argues, consistent with Chomsky's purely biological conception of the FoL. For B-R, the purely conventional (and arbitrary) relation between phonological representations and concepts, which Saussure postulated, can have no place in a biological conception of language, since conventionality is not, and could not be, a biological property. That conventional relationship belongs in the domain of E-language, which is socio-political in nature, not biological.

B-R seeks to exclude from the FoL anything which falls within the domain of E-language, and this leads him to exclude sequencing of any sort from the FoL: sequencing occurs in mind-external events in space and time. This has the radical consequence of excluding from the FoL almost all of what has traditionally counted as syntax, where sequencing is central. This leads B-R to claim that the FoL is not distinct from the Fodorian Language of Thought (LoT, also known as Mentalese), a putative innate set of universal semantic primitives which are said to be the building blocks for all possible concepts. For B-R, the innate language module *is* the LoT, and the LoT is knowledge without grounds.

Given this radical position, the question arises as to what the relationship might be between E-languages as socio-cultural-political objects, on the one hand, and the LoT on the other. B-R claims that this relationship is one of *M-representation*, where 'M' stands for 'Magritte'. The appeal here is to Magritte's painting 'La Trahison des Image', with the words 'Ceci n'est pas une pipe' written under a visual representation of a pipe. The point is that a representation of a pipe is not a pipe, and representations of objects in the LoT are not objects in the LoT. For B-R, when we utter speech sounds, we are producing E-physical M-representations of entirely abstract objects (concepts, built out of innately-endowed semantic primitives) in the LoT. By 'E-physical' is meant physical and mind-external, as distinct from I-physical, meaning internal to a physical mind/brain. This notion of representation is the traditional two-place predicate notion, whereby X represents Y. In the case of the E-physical representations produced by speakers, the relation between the representation and what is represented is not iconic (as it is in the case of Magritte's painting): E-physical M-representations cannot in any way resemble the expressions generate by the LoT. E-languages are said by B-R to be *conventional* systems of E-physical representation. Those conventions are social in nature. This is B-R's Representational Hypothesis.

B-R's metatheoretical position does appear to take Chomskyan assumptions to their logical, very radical, conclusion. But it rests on the validity of the Fodorian notion of a Language of Thought, which in turn relies crucially on the idea that the child is formulating hypotheses and testing them. The logic of Fodor's position is impeccable: if the child is indeed formulating hypotheses about the ambient language, then the child must have at its disposal a conceptual language in which to formulate those hypotheses, and that language must be at least as rich as the ambient language. Thus the Language of Thought. Sampson has a problem here: given that he takes the child to be formulating hypotheses about the structure of the ambient language, he must answer the question 'In *what* are those hypotheses formulated?' This would have to be something like the Language of Thought, but Sampson, being an Empiricist, cannot adopt Fodor's innate LoT. And B-R's position is untenable if the child is not formulating hypotheses. Without hypothesis formulation, the LoT does not exist. While the logic of Fodor's argument is compelling, one of the premisses is flawed: the premiss that the child is engaged in hypothesis formulation.

Having rejected both Fodor's LoT and Chomsky's innate language module, I now turn to an alternative conception of syntactic knowledge conceived of as knowledge of social conventions.

4 Syntactic knowledge as knowledge of social conventions

4.1 Itkonen, autonomism and normativity

Itkonen (1978, 1983 and elsewhere) draws a tripartite distinction between (a) spatiotemporal *events*, which are not intentional, such as rain falling from the sky, (b) actions, such as acts of speaking, which *are* intentional, and are carried out by conscious agents, and (c) socially-constituted norms, also referred to by Itkonen as rules, which constitute the basis for our rule-governed actions. Although linguists at times refer to speech events, these are not events, for Itkonen: recordable speech is the result of intentional actions, based on social norms.

It follows from this that, *pace* Chomsky and many other linguistics specialists, linguistic theory is distinct in kind from physical theory; crucially, linguistic theory is not empirical according to Itkonen, since 'empirical' for him means 'testable against observed events'. Social norms, which, for Itkonen, include the norms which constitute a language, are not observable, and thus theorizing about such norms is not empirical. In Carr (1990), I followed Chomsky in taking 'empirical' to mean 'testable', whether against observable events, as in physics, or against intuitive grammaticality judgements, as in generative linguistics, but let us, for the moment, accept Itkonen's definition of 'empirical' for the sake of discussing his work (it is worth recalling that Itkonen's definition is much more widely accepted than Chomsky's).

The notion of 'correctness' is central to Itkonen's work: it is norms which define correctness. The example 'John speaks fluently French' in English, cited by Smith (1999) is, on Itkonen's terms, incorrect. Correctness has no place in physical theory, since the physical world contains no conventions. Note that, in making such a claim, I am assuming a distinction, adopted by Popper, Itkonen and Burton-Roberts, among others, between purely physical events and structures on the one hand, and the social actions performed by human beings on the other. This distinction is controversial: it may be argued that it constitutes a kind of Cartesian physical/non-physical dualism, but I argue that it is tenable.

The notion of causality is important for Itkonen in this context: linguistic theory, since it does not deal with physical events, makes no appeal to causality. If the sun shines on ice, causing it to melt, there is a causal relationship between the two. But linguistic theory is not theoretical investigation into physical events. Linguistic theory is, for Itkonen, *autonomous* with respect to the investigation of causal events: it is *grammatical* theory, referred to by Itkonen as autonomous linguistics (AL). This conception of autonomism is distinct from that of Anderson (2011): when Anderson objects to autonomism in generative linguistics, he is objecting to the notion of ungrounded syntactic and phonological knowledge. Itkonen's autonomism appeals to groundedness, but rejects the idea that linguistic inquiry can be subsumed under scientific enquiry of causally constituted events.

The distinction between what is social and what is individual is central to Itkonen's claims. Derwing (1973) claims that there is a paradox in Saussure's thinking with respect to this distinction. He claims that, for Saussure, "although language is presumably social, knowledge of it is acquired by means of individual intuition". Itkonen says that this is not a paradox, but a conceptual necessity: "There are compelling reasons to postulate the existence of (social) rules, including rules of language. Because rules are not concrete (spatiotemporal: PC) things, they cannot be observed, but only intuited (by individuals: PC)" (Itkonen, 1983, 8).

I would add the following regarding Derwing's claim. Firstly, Saussure had nothing to say about acquisition, and the claim that we acquire linguistic knowledge via acts of intuition is questionable: as Itkonen says, "the learning of norms of logic or of language is a process equally based on biological maturation and the observation of surrounding social behavior. But once a norm has been learned, it is known on the basis of intuition" (Itkonen, 1983, 58). Secondly, it is *langue* which is social (intersubjective, shared) for Saussure, and *parole* which is individual. Given that langue is social for Saussure, the learning of langue is bound to take place in a social context. Individual acts of speaking constitute *parole*.

For Itkonen, "A language is, first of all, constituted by rules, i. e. norms, which determine the correctness or incorrectness of actual or possible *utterances*" (my emphasis: PC) (Itkonen, 1983, 61). I emphasise the word 'utterance' here since Itkonen is appealing to the idea of the grammaticality of utterances, rather than sentences. This is at odds with an assumption made for many years in Chomsky's thinking, prior to the emergence of the E-language vs I-language distinction, and the relegating of the notion 'set of sentences' to E-language. For the entirety of the Chomskyan tradition prior to that move, starting with Chomsky (1957), Chomsky took a language to be constituted as a set of sentences, and sentences were taken to belong to competence, as opposed to utterances, which were taken to belong to the domain of performance. During that phase of Chomsky's thinking, grammaticality belonged in the domain of sentences and competence, whereby sentences were *generated* by the mentally-constituted grammar and utterances were *produced* by speakers.

There are parallelisms between Saussure's *langue vs parole* distinction and Chomsky's competence *vs* performance distinction, given that both Saussure and Chomsky distinguish between acts of uttering and the system which underlies those acts, but Saussure and Chomsky differ in their conceptions of what is individual: for Chomsky, competence (later, I-language) is individual, whereas, for Saussure, it is acts of parole that are individual. And langue, for Saussure, is social in nature, whereas, for Chomsky, competence is not.

Chomsky's competence *vs* performance distinction has always been controversial. The controversy has been fueled by Chomsky's use of the term 'competence', which has misled some of his critics. This is because he was using the word in a specialized way, distinct from the way it is normally used. Normally, when we speak of someone's competence in a given domain, we are speaking of degree of ability, such as the ability to play tennis or drive a car: we can be more or less competent tennis players, more or less competent car drivers. But Chomsky has always insisted that linguistic knowledge (competence) is not to be conceived of in terms of ability: it does not constitute knowing how to *do* something. Those who have appealed to the idea of communicative competence are not referring to the same notion of competence as Chomsky. Chomsky is, of course, at liberty to use everyday terms in a specialized manner, and it is incumbent upon his readers to understand such specialized usage, but it is, nonetheless, usage which induced misinterpretation.

4.2 Tomasello on socialisation

Let us consider the idea, proposed by Tomasello (2001), that the child is *not* formulating hypotheses about the structure of the ambient language(s). According to Tomasello, what constrains the child's interpretive possibilities when listening to child-directed speech are the social contexts in which those utterances are embedded. On this view, the acquisition of what Tomasello calls linguistic symbols is "a kind of by-product of social interactions with adults, in much the same way that children learn many other cultural conventions" (Tomasello, 2001, 135).

Under this approach, two things are emphasised. Firstly, "the structured social world into which the child is born, full of routines, social games, and other patterned cultural interactions" (Tomasello, 2001, 135). Secondly, "the child's capacities for tuning into and participating in the structured social world" (Tomasello, 2001, 135). In place of the hypothesis formulation and testing approach, he adopts an "experientialist and conceptualist" view of language, according to which "linguistic symbols are used by human beings to invite others to experience situations in particular ways" (Tomasello, 2001, 134). Under this approach, the child must be able to appreciate that other people have intentions, especially social intentions: the child must come to see others as intentional agents. For Tomasello, linguistic symbols are conventional in nature: the child is acquiring linguistic conventions.

On Tomasello's view "word learning is not a hypothesis-testing procedure needing to be constrained at all, but rather is a process of skill learning that builds upon a deep and pervasive understanding of other persons and their intentional actions" (Tomasello, 2001, 148–149). Tomasello argues that

> young infants, in the first year of life are born as social creatures, but do not, during most of those first twelve months, appreciate that other people have intentional relations towards the world. For instance, five-month old children do not follow the gaze of others to objects. Rather, they prefer to focus on their own face-to-face interactions with others. Equally, they do not imitate the behaviours of others on objects.

I should point out here that, in my view, the phenomenon of over-generalisation in child speech (as in 'Three sheeps comed') does not provide evidence for hypothesis formulation by the child: it reflects the utilization of the innately-endowed capacities for forming inductive and analogical generalization, neither of which capacities could form part of a Chomskyan language module, since they are domain-general capacities: the putative language module is species-specific, and specific to language. Other species possess these capacities: if a cat fins milk in the corner of the kitchen day after day, then seeks milk in that same place, the cat had engaged in inductive generalization. Are we to say that, because the cat forms such generalization, it has formulated a hypothesis? I suggest that cats do not have the cognitive capacity to formulate hypotheses. Under Fodorian assumptions, a cat would have to possess the LoT in order to formulate hypotheses, but that is impossible, since the LoT is, by definition, unique to humans.

Tomasello argues for the emergent capacity, during the second year of life, "to understand other people as intentional agents whose attention, emotion and behaviour towards objects may be actively followed into and shared" (Tomasello, 2001, 150). The child, he argues, is not simply mimicking adult body movements: the child is becoming aware of the intentions behind those movements. For instance, Meltzoff (1988) had fourteen-month-old children imitate adults who bent down at the waist, touched their heads on a panel, resulting in the switching on of a light. Tomasello's point is that this awkward piece of behaviour, much harder than simply touching the panel with one's hand, was imitated, showing that the child appreciated that the adult had the intention to switch on the light and that the adult chose one of a set of possible means towards that end. The child also understood that, if the child had the same intentions, (s)he could choose the same end to that means. The child thus comes to have intentions towards other people's intentions.

In lexical acquisition, Tomasello argues that the child in the second year of life, on hearing, say, 'A dog!' uttered by an adult "understands that the adult's intentional action is not directed towards the object, but rather the child, or more specifically, the child's intentions or attention." (Tomasello, 2001, 151). The child understands that the uttering of that sound sequence

> is produced with the intention to get the child to do something intentional (i. e. focus its attention on some aspect of an on-going event). This kind of cultural imitative learning requires role-reversal: the child must understand that, if the adult can get the child to focus on the object in question by uttering those sounds, then the child itself can use those sounds to get an adult to focus on such objects. (Tomasello, 2001, 151)

Tomasello distinguishes ritualisation from the acquisition of linguistic conventions via cultural imitation. An example of ritualized behaviour is the common tendency for pre-linguistic children to raise their arms when they want to be picked up. He argues that there is no evidence that such behaviour emerges via a process of cultural imitation. Rather, he argues, such gestures are ritualised versions of the movements made by an infant as it tries to pull its way up into an adult's arms. Such gestures do not, according to Tomasello, 'stand for' anything: they are not symbolic. He also argues that chimps in their natural habitats have ritualized behaviours, but do not engage in cultural imitation: they seem not to take their fellow chimps to be intention agents. For Tomasello, this is why chimps do not form linguistic conventions.

Tomasello's conception of linguistic knowledge as knowledge of social conventions is compatible, I suggest, with Itkonen's approach to linguistic knowledge: since Itkonen allows that the child learns the social conventions in question, we can take Tomasello's work to constitute an empirical theory of that learning. This does not exclude mind-internal representations of the syntactic and phonological patterns which are created when we follow those conventions. As Burton-Roberts points out, there are *systems* of conventional representation, and it is those systems that can be taken to be the object of grammatical enquiry. And, as Itkonen points out, the learning of linguistic norms is based on biological maturation.

I do not claim to have presented arguments against the LoT here: I have simply pointed to Tomasello's alternative conception of language acquisition.

5 On the acquisition of grounded syntactic knowledge

Since I adopt Tomasello's conception of syntactic rules, constructions and constraints as social conventions (Itkonen's norms), and the claim by B-R that conventionality has no place in nature, it follows that generative syntax, conceived of as the study of social conventions, is not natural science (I leave aside the status of the social sciences here).

To argue that generative syntax is not a branch of the natural sciences is not to argue that there can be no natural scientific study of the neural mechanisms used to process and store knowledge of social conventions in the mind. But it is an error, I suggest, to reduce social conventions to brain states and processes. Social conventions are intersubjective in nature: they are not brain states or processes.

One can reject the Chomskyan idea of an innate language module while accepting that there are innate cognitive capacities which are brought to bear on the acquisition of knowledge of social conventions. As Chomsky has stated, the question is not whether we possess innate cognitive capacities, but, rather, what they might be. I argue here that they do not include an innate language module, and that generative syntax can be decoupled from Chomsky's innateness hypothesis.

I now present a brief overview of some of the work on relevant brain states and mechanisms which figure in child language acquisition.

An interesting approach to modularity of mind, distinct from that of Fodor, can be found in the work of Karmiloff-Smith (2001 and elsewhere; henceforth K-S). She distinguishes between domain-*relevant* and domain-*specific* cognitive mechanisms, and argues that the former may develop into the latter:

> Unlike the domain-general theorist, this position does not argue for domain-general mechanisms simply applied across all domains. Rather, it suggests that biological constraints on the developing brain might have produced a number of mechanisms that do not start out as strictly domain-specific, that is, dedicated to the exclusive processing of one and only one kind of input. Instead, a mechanism starts out as somewhat more relevant to one kind of input over others, but it is useable – albeit in a less efficient way – for other types of processing too. This allows for compensatory processing and makes for development channeled, but far less pre-determined, than the nativist view. Once a domain-relevant mechanism is repeatedly used to process a certain type of input, it becomes domain-specific as a result of its developmental history. (Karmiloff-Smith, 2001, reprinted 2001: 323–333)

A case in point, discussed by K-S, is the recognition of familiar faces by people suffering from Williams Syndrome, who test like normals for this kind of recognition. For pathologically normal people, face recognition relies on an innate endowment which allows new-born infants to recognize face-like visual stimuli. But the input then plays a major role, as the infant is repeatedly exposed to familiar faces. Familiar face recognition is known to be modular: specific damage to the capacity to recognize familiar faces is attested in cases of prosopagnosia. Crucially, the mechanism involved is holistic, not analytical: we recognize the entire face as a whole, and do not break it down analytical into its component parts (eyes, nose, mouth). K-S points out that Williams Syndrome patients use an *analytic* mechanism for familiar face recognition, not a *holistic* one: they can recognize familiar faces just as well as normal, but they use a different mechanism, one which starts out domain-relevant and *becomes* domain-specific. In both cases, the familiar face recognition module is emergent, not innate in the Fodorian sense.

I suggest that we may interpret the notion of a syntax module in the mind, distinct from the phonological module, in terms of this kind of emergent modularity.

6 An alternative to Popper's dualism

Popper's tripartite distinction between World 1 (the mind-external physical world), World 2 (the mind-internal world) and World 3 (objective knowledge, considered as mind-external) was easily interpreted as subsuming a form of Cartesian dualism, whereby the mind is conceived of as something distinct from the body, and thus the brain. In his later work with John Eccles (Popper and Eccles, 1977), Popper explicitly adopted a version of Cartesian dualism, a position later adopted by the linguist Geoffrey Sampson.

Cartesian dualism has been discredited for over a century, since the publication of Gilbert Ryle's *The Concept of Mind*. There are two main objections to Cartesian dualism. Firstly, the dubious ontological status of some non-corporeal mind, as distinct from the brain, which amounts to a form of mysticism. Secondly, even if one is prepared to accept this 'ghost in the machine' mysticism, there is the question of how something non-physical could possible enter into causal relations with something physical (the body, subsuming the brain). Descartes never had a satisfactory answer to this question, precisely because it is conceptually impossible that the non-physical could, in principle, interact with the physical. Descartes' correspondence with Princess Elizabeth (Anscombe and Geach, 1954) show extreme evasiveness on the issue. The postulating of a gland in the brain via which the non-physical mind communicates with the brain is no solution to the problem, since it is question-begging: how could the supposedly non-physical mind interact with a gland?

With a few exceptions (Sampson and some philosophers), we are all physicalists now: the mind *is* the brain. Given a physicalist philosophy, we can re-interpret Popper, using the Burton-Roberts distinction between the E-physical and the I-physical: World 1 is E-physical (the physical world outside of the mind), World 2 is I-physical (the world of the mind, conceived of as the brain) and World 3 is the world of intersubjectivity.

Given this re-interpretation of Popper's trichotomy, we can then adopt the following stance on the ontological and epistemological status of generative syntax: knowledge of syntax is knowledge of social conventions (which are World 3), which are internalized in individual brains (World 2: I-physical). The study of neural mechanisms in the brain (World 2: I-physical) can be subsumed under the natural sciences, but the social conventions cannot, since conventionality does not constitute part of the natural world.

7 Concluding remarks

Generative syntax is the study of syntactic knowledge conceived of here as social conventions: objective knowledge, in Popper's and Itkonen's sense. Social conventions are not the object of naturalistic enquiry, since the object of such enquiry contains no conventions. Generative syntax is not naturalistic science.

However, the internalisation of those social conventions falls within naturalistic enquiry, conceived of as the study of World Two as the I-physical world, which also subsumes the neural mechanisms underpinning those internalisations. World Two has emerged from World One, and World Three has emerged from World Two. None of these ontological realms is reducible: World Three cannot be reduced to World Two, and World Two cannot be reduced to World One.

References

Anderson, John M. 2011. *The Substance of Language: Volume 1, the domain of syntax*. Oxford: Oxford University Press.
Anscombe, Elizabeth & Peter T. Geach, eds. 1954. *Descartes: philosophical writings*. Wokingham: Van Nostrand Reinhold.
Burton-Roberts, Noel. 2011. On the grounding of syntax and the role of phonology in human cognition. *Lingua* 121(14). 2089–2102.
Carr, Philip. 1990. *Linguistic Realities*. Cambridge: Cambridge University Press.
Chomsky, Noam. 1957. *Syntactic Structures*. Mouton: The Hague.
Chomsky, Noam. 1986. *Knowledge of Language*. New York: Praeger.
Chomsky, Noam. 2002. *On Nature and Language*. Cambridge: Cambridge University Press.
Derwing, Bruce. 1973. *Transformation Grammar as a Theory of Language Acquisition*. Cambridge: Cambridge University Press.
Everett, Daniel L. 2005. Cultural constraints on grammar and cognition in Pirahã: Another look at the design features of human language. *Current anthropology* 46(4). 621–646.
Fodor, Jerry A. 1981. *Representations: philosophical essays on the foundations of cognitive science*. Reprinted in 1986. Brighton: The Harvester Press.
Fodor, Jerry A. 1983. *The Modularity of Mind: an essay in faculty psychology*. Cambridge, Mass.: MIT Press.
Gopnik, Alison. 2000. Theories, language and culture: Whorf without wincing. In Melissa Bowerman & Stephen Levinson (eds.), *Language Development: the essential readings*, 45–69. Oxford: Blackwell.
Itkonen, Esa. 1978. *Grammatical Theory and Metascience*. Amsterdam: Benjamins.
Itkonen, Esa. 1983. *Causality in Linguistic Theory*. London: Croom Helm.
Karmiloff-Smith, Annette. 2001. Development itself is the key to understanding developmental disorders. In M. Tomasello & E. Bates (eds.) *Language Development; the essential readings*. Oxford: Blackwell.
Lightfoot, David. 1979. *Principles of diachronic syntax*. Cambridge: Cambridge University Press.
Meltzoff, Andrew. 1988. Imitation, objects, tools and the rudiments of language in human ontogeny. *Human Evolution* 3. 47–66.
Pinker, Steven. 1995. *The Language Instinct: the new science of language and mind*. London: Penguin.
Popper, Karl R. 1972. *Objective knowledge*. Oxford: Clarendon Press.
Popper, Karl R. & John Eccles. 1977. *The Self and its Brain: an argument for interactionism*. London: Routledge and Kegan Paul.
Rose, Steven. 1993. *The Making of Memory: from molecules to mind*. Harmondsworth: Penguin.
Sampson, Geoffrey. 1997. *Educating Eve: the language instinct debate*. London: Cassell.
Sampson, Geoffrey. 2005. *The 'Language Instinct' Debate*. London: Continuum.
Smith, Neil. 1973. *The Acquisition of Phonology: a case study*. Cambridge: Cambridge University Press.
Smith, Neil. 1999. *Chomsky: Ideas and Ideals*. Cambridge: Cambridge University Press.
Tomasello, Michael. 2001. Perceiving actions and learning words in the second year of life. In M. Bowerman & S. J. Levinson (eds.), *Language acquisition and conceptual development*. Cambridge: C.U.P.
Vihman, Marilyn M. 1996. *Phonological development: the origins of language in the child*. Oxford: Blackwell.

Esa Itkonen

16 Hermeneutics and generative linguistics

Abstract: Positivism is the school of thought which advocates methodological monism: the natural sciences, epitomized by Newtonian physics, ought to serve as the model for all sciences. Hermeneutics is the school of thought which focuses on the relations, and in particular the differences, between natural and human sciences, with the understanding that metascience too is a hermeneutic discipline. (The philosophy of a natural science is itself not a natural science.) Generativism is characterized by an internal tension. On the one hand, its official stance is that of all-out positivism. On the other, it turns out to have strong affinities with hermeneutics.

1 Background

Hermeneutics designates that type of philosophy which is meant to do full justice to what is characteristic of *Geisteswissenschaften* 'human sciences'. The antipode of hermeneutics is known as *positivism*. It argues for *methodological monism*, or the view that the natural sciences (= *Naturwissenschaften*) provide the only legitimate model for scientific description. The contrast between positivism and hermeneutics was a lively issue in the 1960's and 1970's, as witnessed by Radnitzky (1970) and von Wright (1971). The fact that the discussion has cooled off since then, should not be taken to mean that the problems involved have been solved; or even if they have (which in my opinion might well be the case), that the solutions have been generally accepted.

It is not only the case that these problems will remain fundamental for the relation between linguistics and philosophy of science; they were fundamental already a long time ago. Just have a look at what August Schleicher had to say in (1863): "The thinking of the modern age goes unmistakably in the direction of monism. Dualism, [understood] as the opposition between spirit and nature, [. . .] has been completely abandoned from today's scientific point of view" (Arens, 1969, 259–260). In the 21st century, however, this is less evident than Schleicher took it to be. It is not without reason that we speak of the *tidewater* of (intellectual) history.

2 Hermeneutics

This is the crux of hermeneutics: How, exactly, is the difference between the human mind and the inanimate nature reflected in the respective descriptions of these two domains? For analytical purposes, the integral domain of hermeneutics must be divided into a number of subdomains.

https://doi.org/10.1515/9783110540253-016

2.1 *Verstehen*: Not only: "*why* did X occur?" but, first of all: "*what* is X?"

To begin with, let us take our cue from Wilhelm Dilthey (1833–1911), the grand master of hermeneutics: "The spirit understands only what it has created. The nature, which is the subject matter of natural science, comprehends the reality uninfluenced by the spirit" (Dilthey, 1927, 148). *Verstehen* 'to understand' is the key concept of hermeneutics. It is opposed to two positivistic concepts at once, namely *beobachten* 'to observe' and *erklären* 'to explain': the occurrence of some physical event is first observed and then explained. But how can *one* notion (= *verstehen*) correspond to *two* notions (*beobachten* and *erklären*) which are located at different levels of abstraction? The answer to this question will emerge gradually.

According to positivism, it is the primary purpose of natural science to produce theories which explain and predict physical events; and explanations, more particularly, answer *why?*-questions. Therefore, if hermeneutics is to be compared with positivism, it may seem logical to concentrate on how *why?*-questions are answered by the method of *Verstehen*. But this would prejudice the comparison from the start.

As will be seen in Section 3, it is not at all uncommon to conceptualize *linguistics* as just one natural science among others. But let us for a moment shift the focus from linguistics to *philology*, the discipline which, thanks to Friedrich Schleiermacher (1768–1834), gave the initial impetus to hermeneutics: "It is actually the 'philologies' which qualify as the genuine hermeneutic sciences [*Geisteswissenschaften*] [...]." (Apel, 1976b, 113). Thus, the hermeneutic effort was taken to concentrate, first and foremost, on interpreting the *meaning* of texts, whether ancient or modern. Within the scope of this task Schleiermacher distinguishes between two types of interpretation (or understanding), which he calls *grammatical* and *psychological*. The former concerns the speaker's or writer's relation to his/her language as a whole while the latter is supposed, ideally, to comprehend his/her entire thinking: "It is the interpreter's task to understand the text first as well as, and then better than, its author originally did" (Schleiermacher, 1977, 94).

It goes without saying that the questions asked by the *grammatical* interpretation are of the type *what?*, not *why?* This is indeed a conceptual necessity. As demonstrated by Itkonen (1991, 2000), the grammatical traditions of all cultures start by trying to answer the question: "*what* is the correct sentence of the language L?" Let it be added that semantic descriptions, in particular, have in general emerged from philological-exegetical work (cf. van Bekkum et al., 1997).

The same is true, *mutatis mutandis*, of the *psychological* interpretation as well. When we approach an author, we first of all ask *what* it is that s/he was thinking when s/he wrote X. Of course, we can reformulate this as a *why?*-question: "*Why* did s/he write X (and not e. g. Y)?" Now, two things should be noticed. First, this question is equivalent to "*What* did s/he mean by writing X (rather than Y)?" As noted by Collingwood (1946, 214), "when [the historian] knows what happened, he already knows why it happened". Second, if we nevertheless insist on asking a *why?*-question (rather than a *what?*-question), it will be answered by a *rational explanation*, which is qualitatively different from natural-science explanations (see below).

First and foremost, the brute fact remains that nobody has ever tried to conceptualize e. g. biblical exegesis in the natural-science terms. (This is possible, of course, but extremely implausible.) The same is true of all exegetical work, for instance of the work that I have done with Chomsky's ([1955] 1975) dissertation (cf. Section 3). In pointing out, and solving, a set of contradictions implicit in this work, I indeed try to follow Schleiermacher's exhortation to "understand the text better than its author did". But in so doing, I am surely not explaining and/or predicting any physical events.

According to standard accounts of the Western history of philosophy, the Greeks endorsed a contemplative, *mirror-of-nature* view of knowledge. This interpretation is misleading, in more than one way. In particular, both Plato and Aristotle also argued for a view which is perfectly captured by the Dilthey-citation at the beginning of this section: we can genuinely understand (and know) only what *we ourselves* have created.

This idea reappears in the *verum* = *factum* doctrine developed by Giambattista Vico (1668–1744): only such statements presented by A which are about what A has made him-/herself, can be (genuinely) known by A to be true, as opposed to statements which are about what A has merely observed. "The emphasis on this contrast, which runs through all Vico's thought, is, in effect, the original formulation of the familiar and much controverted distinction between the methods and goals of *Naturwissenschaft* and *Geisteswissenschaft* [. . .]" (Berlin, 1976, 24). Hobbes and Kant, for instance, also made use of this notion of agent's (as opposed to observer's) knowledge. Different aspects of this topic have been discussed by Hintikka (1974, Ch. 2) and Itkonen (1978, 193–198; 1983, 298–302; 1991, 169, 189–191; 2003b, Ch. 11).

2.2 *Verstehen*: Not only: 'Subject → Object', but first of all: 'Subject ↔ Subject'

Let us quote Dilthey once again: "The nature is alien to us. For it is just an exterior, nothing internal. The society is our world" (Dilthey 1914, 36). The seemingly innocuous statement "the society is our world" proves to be crucially important. Who are *we*? What is meant, is the community constituted both by the (social) scientists (= human beings) and by their research objects (= other human beings). In the natural sciences the so-called subject – object divide is a necessity: it is not possible to mistake an astronomer for a planet, or vice versa. Positivism demands that the boundary between scientists and their research objects should be equally strict in the human sciences as well. Hermeneutics counters that this is not only artificial but also logically impossible.

From the hermeneutic point of view, each human science involves two levels of knowledge, pre-theoretical and theoretical, and scientific work can be conceptualized as an *ascent* from the former to the latter: “Psychological understanding becomes of itself psychological investigation”, as Dilthey once put it. Accordingly, human sciences can be called two-level sciences (cf. Itkonen, 1978, 198–219). As far as knowledge is concerned, a natural science like astronomy, by contrast, remains a one-level science. Planets have no knowledge, and there is no way that a planet can *ascend* so as to become an astronomer.

The hermeneutic position is supported by the fact that as long as sociologists or linguists engage in conversations with their research objects (as is typically the case), the latter cease to be *objects* and become, technically speaking, *co-subjects*. Thus, their status vacillates between *object* and (*co-*)*subject*. Acknowledging this fact is the starting point of, e. g., the *verstehende Soziologie* of Max Weber, the symbolic interactionism of G. H. Mead, and the phenomenological sociology of Alfred Schutz. *Kommunikationsgemeinschaft* ‘community of communication’, whose members are co-subjects by definition, is the *a priori* presupposition of modern hermeneutics: “Hermeneutic understanding assigns the role of a dialogue partner to the interpreter” (Habermas, 1968, 226). “The validity of hermeneutic hypotheses can be confirmed or falsified only if the scientist abandons, at least to some extent, the role of an observer and knowingly participates at the language game s/he is supposed to understand” (Apel, 1976b, 271).

Moreover, the notion of *co-subject* can be given a stronger interpretation according to which a theoretical description of human behaviour should be accessible to those who it is about. This principle has supporters within emancipatory sociology and psychoanalysis: “even explanations of the Freudian type, if they are to be acceptable, must be in terms of concepts which are familiar to the agent as well as to the observer” (Winch, 1958, 48). The background metaphor is the understanding of a common language: “Freud has always conceptualized the interpretation of dreams on the hermeneutic model of philological research” (Habermas 1968, 263). – Bühl (1972) is a good collection of articles on *Verstehen*.

2.3 Historical Understanding: *Einfühlung*/Empathy

“Our historical understanding is exactly the same as when we understand someone who is talking with us.” The real significance of this apparently simple statement becomes evident when we add that it was made, in 1857, by J. G Droysen (1808–1884), the leading German historian of his time (who, incidentally, was the first to coin the terminological *verstehen* vs. *erklären* distinction). In order to achieve a *conversation* with those who have lived a long time ago, the historian must endeavour to bridge the gap produced by the temporal and cultural distance: “According to the older hermeneutics (Schleiermacher, Droysen, Dilthey), the historian must transfer him/herself into the total situation of those actions which s/he wants to understand" (Apel, 1976b, 109).

After Dilthey, this position was again reasserted by Collingwood (1946), in terms which clearly place him in the same agent's knowledge tradition with Plato and Vico: "For the historian, the activities whose history he is studying are not spectacles to be watched, but experiences to be lived through in his own mind; they are objective, or known to him, only because they are also subjective, or activities of his own" (1946, 218).

Within the philosophy of history, the Collingwood-type *re-enactment* of past actions has been ridiculed as being some near-mystical ability to penetrate an *individual* mind located in a perhaps very distant past. This criticism is quite unfair, as Dray (1980, 21–25), for instance, has pointed out. Trying to understand any action, historical or not, means trying to find the *reason* for doing it; a reason (qua goal-belief constellation) constitutes a *practical* argument (or inference); and "arguments, of course, have *generality*" (Dray, 1980, 23, emphasis added). First, there is no alternative to doing what Collingwood does. Second, it is not individual but general (and potentially even universal) in character.

At this point we must insert a comment on the nature of beliefs (and goals): "Every belief must have both a history and a logic; for they are concerned each with a different element of the belief. 'Believe' is a psychological verb and the history of a belief is therefore a psychological story; what is believed, a proposition, is a logical entity, having only logical properties and relations, which are non-temporal" (Edgley, [1965] 1978, 24). It is only in their latter capacity that beliefs (and goals) can possess conceptual (= necessary) relations and thus be contained in (practical) *arguments* (or inferences) (see below).

In the Popperian terminology, goals and beliefs are inhabitants both of (psychological) world-2 and (social-normative) world-3. This duality was obvious to Collingwood: "In a sense, these thoughts are no doubt themselves events happening in time; but [. . .] there is another sense [. . .] in which they are not in time at all" (1946, 217). In fact, this is what was meant, in the Collingwood-quotation given above, by the subjective-*cum*-objective nature of what the historian is investigating.

Now we are in a better position to understand what Dray ([1963] 1974) means by *rational explanation*: "we cannot help certifying the agent's reasons as good ones, from his point of view, for doing what he did. [. . .] It is out of a very proper recognition of this fact, it seems to me, that controversial theories of 'empathy' in history derive their plausibility" (Dray, [1963] 1974, 73).

This account of historical understanding remains seriously one-sided, as long as we fail to duly account for the idea of *Gesamtsituation* 'total situation', mentioned in the Apel-quotation above. Certainly the antecedents of the World War II cannot be adequately described by any historian who does not try to re-enact Hitler's thought processes. But these must be seen against the larger background which contained at least the following elements: the resentment of the German population after the treaty of Versailles; the global economic depression; Stalin's injunction for German communists not to co-operate with social democrats; the ambiguous attitude of the German generals who were critical of Hitler's personality but approved of his rehabilitation of the German military. "The historian chooses the total situation (economic, strategic, and cultural) which makes it possible to explain the event" (Habermas, 1970, 111).

At the limit, *total situation* becomes identical with an entire culture. This idea goes back at least to J. G. Herder (1744–1803): "But what lies at the heart of the whole of [Herder's] thought, [. . .] is the theme to which he constantly returns; that one must not judge one culture by the criteria of another; [. . .] so that to understand them one must perform an imaginative act of 'empathy' into their essence, understand them 'from within' as far as possible, [. . .]" (Berlin, 1976, 210). The German words corresponding to empathy are *Einfühlung* and *Nacherleben*.

Now, it is easier to understand that it is always in some larger context that empathy-based explanations of historical actions are first placed and then continually revised (or *disconfirmed*) as more and more facts are taken into account, facts which may in turn be revised in the light of improved action-explanations; and so on.

Explicit references to empathy are quite rare in linguistics. I think this is a mistake, as I have tried to show with many examples taken from diachronic and/or typological linguistics in, e. g., Itkonen (1983, 216–219; 2004, 27–32; 2013, 46–53). Rational means–ends explanation is ubiquitous in linguistics, and so is empathy, because the former is based on the latter. More precisely, empathy and (linguistic) intuition may be characterized, respectively, as vicarious introspection and conventionalized empathy (2008a, 25–27; 2013, 58–60). Dray ([1963] 1974) reminds us of the slogan "Historical study is vicarious experience": "like all such slogans, [it] fastens only on one facet of the truth. But it *is* a facet of the truth [. . .]" (Dray, [1963] 1974, 89). We shall take up this topic below.

Let us in conclusion return to Droysen, with the aid of Winch (1958): "Historical explanation [. . .] is like applying one's knowledge of a language in order to understand a conversation rather than like applying one's knowledge of the laws of mechanics to understand the workings of a watch" (1958, 133).

2.4 Social Understanding: Normativity

Dilthey (1914, 30–36) points out this fundamental difference: part of the data investigated within the *Geisteswissenschaften* are of *normative* character whereas this is never the case within the *Naturwissenschaften*. The same or similar topic – under the designation of "what it is to follow a rule" – is a central concern in Wittgenstein's later philosophy; and it has been further developed e. g. by Winch (1958). Wittgenstein is too complex a figure to admit of any easy label, but Winch's contribution can be characterized in less equivocal terms: "This is how Winch actually manages to accomplish something amounting to the foundations of philosophical hermeneutics" (Apel, 1976a, 80).

A preliminary notion of rule/norm emerges from the following statements: "Understanding something involves understanding the contradictory too: I understand what it is to act honestly just so far as and no farther than I understand what it is not to act honestly" (Winch, 1958, 65). "[T]he notion of following a rule is logically inseparable from the notion of *making a mistake*" (p. 32). "[T]he test of whether a man's actions are the application of a rule is not whether he can *formulate* it but whether it makes sense to distinguish between a right way and a wrong way of doing things in connection with what he does" (p. 58). It follows, among other things, that rules (of language) are inherently *social*. Why? Because the person who would want to follow the rules of an entirely private language has no way of knowing whether or not s/he has made a mistake. This is, in a nutshell, Wittgenstein's famous private-language argument (cf. Winch, 1958, 33). Its implications for the philosophy of linguistics have been spelled out by Itkonen (1974, 50–66; 1978, 109–113; 2008b, 280–283).

As far as linguistics is concerned, the general notion of norm may be divided into *rules of correctness* and *principles of rationality*, as suggested by Itkonen (1983). The former are called constitutive rules (of language) by Searle (1969, 33–42). The latter, also called technical norms, are exemplified by "if you want to make the hut habitable, you ought to heat it" (cf. von Wright, 1963, 9–11). Informally formulated, their applications are like this: "He heated the hut *in order to* make it habitable". If explicitly formulated, such applications are in turn identical with *rational explanations* (cf. below).

Rules (of correctness) either exist or do not exist whereas rule-formulations are either true or false. A sentence that describes a simple well-established rule (like "In English the definite article precedes the noun") has the *prima facie* peculiar property that it cannot be falsified by what occurs in space and time: correct actions are irrelevant because they conform to the rule while incorrect actions are irrelevant because they deviate from it. (Analogously, a principle of rationality is falsified neither by rational nor by irrational actions.) This claim is open to 12 standard objections, all of which are answered by Itkonen (2003b, 18–21, 43–36). For now, suffice it to say that grammatical descriptions, i. e. *theoretical*-descriptive generalizations about (large) sets of rules, are of course falsifiable.

"Look at the sentence as an instrument and at its sense as its employment" (Wittgenstein, 1958, § 421). "This is why there is a correspondence between the concepts 'rule' and 'meaning' " (Wittgenstein 1969, § 62). Summarized as a slogan: "meaning is *use*"; and, slightly more elaborately: "meaning is *correct* use" (as indicated by the constant reference to *rules*). The meanings of such words as *buying*, *selling*, *possession*, or *money*, are, qua normative entities, shown by how they are used. The relations that obtain between meanings are typically of *necessary* character: it is a necessary truth that if (and only if) A bought B from C, then C sold B to A and that B, having been a property possessed by C, came to be possessed by A. But the same is also true of (the meanings of) the actions described by these words: "if social relations between men exist only in and through their ideas [such as 'buying', 'selling', 'property', 'money'], then, since the relations between ideas are internal [= necessary] relations, social relations must be a species of internal relation too" (Winch, 1958, 128). Thus, such terms as social and semantic turn out to be near-synonymous.

From what precedes, Winch (1958) makes a generalization which at first seems excessive: "all behaviour which is meaningful (therefore all specifically human behaviour) is *ipso facto* rule-governed" (p. 52). But, on reflection, this generalization is justified (or entailed) by the very notion of *meaningful* behaviour. Meaning is not arbitrary but determined by rules. These rules do not determine which actions we in fact do; they determine the way we *understand* whatever it is that we do. To open a window is different from closing it; the one who thinks that s/he is opening a window when s/he is in fact closing it commits a mistake; and (as we have just learned) there is no mistake without a corresponding norm. Notice also what Winch is *not* saying. Although the notion of *rule* applies to all meaningful behaviour (in the sense specified here), it does *not* follow that the same behaviour could not also exemplify some (non-normative) *regularities*, manifested as correlations like the one between economic recession and social upheaval (cf. Itkonen, 1978, 156–166, 182–86).

Natural science describes what is physical and thus inherently non-normative. If science is equated with natural science, then a science purporting to describe rules/norms becomes logically impossible: "If rules involve the concepts of right and wrong, they introduce a normative aspect that has always been avoided in the natural sciences. [...] To admit that language follows rules seems to put it outside the phenomena accessible to *scientific* investigation" (George A. Miller, quoted from Koestler, 1967, 42; emphasis added). For a consistent positivist, this conclusion is indeed unavoidable. The other option is to assume that language belongs to the phenomena accessible to *hermeneutic* investigation (or hermeneutic science).

While natural science describes what is non-normative, those who practice natural science must of course follow (scientific) norms of their own; and, qua norms, *these* must be outside the phenomena accessible to natural science, which means that they must by definition be accessible to hermeneutic science: "The human sciences now achieve the status of a metascience" (Habermas, 1970, 80). Apel (1976b) agrees: "In order to know 'what' it is that s/he must explain, the natural scientist must have reached a previous understanding on it with his/her fellow scientists [...]. This kind of mutual understanding can never be reached (or replaced) by the methods of physical science" (Apel, 1976b, 112). Physics describes (e. g.) the motion of bodies in space and time, but physics itself is not described by physics, but by philosophy of physics, which qua study of norms is a hermeneutic discipline.

2.5 Philosophical and Logical Understanding

Each and every kind of philosophy is concerned with norms of correct thinking: "The problem for epistemology is not 'why *do* I believe this or that?' but 'why *should* I believe this or that?' " (Russell, [1940] 1967, 14). As far as philosophy of science is concerned, this remark was given an abstract justification in the last paragraph of 2.4. It is justified in more concrete terms by Cohen (1986). As he sees it, science has the goal of "determining whether or not a given set of beliefs is well-attested". After enumerating several techniques and principles which scientists make use of in order to achieve this goal, he concludes: "they are all norms, not factual beliefs" (Cohen, 1986, 44).

More concretely still, the so-called *proto*-physics, created in the 1960's by Paul Lorenzen, has as a matter of fact formalized those norms which constitute the core of *physics*:

> The reconstruction of scientific measurement is called 'protophysics'. It is subdivided into the theories of measuring space, time, and mass, or geometry, chronometry, and 'hylometry' [. . .], which are sciences of increasing complexity. The objective character of (Galilean) physics is due to the intersubjective agreement upon those *ideal norms*, formalizable or explicable as the axioms of protophysics, which govern actual measurement. Protophysics [. . .] is an *a priori* science which, instead of investigating actual physical events, investigates possible physical events, i. e. the concept 'physical event', as defined by the ideal norms of measurement. (Lorenzen, 1969 and Böhme, 1976) (Itkonen 1978, 45; also 2005, 40)

In sum, protophysics defines "the normative foundation of observation" insofar as "objects of possible observation are conceptualized by us as measurable bodies" (Habermas and Luhmann, 1971, 127).

Let us move from philosophy of science to philosophy proper. Ever since antiquity, philosophers have been concerned with analyzing such *concepts* as knowledge and justice. These are identical with the *meanings* of the corresponding words, as determined by the *norms* for their use (again, cf. 2.4). Three things follow. First, philosophy proves to be a hermeneutic discipline (or "science") *par excellence*, given that meanings and norms are the traditional concern of hermeneutics.

Second, there is – *prima facie* – no difference between philosophy as conceptual analysis and linguistic semantics. The latter point has been argued at length by Itkonen (2016), against attempts to reduce philosophical-*cum*-linguistic semantics to empirical psychology. (It is a different matter that any area of linguistics, from phonology to semantics, can be complemented *ad libitum* by psychological considerations.)

But third, philosophical analysis can*not*, on reflection, be simply identical with an analysis of concepts/meanings as embodied in existing uses of language. It is the traditional task of philosophy to develop *better* concepts/meanings and thus to *improve* upon existing uses of language: "As a tool of analysis, the use theory of meaning can provide us only with certain data, i. e., raw material for philosophical analysis; [. . .]"

(Searle, 1969, 148–149). More recently, the same sentiment has been expressed by Dennett (2014, 423).

Just as there is a distinction between physics (= A) and philosophy of physics (= A*), there is a distinction between philosophy (= B) and philosophy of philosophy, or *meta*philosophy (= B*). By now it should be clear that while A and A* are, respectively, non-hermeneutic and hermeneutic, both B and B* are hermeneutic (or, perhaps, B* is hermeneutic "to the power 2").

Pap (1958), the modern classic of metaphilosophy, endorses the traditional view that it is the task of philosophical research to produce conceptual analysis or *explication*. This is, in outline, how it is done: "The concept of entailment (and the related concepts of self-contradiction and logical incompatibility) is the primary tool by means of which analytic philosophers undertake to analyze concepts" (p. 92). "[S]emantic analysis of natural language involves *intuitive* knowledge of necessary propositions, [. . .]" (p. 396). For example, consider this entailment (= necessarily true implication): "If X knows that p, then *p* is true". It is not necessary in the sense that it cannot be denied, but in the sense that it *ought* to be accepted, i. e. *not* to be denied: the one who denies it makes a *mistake*. This reveals, in a nutshell, the normative basis of philosophy.

Cohen (1986) gives the same account of the methodological status of analytical philosophy. All of its problems are "normative problems connected with the rationality of judgment, attitude, procedure, or action" (p. 46–47), and it is only by means of intuition that they come to be known, first of all. Thus, analytical philosophy is based on the use of intuition, as confirmed by quotations taken from nine leading representatives of this school (p. 77–79). The philosophical method itself can be summarized as intuition-*cum*-reflection, i. e. an ascent from intuitive pre-understanding (*Vorverständnis*) to theoretical reflection (cf. also Pap, 1958, 274).

The normative basis of philosophy, though undeniable, may not be immediately obvious. By contrast, axiomatic logic endeavours to formalize the concept valid formula in system S, which is based on the *openly* normative valid vs. non-valid distinction. In the same way, grammatical description is meant to define the concept correct sentence of language L, by systematizing the normative correct vs. incorrect distinction. In both cases, it is by means of (logical or linguistic) *intuition* that these distinctions and concepts are grasped, in the first place.

Insofar as a system of *logic*, with its axioms, definitions, and inference rules, is considered as a formal object, it needs to be interpreted by *meta*logic, just as – more generally – all mathematical objects are interpreted by *meta*mathematics:

> The metatheory will be expressed in ordinary language, with mathematical symbols, such as metamathematical variables, introduced according to need. [. . .] [The deductions] must proceed by *intuitive* inferences, and not, as the deductions in the formal theory, by applications of stated rules. Rules have been stated to formalize the object theory, but now we must understand without rules how those rules work. An *intuitive* mathematics is necessary even to define the formal mathematics. (Kleene 1952, 62; emphasis added)

An axiomatic system of logic is sound if it generates *only* valid formulae; it is complete if it generates *all* valid formulae; and it is adequate if it generates all and only valid formulae. (This is exactly how an adequate generative grammar, too, is defined, once validity is replaced by correctness/grammaticality.) The questions answered by soundness and completeness are, respectively: "this formula is generated by the system; is it valid?" and "this formula is valid; is it generated by the system?" Soundness and completeness are metalogical properties. Insofar as a system can be *formally* proved to possess them, this formalization needs in turn to be interpreted by intuitive (= non-formal) *meta*metalogic (cf. Itkonen, 2003a, Ch. VII).

Logic (= C) → metalogic (= C*) → metametalogic (C**): It may seem surprising to learn that all this belongs to the realm of hermeneutics, given that representatives of hermeneutics have traditionally preferred not to make much use of formalization. But there is no principled reason for this preference; it is just a matter of taste. Surely the descriptive practice of a (meta)logician is far removed from the ideal of positivism, namely explaining and predicting spatio-temporal events. And let us not forget that Husserl, the anti-positivist *par excellence*, had a life-long interest in the foundations of logic (cf. Itkonen, 1978, 53–54; 1991, 285–286).

2.6 Consciousness

To fix our ideas, let us make, first, a clear distinction between *ontology* (= what exists?) and *epistemology* (= how is it known?). Second, let us categorize ontology into the three Popperian worlds: physical (w-1), psychological (w-2), social/normative (w-3). Third, let us categorize epistemology into three types of gaining information, depending on which world they are directed at: w-1 = *observation* (i. e. sense-perception), w-2 = *introspection*, w-3 = *intuition*. This overall view was already expressed by Frege ([1918] 1967), in his own terminology: "One sees a thing [w-1], one has an idea [w-2], one apprehends or thinks a thought [w-3]" ([1918] 1967, 29). (Notice the idiosyncratic use of *thought*.) The same two-way trichotomy was endorsed by Itkonen (1981a, 131–132; 1983, 8–9) and Katz (1981, 194–196). To repeat, it exemplifies the Popperian tripartite ontology.

On reflection, however, the situation is slightly more complicated. It makes sense to assume that the human mind is ontologically homogeneous, but epistemologically it is heterogeneous, namely either conscious or unconscious (with gradations); and it is only (what is experienced by) the conscious mind that can become an object of introspection. At least in experimental psychology, the unconscious mind is investigated by the same hypothetical method as are unobservable entities in physics. Moreover, information about *other* minds is gained by empathy, defined as *vicarious* introspection/experience, rather than by immediate introspection (cf. 2.3). Let it be added that the *existence* of (as opposed to specific *information* about) other minds is a necessary part of every *consistent* conceptual framework. This is the lesson of the private-language argument.

Each of the three information-gaining acts is subjective in the sense of emanating from w-2. Observation and intuition are directed, respectively, at w-1 and w-3, while introspection is directed, self-reflectively, back at w-2. Observations of w-1, introspections about w-2, and intuitions about w-3 can of course become fully conscious. By contrast, the data of natural science, i. e. w-1-itself, lacks any form of consciousness just as it lacks any form of normativity (cf. 2.4). How about w-2 and w-3?

As for w-2, the question was already answered: by definition, introspection is about conscious experiences only. (This is why the object of introspection is also called content of consciousness.) As for w-3, this is the question to be asked: Can (or must) norms qua data be accessible to conscious intuition?

Let us, once again, take our cue from Winch (1958). As argued in 2.4, rules/norms are social or public; they entail the possibility of making a mistake; and mistakes must be potential objects of consciousness: "A mistake is a contravention of what is *established* as correct; as such, it must be *recognisable* as such a contravention. That is, if I make a mistake in, say, my use of a word, other people must be able to point it out to me" (1958: 32). Insofar as we are able to become conscious of deviations from a norm, we must *eo ipso* be able to become conscious of the norm which they are deviations from.

Thus, consciousness is involved not only in philosophical-*cum*-logical understanding (cf. 2.5), but also in social understanding (cf. 2.4). It may be less obvious that consciousness is involved also in historical understanding (cf. 2.3), even at the level of data: "The historical process is itself a process of thought, and it exists only in so far as the minds which are parts of it *know* themselves for parts of it" (Collingwood, 1946, 226; emphasis added).

Due to the influence of positivism, there has been a persistent tendency to deny the role of consciousness in *linguistics*. But: "sooner or later we will have to restore conscious experience to the central role it enjoyed in the human sciences a hundred years ago" (Chafe, 1994, 7). "Meaning is located in conscious experience" (Talmy, 2000, 5). Arguments to the same effect have been offered e. g. by Zlatev (2008).

But this is not enough. In addition, a clear distinction has to be made between (conscious) *intuition* and (conscious) *introspection*. To put it briefly, subjective experience, i. e. what introspection is about, is not correct or incorrect in the same sense as are linguistic forms and meanings, i. e. what linguistic intuition is about. Quite dissimilar results are achieved in linguistic research, depending on whether one is describing normative intuitive knowledge or non-normative introspective knowledge (cf. Itkonen, 2008a, 2016).

It was in connection with the philosophy of *logic* that the importance of this topic was realized, first by Frege and then, following him, by Husserl. Frege (1893) argues that logic deals with laws of thought (*Denkgesetze*), in the sense of prescribing how people ought to think, instead of describing how they think in fact. This normative dimension is independent from factual opinions just as being-true (*Wahrsein*) is from being-held-to-be-true (*Fürwahrgehaltenwerden*) (Frege, 1893, xv–xvi). Husserl (1913) realizes that this must be true more generally as well: “Thinking about mathematico-logical validity, Husserl became acutely aware of the fact that meanings in general are distinct from actual experiences” (Apel, 1981, 87).

Above, it was briefly indicated how the same semantics vs. psychology distinction is conceptualized by Frege ([1918] 1967). It is, again, the same distinction that was conceptualized by Edgley ([1965] 1978, 24) by saying that every belief has both a “history” and a “logic” (cf. 2.3).

2.7 Explanation

“Two main traditions can be distinguished in the history of ideas, differing as to the conditions an *explanation* has to satisfy in order to be scientifically respectable. The one tradition is sometimes called *aristotelian*, the other *galilean*” (von Wright 1971, 2; the first emphasis added). The two traditions are identified with hermeneutics (1971, 5) and positivism (1971, 3), respectively. Let us have a look at their respective notions of explanation.

“[The positivistic explanation] consists, more specifically, in the subsumption of individual cases under hypothetically assumed general laws of nature [. . .]” (p. 4). Hence, these explanations must be *lawlike* (nomic, nomological, nomothetic) in character. Hempel (1965) expresses the received view on this issue: he presents two distinct explanatory models, i. e. deductive-nomological and inductive-statistical, depending on whether the laws involved are of deterministic or of statistical nature (cf. Itkonen, 1978, 4–12; 2003a, 201–219). These are also called covering-law explanations.

Hermeneutic explanations qualify as teleological (or finalistic) because they rely on the notion of *goal*. Living organisms display processes that are both goal-directed and nomic (and are explained accordingly) (cf. Itkonen, 2013–2014). But they are not typical data for hermeneutics. Instead, genuine hermeneutic explanations are both teleological and *non*-nomic; they are called *rational explanations* or practical inferences/arguments.

First, this is the basic *justification* for rational explanations: “To explain an action as an action is to show that it is rational. This involves showing that on the basis of the goals and beliefs of the person concerned the action was the means he believed to be the most likely to achieve his goal” (Newton-Smith, 1981, 241). More succinctly: “rational action is its own explanation” (Hollis, 1977, 128). Second, the (tentative) *discovery* of rational explanations is based on empathy: “We must assume a common rationality, and argue from what we would do to what others would do” (Gibson, 1976, 116; cf. here 2.3).

Next, the general schema of rational explanation may be summarized as follows: $(G \;\&\; B) \rightarrow A$, i. e. “the goal-*cum*-belief causes the action”, where the relevant (psychological-*cum*-logical) *G* & *B* is called the *reason* for doing *A*. The actual structure of rational explanation is more complicated (cf. Itkonen, 1983, 49–54; 2013, 54–58). Rational explanations cannot be discredited by objecting (as is often done) that "not all actions are rational". This is a trivial misunderstanding. Rational explanations explain not only actions that *are* rational but also (irrational) actions which (wrongly) *seem* to be rational. To be sure, an extra "burden" is imposed on rational explanations applied to irrational actions (cf. Davidson, 1975, 11). In the light of new evidence, moreover, what first seems irrational may later prove to be rational, after all.

One and the same bit of behaviour admits of several act-identifications, which reveals the relative nature of the boundary between *what?* vs. *why?* -questions and -answers (as suggested in 2.1): "He pulled the trigger in order to fire a gun / to assassinate a foreign diplomat / to start a war / to secure his place in the history books": "*What* did he do?" – "He pulled the trigger." > "*Why*?" – "In order to fire a gun." > "*What* did he do?" – "He fired a gun." > "*Why*?" – "In order to assassinate a foreign diplomat." > [. . .] > "*What* did he do?" – "He secured his place in the history books." *Ergo*: to know what he did is to know why he did it.

Teleological explanations are often contrasted with causal (and nomic) explanations, but this cannot be quite right. There would seem to be no point in *explaining* actions by corresponding reasons, unless the goals and beliefs involved are taken to be ontologically real and causally effective entities. Thus, teleological turns out to be a subtype of causal (cf. Itkonen, 1983, 31–34). The causal element remains hidden in a formulation like "X did A in order to achieve G", but it becomes visible in the following formulation: "It is *because* X had the goal G and the belief B of A being the best means to achieve G that s/he set out to do A".

Traditionally, non-nomic causation was taken to be a *contradictio in adiecto*. But causality was (correctly) thought to be the only genuinely explanatory notion. Therefore, it seemed to follow that any type of *scientific* historiography must be based on universally valid *laws* of history. This, however, turned out to be a complete failure:

> All seemed ready, particularly in the nineteenth century, for the formulation of [a natural science of history] [. . .]. The stage was set, but virtually nothing materialised. No general laws were formulated – not even moderately reliable maxims – from which historians could deduce (together with knowledge of the initial conditions) either what would happen next, or what had happened in the past [. . .]. Neither sociologists nor psychologists had been able to create the new mechanism: the 'nomothetic' sciences [of history] remained stillborn". (Berlin, [1960] 1980, 110)

Exactly the same conclusion was reached by Dray ([1963] 1974) as far as the scientific status of historiography is concerned. There are no universal laws of history. If one tries to explain a particular historical event by invoking *prima facie* relevant laws, these turn out to be more and more specific, until they apply only to the event to be explained (for discussion, see Itkonen, 1983, 95–102). And yet, the historical explanation itself may be impeccable. Conclusion: Historical explanations are non-nomic, which just reflects the fact that, if there is causation in history, it too is of non-nomic nature.

The same remarks apply to the scientific status of sociology. Certainly a huge number of statistical regularities valid in this or that society have been discovered, but no universal laws of society. Giddens (1976) draws the inevitable conclusion: “I shall dogmatically assert the need for an account of *agent causality* [. . .], according to which causality does not presuppose ‘laws’ of invariant connection (if anything, the reverse is the case) [. . .]” (p. 84).

Again, the same remarks apply to diachronic linguistics: “The laws of general-historical phonetics and morphology do not suffice to explain a single fact [. . .]. We are not able to predict a single future development” (Meillet, 1921, 16). Hence, we are faced with a choice: either to endorse the idea of non-nomic explanation or to abandon the idea of explanation altogether. Lass (1980) opts for the latter, rather gloomy alternative and, rejecting the friendly advice offered by Itkonen (1981b), continues to expound this line of thinking in his (1997) book. On the one hand, he admits defeat: “we still have no convincing explanations for change” (p. 386). On the other, he refuses to change his mind: “my fundamental error, as people like Esa Itkonen and Raimo Anttila have never ceased reminding me, both in print and in person, is that I have failed to see that linguistics (of all kinds) is not a causal but hermeneutic science [. . .]” (pp. 332–333). “I reject completely [. . .] ‘the hermeneutic challenge’ [. . .]” (p. 337).

For the sake of completeness, let us add that Lass has not been the only one to express such misgivings: “Hermeneutics seems to me nothing more than a licence for sloppy thinking; [. . .]” (Sampson, 1975, 598). To be sure, this verdict was softened to some extent: “Itkonen believes in hermeneutics, although he is not himself a sloppy thinker; and his case is constructed about as skilfully as it could be” (Sampson, 1975, 598).

2.8 Hermeneutic circle

The contrast between positivistic explanations and hermeneutic ones can also be formulated somewhat differently, namely as a contrast between vertical and horizontal. *Vertical* explanation presupposes a clear difference between levels of abstraction and proceeds from general to particular, typically by means of deduction. *Horizontal* explanation takes a set of phenomena located, in principle, on the same level of abstraction and shows that, instead of just being disconnected items in a list, they constitute a coherent whole or, alternatively, form a *pattern*, and are by the same token *explained* (in a non-deductive sense); hence the designation pattern explanation. This notion was developed for behavioural or social sciences by Kaplan (1964) and Diesing (1972). Its importance for linguistics has been acknowledged by Itkonen (1983) and Anttila (1989). Each and every linguist makes unwittingly use of it all the time.

In reality, we are dealing here with a figure of thought that has been part and parcel of hermeneutics ever since Schleiermacher. More recently it has been formulated as follows:

> "The meaning of the part is determined from its place in the whole, and that of the whole from the way it contextualizes the parts. The movement back and forth, from parts to whole, is one way of characterising what is often described as the 'hermeneutic circle'." (Anderson et al., 1986, 69)

The connection between rational explanation and pattern explanation may not be obvious at once. The following serves as a clarification: "The cogency of a teleological [= rational] explanation rests on its ability to discover a *coherent pattern* in the behaviour of an agent. Coherence here includes the idea of *rationality* both in the sense that the action to be explained must be reasonable in the light of assigned desires and beliefs, but also in the sense that the assigned desires and beliefs must fit with one another" (Davidson, 1975, 11; emphasis added). Next, the (hypothetical) goal & belief complex must fit with the total situation (cf. 2.3). If it does not, the description needs to be revised, perhaps repeatedly. This is why the metaphor of (ascending) spiral seems more befitting than circle, which invokes unwanted connotations of circular thinking.

Typological research, for instance, involves a huge number of hermeneutic spirals, exemplified by the back-and-forth movement between e. g. noun vs. verb or word class vs. sentence structure, and ultimately particular language vs. universal framework (cf. Itkonen, 2011b, 2011d). To give another example, consider the philosophical problem known as the paradox of analysis: Let A and B stand for analysandum and analysans, respectively. Now, the analysis is trivial, if its result is A = B; and it is false, if its result is A ≠ B. The answer to this problem is provided by (one version of) the hermeneutic spiral: "A and B they represent different stages of a *process*. A represents a body of knowledge in its prescientific, atheoretical state. B represents a *different* state of the *same* body of knowledge, viz. its scientific or theoretical state. This explains why the two are simultaneously identical and different; [...]" (Itkonen, 1978, 301).

This passage complements the remarks made on philosophical understanding in 2.5. What Pap (1958) says about the nature of explication is correct in itself, but fails to focus on the process aspect.

There is an interesting addition to be made here. Rescher (1979) regards science as "cognitive systematization" and, just like Kaplan (1964) and Diesing (1972), he too makes a distinction between a vertical approach and a horizontal one, but in a slightly different sense. For him, the vertical approach equals traditional axiomatics (rather than nomic explanation), whereas the horizontal approach equals what he calls coherentist inductivism, easily identifiable as an application of the hermeneutic spiral:

> Here there is a definite place for a *dialectical* process of *cyclical* structure, where one returns repeatedly to an item already 'established'. For the process of confirmation is now more complex [than in the vertical approach, because we are not] proceeding linearly, by fresh deductions from

> novel premises. [. . .] This cyclic process of reappraisal is such that one can even – in suitable circumstances – dispense with the need for 'new' data-inputs in an endeavour to *squeeze more information* out of the old. (Rescher, 1979, 75; emphasis added)

3 Generative Linguistics

The relation between hermeneutics and generativism is asymmetrical insofar as the former is a general philosophy of *all* human sciences whereas the latter is a school of thought belonging to *one* particular (human) science, namely linguistics. In order to have a meaningful comparison, only such aspects of generative linguistics will be singled out here as are general enough to have some relevance for hermeneutics, as discussed in Section 2. More technical large-scale assessments of generativism have been given in Itkonen (1975) and (1996).

"My own work has just been wildly distorted, in most amazing ways" (Chomsky, 1982, 38). In order not to attract such animadversions, I prefer to give direct quotations by Chomsky (and occasionally by others) and let them speak for themselves. A)-quotations represent the orthodox positivist line whereas B)-quotations reveal some affinity with hermeneutics. Taken together, A) and B) suggest a certain amount of conceptual incoherence. Brief C)-commentaries will also be added.

3.1 Subject vs. Object

A) "Our scientist S [. . .] studies language exactly as he studies physics, taking humans to be *natural objects*" (Chomsky, 1976, 183; emphasis added). "If ['the strong minimalist thesis'] were true, language would be something like a snowflake, taking the form it does by virtue of natural [= physical] law, in which case UG would be very limited" (Chomsky 2011, 26).

B) "The problem for the grammarian is to construct a description and, where possible, an explanation for the enormous mass of unquestionable data concerning the linguistic intuition of the native speaker (often, *himself*)" (Chomsky, 1965, 20; emphasis added).

C) In sum, A: Subject → Object vs. B: Subject ↔ Object, or even Subject = Object. What we have here is an implicit contradiction: Unlike humans, snowflake-like natural objects can describe neither each other nor themselves.

3.2 Linguistic Intuition

A) "The linguist interested in constructing a general theory of linguistic structure [. . .] should try to avoid such notions as 'intuition' [. . .]" (Chomsky, [1955] 1975, 87).

"It is also quite clear that the major goal of grammatical theory is to replace this obscure reliance on intuition by some objective and rigorous approach" (Chomsky, 1957, 94).

B) "It appears then that in a certain sense the ultimate criterion remains the speaker's linguistic intuition about linguistic form, since only this can tell us which behavioural tests are to the point" (Chomsky, [1955] 1975, 104). "There is no way to avoid the traditional assumption that the speaker-hearer's linguistic intuition is the ultimate standard that determines the accuracy of any proposed grammar, linguistic theory, or operational test [...]" (Chomsky, 1965, 21).

C) A vs. B = contradiction.

3.3 Knowing-that

A) "Obviously knowledge of a language is not a matter of knowing that" (Chomsky, 1972, 190). "We must avoid the temptation to assume some notion of 'accessibility to consciousness' with regard to mental states and their contents" (Chomsky, 1986, 230). "But knowing English is not really 'knowing that' *anything*" (Jackendoff, 2002, 28).

B) "Any speaker of English *knows that* (a) *keep* and *coop* begin with the same 'sound', [...]; (b) [the relation between *see* and *sight*] is a relation similar to that between *refuse* and *refusal* [...]; (c) *Are they coming?* is the question corresponding to *They are coming*. Similarly, *Did they come?* is the question corresponding to *They came* and not (despite the similarity of morphemic constitution) to *They did come;* [...] [three additional items (d)–(f) now follow]. [We ought to be able to account] for each of these and *thousands* of other intuitions about linguistic form" (Chomsky 1975: 62; emphasis added; similarly Chomsky, 1965, 75–76).

C) Again, A vs. B = contradiction. (It is conceivable that Chomsky uses the expression 'knowing that' in two different meanings, to refer either to what is conscious or to what is unconscious. But why would he do so?)

3.4 Normativity

A) "In defining elements for each language, the linguist relates them to the physiological activities or sound waves of speech, [...]" (Harris 1951, 16). "[...] such notions as 'phoneme in L', 'phrase in L', 'transformation in L' are defined for an arbitrary language L in terms of *physical* and distributional properties of utterances of L and formal properties of grammars of L" (Chomsky, 1957, 54; emphasis added).

B) "We can [...] state methods for excluding parts of the corpus (e. g., *mistakes*, [...]) from the set of fully grammatical sentences" (Chomsky, [1955] 1975, 68; emphasis added). "[. . .] actual linguistic behaviour can easily be characterized as a special *deviation* from underlying *norms*" (Chomsky, [1955] 1975, 149, n. 21; emphasis added).

C) Mistakes are committed by speakers investigated by linguists whereas no mistakes are committed by 'natural objects' investigated by physicists. Physical data are inherently *non*-normative (cf. the Miller-quotation in 2.4). Once again, A vs. B = contradiction.

According to Putnam (1989), Chomsky has envisaged the possibility that "our idealized or 'competence' description is a description of correct thinking in the *normative* sense" (Putnam, 1989, 216; emphasis added). This is OK.

There are clear intersubjective criteria for separating mistakes from correct speech (cf. Hokkanen, 2001). Hence, contrary to a common misconception, ungrammatical data do come labelled as ungrammatical. With lots of such negative evidence around, (analogy-based) language-acquisition proves to be a child's play: "young children make analogies across whole utterances" (Tomasello, 2003, 144; similarly Itkonen, 2005, Ch. 2). The poverty of the stimulus argument vanishes into thin air.

3.5 Private vs. Public Rules

A) "As for the fact that the rules of language are public rules, this is indeed a *contingent* fact [. . .]. We thus share rules of language with others as we share an organization of visual space with them" (Chomsky, 1976, 71; emphasis added).

B) "The expressions [like *the boy elapsed*] deviate in some manner [. . .] from the *rules of English*. [. . .] There are fairly clear-cut cases of *violation* of purely syntactic *rules*, for example *sincerity frighten may boy the* [. . .]" (Chomsky 1965, 76; emphasis added). "The rules of the language are [. . .] like rules of chess" (Chomsky, 1986, 27).

C) Let us hear what I had to say 35 years ago: "In Chomsky's sense, two speakers share the same language just as two pieces of iron share the same internal structure. In this type of situation, when A and B share C, there are in fact *two* (similar) C's, *viz*. C-1 possessed by A and C-2 possessed by B. It is a matter of necessity, however, that successful communication (which Chomsky explicitly mentions) requires a stronger sense of 'sharing', i. e. the possibility of appealing, in case of doubt, to rules that are possessed *jointly*, and not just separately, by the speakers. Consequently, two speakers in reality share the same language in the sense in which they might share a secret. In this type of situation, when A and B share C, there is only one C which is possessed both by A and B. Since there is only *one* language, to which all of its speakers equally have access, it is [. . .] a *necessary* fact that its rules are public rules" (Itkonen, 1983, 228).

3.6 Functionalism

A) "To account for or somehow explain the structure of UG, or of particular grammars, on the basis of functional considerations is a pretty hopeless prospect, I would think; it is, perhaps, even 'perverse' to assume otherwise" (Chomsky, 1976, 58). "The 'instrumental' analysis of language as a device of achieving some ends is seriously inadequate" (Chomsky, 1976, 69). "[S]tructures [...] can be characterized as optimal *irrespective of any functional correlate*" (Piattelli-Palmarini and Uriagereka, 2008, 209).

B) "In §§ 3–7, then, we were studying language as an instrument or a tool, [...]" (Chomsky, 1957, 103). "[T]here are intimate connections between structure and function. How could this fail to be true? Who has ever doubted it?" (Chomsky, 1976, 235).

C) Who indeed?

3.7 Explanation vs. Explication

A) "A grammar of a language L is essentially a theory of L. Any scientific theory is based on a finite number of observations, and it seeks to relate the observed phenomena and to predict new phenomena by constructing general laws [...]" (Chomsky, 1957, 49). "In the natural sciences, the facts have no interest in themselves, but only to the degree they have bearing on explanatory principles [...]" (Chomsky, 1979, 58).

B) "We thus face a familiar task of explication of some intuitive concept – in this case, the concept 'grammatical in English', [...]" (Chomsky 1957, 13).

C) *Explication* equals conceptual analysis as practised within the tradition of analytical philosophy, epitomized by Pap (1958) (cf. Itkonen 1978, Ch. 11). *Explanation*, by contrast, applies (primarily) to physical events in space and time: it is a method explicated (*sic!*) e. g. by Hempel and Oppenheim's (1948) deductive-nomological model.

3.8 Axiomatics vs. Psychology

A) "A grammar is a device for generating sentences [...] [i. e.] a sequence of statements of the form $X - i \rightarrow Y - i$ interpreted as the instruction 'rewrite *X-i* as *Y-i*', where *X-i* and *Y-i* are strings. [...]. Call each statement of [this] form a *conversion* [...]. A derivation is roughly analogous to a proof, with *Sentence* playing the role of the single *axiom*, and the conversions corresponding roughly to rules of inference" (Chomsky, [1955] 1975, 67; the latter emphasis added). "The idea of generative grammar emerged from an *analogy* with categorial systems in logic. The idea was to treat grammaticality like theoremhood in logistic [= axiomatic] systems and to treat grammatical structure like proof structure in derivations" (Katz, 1981, 36; emphasis added; quoted and discussed in Itkonen, 2005, 19–20; on analogy in scientific discovery/invention, see also Itkonen, 2005, 176–198). "We must analyze and define the notion of *simplicity* that we intend to use in choosing among grammars" (Chomsky, 1957, 54; emphasis added). "We may tentatively define the phonemes and morphemes of a language as the tentative phonemes and morphemes which [...] jointly lead to the simplest grammar" (Chomsky, 1957, 57). "[My] new analysis of 'significance' [...] will rule out mentalism for what were essentially Bloomfield's reasons, i. e., its obscurity and inherent testability" (Chomsky, [1955] 1975, 86).

B) "Hence, in the technical sense, linguistic theory is mentalistic, [...]" (Chomsky, 1965, 4). "A fully adequate grammar must assign to each of an infinite range of sentences a structural description indicating how the sentence is understood by the *ideal speaker-hearer*" (Chomsky, 1965, 4–5; emphasis added).

C) In the 1950s, a generative grammar was defined as a non-mentalistic mathematical structure. In the 1960s, while remaining essentially the same, it was reinterpreted as a mentalistic mathematical structure, i. e. a structure supposedly embodied by the ideal speaker-hearer. Conclusion: "The 'ideal speaker' possesses no properties over and above those belonging to an axiomatic system: in fact, *the two are identical*" (Itkonen, 1976, 214; also 2011c). Next, this "ghostly" new entity started to live a life of its own: "[An adequate grammar] predicts correctly how the idealized native speaker would understand arbitrary sentences, [...]" (Chomsky, 1965, 40). But this is like predicting that a description containing the rule $S \rightarrow NP + VP$ will indeed contain the rule $S \rightarrow NP + VP$.

Additional confirmation is provided by a consideration of Pānini's grammar. Axiomatics has the goal of reducing the number of axioms, definitions, and rules of inference. Pānini pursues the same (axiomatic) goal of economy/simplicity (cf. Itkonen, 1991, 38–44). On the one hand: "Modern linguistics acknowledges [Pānini's grammar] as the most complete generative grammar of any language yet written, [...]" (Kiparsky, 1993, 2912; also Chomsky 1986, 13). On the other: "Pānini's grammar has nothing to do with psychological reality!" (Paul Kiparsky, p.c., Los Angeles, June 1982). It is a simple truth that the (untrained) human mind does *not* operate axiomatically.

3.9 Psychologism/Physicalism vs. Necessary Truth

A) "The standpoint of generative grammar is that of individual *psychology*" (Chomsky, 1986, 3; emphasis added). "[Elements of I-language] are real elements of particular minds/brains, aspects of the *physical* world, [...]" (Chomsky, 1986, 26; emphasis added).

B) “Thus I agree with Katz that certain *analytic* connections exist among linguistic expressions, certain truths hold solely by virtue of linguistic facts: for instance the relation between *I persuaded him to leave* and *He intends to leave*, [. . .]” (Chomsky, 1979, 145; emphasis added). “There exist principles that are completely linguistic. For example, in *John sees him*, *John* and *him* cannot be taken to refer to the same person, [. . .] That is a linguistic rule” (Chomsky, 1979, 146).

C) Once again, we have a contradiction. As noted by Katz and Postal (1991), the psychologism (and/or physicalism), endorsed in A), allows “no place for necessary [= analytic] connections in grammatical structure” (Katz and Postal, 1991, 521), endorsed in B). They correctly add that “[no] other version of conceptualism [= psychologism] can escape the defects of Chomsky’s version” (Katz and Postal, 1991, 550). This judgment applies e. g. to the Langacker-type *cognitive* grammar (cf. Itkonen, 2016).

3.10 Autonomy of Syntax

A) “[. . .] meaning was not used as a criterion in setting up the morphemic segments, [. . .]” (Harris, 1951, 173). “Grammar is best formulated as a self-contained study independent of semantics” (Chomsky 1957, 106). “It seems to me that [the thesis of the autonomy of syntax] has held up pretty well. I don’t see any reason to question it right now” (Chomsky, 1982, 115).

B) “There are striking correspondences between the structures [. . .] discovered in formal, grammatical analysis and specific semantic functions”. “[. . .] some [of the six pro-semantic claims] are very nearly true” (Chomsky, 1957, 101).

C) Significant concessions are made in B), but they are far from enough. In hindsight, it is easy to see that, from the start, we have been dealing here with a fallacy.

Lexical forms are *non*-autonomous vis-à-vis corresponding meanings to the point that the two types of entities are designated by one and the same sign, with only typographical variation: the form *book* expresses the meaning ‘book’.

The autonomy of forms is even more unnatural in the *grammatical* domain. Consider the Latin verb form *ama-ba-s* ‘you loved’. We say that (considered as a word, and not as a sentence) it is active indicative imperfect second-person singular. This is its *formal* characterization. Why? Because it expresses the following grammatical *meanings*: ‘active’, ‘indicative’, ‘imperfect’, ‘second-person singular’. Considered as a sentence-form, it is also an affirmative assertion. Why? Because it expresses the *meanings* ‘affirmative’ and ‘assertion’.

Against this background, it makes no sense at all to initiate “a purely negative discussion of the possibility of finding semantic foundations for syntactic theory” (Chomsky, 1957, 93). Chomsky mentions, in particular, two reasons for doing so. First, as he sees it, it is not true that “the grammatical relation subject–verb corresponds to the general 'structural meaning' actor–action" (Chomsky, 1957, 94). Second, it is not true that "the grammatical relation verb-object corresponds to the structural meaning action-goal or action-object of action" (Chomsky, 1957, 94). These claims are supported by examples like *the fighting stopped* (= no action), on the one hand, and *I missed the train* (= no object of action), on the other.

What we have here is a falsity wrapped inside a confusion. First, no sane person has ever claimed that all verbs express genuine actions. Second, what is it, precisely, that we reject when we reject meanings? Meanings of *what*? Grammatical morphemes or words or word-classes or sentence-constituents (*Satzglied*)? No coherent notion of semantics ensues from conflating these four very different entities.

4 Hermeneutic Reception of Generativism

For Habermas (1970), there is no language without norms, and no norms without norm-violating behaviour (*normwidriges Verhalten*). Being the very opposite of natural laws (*Naturgesetze*), "norms are always based on mutual agreement" (Habermas, 1970, 160). He thinks generativism will soon come to accept these self-evident truths.

For Apel (1976c), a grammar is above all "the reconstruction of a norm competence" (Apel, 1976c, 277) or "of a norm awareness" (Apel, 1976c, 285). He assumes that Chomsky agrees, given his constant use of such openly normative terms as *grammaticality* and *acceptability* (Apel 1976c, 285). This would be the natural interpretation.

Let us recall that the primary concern of hermeneutics is *meaning*. Habermas and Apel cannot help finding the emphasis on linguistic *form* just a temporary aberration or, to put it more politely, *one-sidedness* (*Einseitigkeit*). The generative theory needs to be extended so as to encompass semantics and pragmatics, in addition to syntax. The end result is supposed to be a theory of *communicative competence*, i. e. a competence possessed by the members not just of every actual *Kommunikationsgemeinschaft*, but also of the *ideal* one.

5 Conclusion

Genuine science is value-free: facts must be taken as they come. Yet in the controversy between hermeneutics and positivism it may be easy to forget this fundamental truth. It is not uncommon that one position comes to be regarded as somehow *better* or *more worthwhile* than the other. Let us illustrate.

"The doctrine of *Verstehen* is hermeneutics" (Sombart, [1929] 1972, 171). "I defend the position that the epistemological monism of natural science must be contrasted with epistemological pluralism" (172). "[What follows is] the superiority of that type of knowledge which is characteristic of the human sciences, the superiority of *Verstehen*" (173). Notice the affinity with the Vico's *verum* = *factum* doctrine (cf. 2.1).

"For me the field would completely lose interest, I would see no point in studying grammar, if it did not turn out that there were a unifying, rather abstract explanatory theory" (Chomsky, 1982, 86). And, as noted in 3.4 A) & 3.7 A), this would by definition be a *physical* theory. In other words, if positivism in linguistics proves to be false, Chomsky rejects this result. Why? Because he does not find it worthwhile.

References

Arens, Hans. 1969. *Sprachwissenschaft: Der Gang ihrer Entwicklung von der Antike bis zur Gegenwart, Band 1*. Frankfurt/Main: Athenäum.

Anderson, Robert John, John A. Hughes & Wes W. Sharrock. 1986. *Philosophy and the human sciences*. London: Croom Helm.

Anttila, Raimo. 1989. Pattern explanation: The survival of the fit. *Diachronica* 6. 1–21.

Apel, Karl-Otto. 1976a. Die Entfaltung der 'sprachanalytischen' Philosophie und das Problem der 'Geisteswissenschaften. Karl-Otto Apel, *Transformation der Philosophie, Band 2: Das Apriori der Kommunikationsgemeinschaft*, 28–95. Frankfurt/Main: Suhrkamp.

Apel, Karl-Otto. 1976b. Szientistik, Hermeneutik, Ideologiekritik: Entwurf einer Wissenschaftslehre in erkenntnisanthropologischer Sicht. Karl-Otto Apel, *Transformation der Philosophie, Band 2: Das Apriori der Kommunikationsgemeinschaft*, 96–127. Frankfurt am Main: Suhrkamp.

Apel, Karl-Otto. 1976c. Noam Chomskys Sprachtheorie und die Philosophie der Gegenwart. Karl-Otto Apel, *Transformation der Philosophie, Band 2: Das Apriori der Kommunikationsgemeinschaft*, 264–310. Frankfurt am Main: Suhrkamp.

Apel, Karl-Otto. 1981. Die beiden Phasen der Phänomenologie in ihrer Auswirkung auf das philosophische Vorverständis von Sprache und Dichtung in der Gegenwart. Karl-Otto Apel, *Transformation der Philosophie, Band 1: Sprachanalytik, Semiotik, Hermeneutik*. 2nd edn., 79–105. Frankfurt am Main: Suhrkamp.

Berlin, Isaiah. 1976. *Vico and Herder: Two studies in the history of ideas*. London: Chatto & Windus.

Berlin, Isaiah. 1980 [1960]. The concept of scientific history. Isaiah Berlin, *Concepts and categories: Philosophical essays*, 103–142. Oxford: Oxford University Press.

Böhme, Gernot, ed. 1976. *Protophysik*. Frankfurt am Main: Suhrkamp.

Bühl, Walter L., ed. 1972. *Verstehende Soziologie: Grundzüge und Entwicklungstendenzen*. München: Nymphenburger Verlagshandlung.

Chafe, Wallace. 1994. *Discourse, consciousness, and time*. Chicago: The University of Chicago Press.

Chomsky, Noam. 1957. *Syntactic structures*. Mouton: The Hague.

Chomsky, Noam. 1965. *Aspects of a theory of syntax*. Cambridge Mass.: The MIT Press.

Chomsky, Noam. 1972. *Language and mind*. Enlarged edition: San Diego: Harcourt Brace Jovanovich, Publishers.

Chomsky, Noam. 1975 [1955]. *The logical structure of linguistic theory*. New York: Plenum Press.

Chomsky, Noam. 1976. *Reflections on language*. Glasgow: Fontana.

Chomsky, Noam. 1979. *Language and responsibility*. New York: Pantheon Books.

Chomsky, Noam. 1982. *The generative enterprise*. Dordrecht: Foris.

Chomsky, Noam. 1986. *Knowledge of language: Its nature, origin, and use*. New York: Praeger.

Chomsky, Noam. 2011. Opening remarks. In Massimo Piattelli-Palmarini, Pello Salaburu & Juan Uriagereka (eds.), *Of minds and language: A dialogue with Noam Chomsky in the Basque country*, 13–32. Oxford: Oxford University Press.

Cohen, Jonathan L. 1986. *The dialogue of reason: An analysis of analytical philosophy*. Oxford: Clarendon Press.
Collingwood, R. G. 1946. *The idea of history*. Oxford: Clarendon Press.
Davidson, Donald. 1975. Thought and talk. In Samuel Guttenplan (ed.), *Mind & language*, 7–24. Oxford: Oxford University Press.
Dennett, Daniel C. 2014. *Intuition pumps and other tools for thinking*. London: Penguin Books.
Diesing, Paul. 1972. *Patterns of discovery in the social sciences*. London: Routledge.
Dilthey, Wilhelm. 1914. *Gesammelte Schriften*. G. Misch et al. (eds.), Vol. 1. Leipzig: B.G. Teubner.
Dilthey, Wilhelm. 1927. *Gesammelte Schriften*. G. Misch et al. (eds.), Vol. 7. Leipzig: B.G. Teubner.
Dray, William. [1963] 1974. The historical explanation of actions reconsidered. In Patrick Gardiner (ed.), *The philosophy of history*, 66–89. Oxford: Oxford University Press.
Dray, William. 1980. R. G. Collingwood and the understanding of actions in history. In William Dray, *Perspectives on history*, 9–26. London: Routledge.
Edgley, R. [1965] 1978. Practical reason. In Joseph Raz (ed.), *Practical reasoning*, 18–32. Oxford: Oxford University Press.
Frege, Gottlob. 1893. *Grundgesetze der Arithmetik I–II*. Jena: Hermann Pohle.
Frege, Gottlob. [1918] 1967. Thought. In P. F. Strawson (ed.), *Philosophical logic*, 1–31. Oxford: Oxford University Press.
Gibson, Q. 1976. Arguing from rationality. In S. I. Benn & G. W. Mortimore (eds.), *Rationality and the social sciences*, 111–131. London: Routledge.
Giddens, Anthony. 1976. *New rules of sociological method*. London: Hutchinson.
Habermas, Jürgen. 1968. *Erkenntnis und Interesse*. Frankfurt/Main: Suhrkamp.
Habermas, Jürgen. 1970. *Zur Logik der Sozialwissenschaften*. Frankfurt/Main: Suhrkamp.
Habermas, Jürgen & Niklas Luhmann. 1971. *Theorie der Gesellschaft oder Sozialtechnologie*. Frankfurt am Main: Suhrkamp.
Harris, Zellig S. 1951. *Structural linguistics*. Chicago: The University of Chicago Press.
Hempel, Carl G. 1965. *Aspects of scientific explanation and other essays in the philosophy of science*. New York: Free Press.
Hempel, Carl G. & Paul Oppenheim. 1948. Studies in the logic of explanation. *Philosophy of Science* 15. 135–175.
Hintikka, Jaakko. 1974. *Knowledge and the known: Historical perspectives in epistemology*. Dordrecht-Holland: Reidel.
Hokkanen, Tapio. 2001. *Slips of the tongue: Errors, repairs, and a model*. (Studia Fennica Linguistica 10).
Hollis, Martin. 1977. *Models of man*. London: Cambridge University Press.
Husserl, Edmund. 1913. *Logische Untersuchungen I–II*, 2nd edn. Tübingen: Max Niemeyer.
Itkonen, Esa. 1974. *Linguistics and metascience*. (Studia Philosophica Turkuensia II).
Itkonen, Esa. 1975. Transformational grammar and the philosophy of science. In E. F. K. Koerner (ed.), *The transformational-generative paradigm and modern linguistic theory*, 381–445. Amsterdam: Benjamins.
Itkonen, Esa. 1976. The use and misuse of the principle of axiomatics in linguistics. *Lingua* 38. 185–220.
Itkonen, Esa. 1978. *Grammatical theory and metascience*. Amsterdam: Benjamins.
Itkonen, Esa. 1981a. The concept of linguistic intuition. In Florian Coulmas (ed.), *A festschrift for native speaker*, 127–140. Mouton: The Hague.
Itkonen, Esa. 1981b. Review article on Lass (1980). *Language* 57. 688–697.
Itkonen, Esa. 1983. *Causality in linguistic theory*. London: Croom Helm.
Itkonen, Esa. 1991. *Universal history of linguistics: India, China, Arabia, Europe*. Amsterdam: Benjamins.

Itkonen, Esa. 1996. Concerning the generative paradigm. *Journal of Pragmatics* 25. 471–501.
Itkonen, Esa. 2000. *Tolkaappiyam*: The basic work of ancient Tamil language and culture. *SKY Journal of Linguistics* 13. 75–99. The Yearbook of the Linguistic Association of Finland.
Itkonen, Esa. 2003a. *Methods of formalization beside and inside both autonomous and non-autonomous linguistics* (Publications in General Linguistics 6). University of Turku.
Itkonen, Esa. 2003b. *What is language? A study in the philosophy of linguistics* (Publications in General Linguistics 8). University of Turku.
Itkonen, Esa. 2004. Iconicity and typological explanation. *Logos & Language* V(1). 21–33.
Itkonen, Esa. 2005. *Analogy as structure and process: Approaches in linguistics, cognitive psychology, and philosophy of science*. Amsterdam: Benjamins.
Itkonen, Esa. 2008a. Concerning the role of consciousness in linguistics. *Journal of Consciousness Studies* 15(6). 17–33.
Itkonen, Esa. 2008b. The central role of normativity in language and linguistics. In Jordan Zlatev, Timothy P. Racine, Chris Sinha & Esa Itkonen (eds.), *The shared mind: Perspectives on intersubjectivity*, 279–305. Amsterdam: Benjamins.
Itkonen, Esa. 2011b. Concerning (radical) construction grammar. In Esa Itkonen, *Papers on typological linguistics* (Publications in General Linguistics 15). 62–75. University of Turku.
Itkonen, Esa. 2011c. Prolegomena to any discussion of simplicity vs. complexity in linguistics. In Esa Itkonen, *Papers on typological linguistics* (Publications in General Linguistics 15). 140–152. University of Turku.
Itkonen, Esa. 2011d. On the role of induction in typological linguistics. In Esa Itkonen, *Papers on typological linguistics* (Publications in General Linguistics 15), 177–190. University of Turku.
Itkonen, Esa. 2013. Functional explanation and its uses. In Shannon T. Bischoff & Carmen Jany (eds.), *Functional approaches to language*, 31–69. Berlin: DeGruyter.
Itkonen, Esa. 2013–2014. On explanation in linguistics. *Energeia* V, 10–40.
Itkonen, Esa. 2016. An assessment of (mentalist) cognitive semantics. *Public Journal of Semiotics* 7(1). 1–42.
Jackendoff, Ray 2002. *Foundations of language*. Oxford: Oxford University Press.
Kaplan, Abraham. 1964. *The conduct of inquiry*. San Francisco: Chandler.
Katz, Jerrold J. 1981. *Language and other abstract objects*. London: Blackwell.
Katz, Jerrold J. & Paul M. Postal. 1991. Realism vs. conceptualism in linguistics. *Journal of Philosophy* 14. 515–554.
Kiparsky, Paul. 1993. Paninian linguistics. In R. E. Asher (ed.), *The encyclopedia of language and linguistics. Vol. 1(6)*, 2918–2923. Oxford: Pergamon Press.
Kleene, Stephen Cole. 1952. *Introduction to metamathematics*. Princeton, N.J.: Van Nostrand.
Koerner, E. F. K., ed. 1975. *The transformational-generative paradigm and modern linguistic theory*. Amsterdam: Benjamins.
Koestler, Arthur. 1967. *The ghost in the machine*. London: Hutchinson.
Lass, Roger. 1980. *On explaining language change*. Cambridge: Cambridge University Press.
Lass, Roger. 1997. *Historical linguistics and language change*. Cambridge: Cambridge University Press.
Lorenzen, Paul. 1969. *Systematisches Denken*. Frankfurt am Main: Suhrkamp.
Meillet, Antoine. 1921. *Linguistique historique et linguistique générale*. Paris: H. Champion.
Newton-Smith, W. H. 1981. *The rationality of science*. London: Routledge.
Pap, Arthur. 1958. *Semantics and necessary truth: An inquiry into the foundations of analytic philosophy*. New Haven: Yale University Press.
Piattelli-Palmarini, Massimo & Juan Uriagereka. 2008. Still a bridge too far? Biolinguistic questions for grounding language on brains. *Physics of Life Reviews* 5. 207–224.

Putnam, Hilary. 1989. Reflexive reflections. In Stuart Silvers (ed.), *Rerepresentation*, 210–232. Dordrecht: Kluwer.
Radnitzky, Gerard. 1970. *Contemporary schools of metascience*. 2nd edn. Göteborg: Akademiförlaget.
Rescher, Nicholas. 1979. *Cognitive systematization*. Oxford: Blackwell.
Russell, Bertrand. 1967 [1940]. *An inquiry into meaning and truth*. Harmondsworth: Pelican Books.
Sampson, Geoffrey. 1975. Review of Koerner (ed.) 1975. *Language* 52(4). 584–601.
Schleiermacher, Friedrich. 1977. *Hermeneutik und Kritik*. M. Frank (ed.), Frankfurt/Main: Suhrkamp.
Searle, John. 1969. *Speech acts: An essay in the philosophy of language*. Cambridge: Cambridge University Press.
Sombart, Werner. 1929 [1972]. Sinnverstehen – Sachverstehen – Seelverstehen. In Walter L. Bühl (ed.), *Verstehende Soziologie: Grundzüge und Entwicklungstendenzen*, 171–187. München: Nymphenburger Verlagshandlung.
Tomasello, Michael. 2003. *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
Talmy, Leonard. 2000. *Towards a cognitive semantics*. Cambridge, MA: The MIT Press.
Van Bekkum, Wout, Jan Houben, Ineke Sluiter & Kees Versteegh. 1997. *The emergence of semantics in four linguistic traditions: Hebrew, Sanskrit, Greek, Arabic*. Amsterdam: Benjamins.
Winch, Peter. 1958. *The idea of a social science and its relation to philosophy*. London: Routledge.
Wittgenstein, Ludwig. 1958. *Philosophical investigations*, 2nd edn. Oxford: Blackwell.
Wittgenstein, Ludwig. 1969. *On certainty*. Oxford: Blackwell.
von Wright, Georg Henrik. 1963. *Norm and action*. London: Routledge.
von Wright, Georg Henrik. 1971. *Explanation and understanding*. London: Routledge.
Zlatev, Jordan. 2008. The dependence of language on consciousness. *Journal of Consciousness Studies* 15(6). 34–62.

András Kertész and Csilla Rákosi

17 The uncertainty of syntactic theorizing

Abstract: In this paper, we present a metatheoretical model and show how it serves to capture some basic properties of generative linguistic theorizing. By 'generative linguistic theorizing', we mean actual research practice, i. e. research activities performed while elaborating, putting forward, applying, and testing particular versions of generative linguistics. We call our metatheoretical approach the p-model of plausible argumentation. With the help of the p-model, we will show that the way in which generative linguistics has proceeded over the past decades and is functioning today differs significantly from the self-image of generative linguists as documented in the literature. Thus, we will show that generative linguistic research is built on uncertainty rather than certainty; it tolerates inconsistency instead of defending consistency at all costs; the structure of its theories is not static, but dynamic; and its development is not linear, but cyclic and prismatic.

1 Introduction[1]

In (1) we quote *Syntactic Structures* in which Chomsky characterizes a grammar in the following way:[2]

(1) a. A grammar of the language L is essentially a *theory* of L.
b. Any scientific theory is based on a finite number of *observations*, and it seeks to relate the observed phenomena and to *predict* new phenomena by constructing *general laws* in terms of *hypothetical constructs* such as (in physics, for example) 'mass' and 'electron'.
c. Similarly, a grammar of English is based on a finite corpus of utterances (observations), and it will contain certain grammatical rules (laws) stated in terms of the particular phonemes, phrases, etc., of English (hypothetical constructs). These rules express structural relations among the sentences of the corpus and the indefinite number of sentences generated by the grammar beyond the corpus (predictions). Our problem is to develop

1 Two terminological remarks should be made in order to avoid any misunderstanding of the subject matter of this chapter. First, the term 'generative linguistics' is ambiguous because it may refer to a series of theories, approaches or models, some of which compete with each other. Second, we do not go into the discussion of whether particular versions of generative linguistics are 'theories' or 'programs', etc. We will see that from the point of view of the p-model these notions are not crucial.

2 We have subdivided this quotation into units in order to make references to relevant parts of the quotation easier.

https://doi.org/10.1515/9783110540253-017

> and clarify the criteria for selecting the correct grammar for each language, that is, the correct theory of this language. (Chomsky, 1957, 49; emphasis added)

Although Chomsky does not cite the literature which the ideas summarized in this passage are based on, it goes without saying that the terms introduced in (1b) have been borrowed from the standard view of the philosophy of science. Basically, the standard view of the philosophy of science consisted of two main branches, namely *logical positivism* (whose most prominent personality was Rudolf Carnap and which was based on the inductive method) and Popperian *falsificationism* focusing on the deductive testing of scientific theories (for an overview of the standard view see e. g. Hung, 2014, 311). Besides their differences, these two branches share a set of common features, the most important of which is the assumption that there are criteria of rationality that any kind of scientific inquiry should meet. (1b) is a concise summary of the picture which the Popperian deductive branch of the standard view of the philosophy of science drew of successful physical theories. (1a) and (1c) conceive of grammars in analogy to this account of scientific theories. Nevertheless, as later developments witness, the relation between grammars and scientific theories in this sense seems to be quite challenging, both for linguists themselves and for philosophers of science as well for at least the following reasons:

(i) A great amount of the literature reflecting on the methodology of generative linguistics questions the successful realization of the project illustrated by (1). For example, Behme's (2014, 672) analysis concludes that Chomsky's "recent work fails to meet serious scientific standards because he rejects scientific procedure, inflates the value of his own work, and distorts the work of others [...]." Sampson goes a step further and assumes that "[...] linguistics will not move forward healthily until the generative approach is bypassed as *the pseudoscience it is*, with its remaining practitioners retiring from academic life and new recruits to the discipline ignoring its ethos, assumptions, and alleged results" (Sampson, 2007, 122; emphasis added).

(ii) In contrast, other linguists still argue for the claim that the current version of generative linguistics "is well on its way to becoming a full-blown natural science, offering a serious promise of an advanced field of scientific inquiry whose idealizations, abstractions, and deductions will eventually *match in depth and subtlety those of the most advanced domains of modern science*" (Piattelli-Palmarini, 1998, xxv; emphasis added).

(iii) While the standpoints in (i) and (ii) presuppose its correctness, the standard view has become obsolete in the current state of the art of the philosophy of science. Primarily as a result of Kuhn's and Lakatos' impact, today there is general consensus among philosophers of science that the standard view is unrealistic insofar as it is incapable of describing the history of science because no scientific theory put forward so far fulfils its norms. Therefore, it cannot be applied as a guide for the elaboration of new theories, either.

Comparing Chomsky's project as illustrated in (1) with the evaluations as exemplified above in (i)–(iii), the question arises:

(Q) What is the nature of generative linguistic theorizing?

It is this question which the present chapter focuses on. Nevertheless, we will take sides neither for the Chomskyan nor the anti-Chomskyan stance. Rather, we will outline a metatheoretical model which is crucially different from the standard view and which might be capable of capturing some basic properties of generative linguistic theorizing in an *unbiased and balanced* manner. By 'generative linguistic theorizing', we mean *actual research practice*, i. e. the research activities performed while elaborating, putting forward, applying, and testing particular versions of generative linguistics. We call our metatheoretical approach *the p-model* of plausible argumentation.[3]

The p-model of plausible argumentation rejects, in harmony with current trends in the philosophy of science, a series of methodological prejudices which mainstream linguistics takes implicitly or explicitly for granted, but which are at variance with the research practice. The p-model answers the question (Q) by accounting for, among others, the following properties of generative linguistic theorizing:

(P1) Generative linguistic theorizing is *based on the uncertainty of information* instead of its certainty.

(P2) Generative linguistic theorizing *tolerates inconsistency* instead of requiring consistency at all costs.

(P3) Generative linguistic theorizing is a *dynamic*, *cyclic and prismatic process* instead of being based on a static logical structure and resulting from a linear development.

In Section 2 we will elaborate on the properties (P1)–(P3) and put forward a series of theses which will jointly yield our answer to the question (Q). In order to keep our train of thought within reasonable limits and to show the overall workability of the p-model, in Section 3 we will illustrate the notions and the claims we introduce by a case study. We have chosen Zubizarreta's (1982) analysis of Spanish modal verbs which in the 1980s was considered a valuable contribution to Government-Binding Theory and which is, on the one hand, well suited to exemplify certain relevant features of generative linguistics. On the other hand, since Zubizarreta's analysis is out of date now, we will not be compelled to take sides in the acceptance or rejection of Government-Binding Theory as well as Zubizarreta's analysis; such a debate would blur the focus of a balanced metatheoretical approach. Finally, in Section 4 we will summarize our stance by comparing the p-model's theses to other metatheoretical approaches.

3 See Kertész and Rákosi (2012) for a detailed presentation of the p-model and Kertész and Rákosi (2014) for its concise summary.

2 The p-model of generative linguistic theorizing

2.1 On property (P1): The uncertainty of information

2.1.1 Plausible statements

The literature that evaluates generative linguistic theorizing unanimously highlights the changing nature of its hypotheses, basic terms, and the details of the theoretical framework. Proponents of generative linguistics evaluate these changes as a progressive feature that speaks for its continuous development and renewal. In contrast, its opponents criticize it for not even being able to reveal a single rule or principle or term or anything else that has lasted over the six decades of its history. The motivation for these extreme evaluations is that the self-image of generative grammar is still based on the standard view of the philosophy of science. This means that, on the one hand, the search for the firm empirical basis of generative linguistic theories presupposed the certainty of data, the truth of the hypotheses, and the consistency of the theory; on the other hand, the practice of generative linguistic theorizing departs from these features insofar as most hypotheses are not statements the truth of which is guaranteed by empirical evidence or theoretical considerations. Therefore, one central task of any metatheoretical reflection on the nature of generative linguistic theorizing is to capture its *uncertainty*. However, although uncertainty is without doubt one of the constitutive properties of generative linguistic theorizing, the p-model does not deem this a shortcoming, but sets out to reconstruct it and to reveal the consequences it leads to.

The p-model explicates the notion of 'uncertainty' as 'plausibility'. Thus, we claim:

(T) a. The hypotheses which generative linguistics put forward – for example, rules, principles, constraints – are plausible statements.

A *plausible statement* is a pair consisting of an *information content* and a *plausibility value*. The plausibility value of a statement shows that one is ready to accept the statement on the basis of *sources* that support it and that are considered to be reliable to some extent; that is, the plausibility value of a statement is *source-dependent*. This means that a statement may be very plausible according to one source, and less plausible, or implausible, with respect to others. For example, the plausibility value of a statement asserting the grammaticalness of a given sentence of English may be judged differently depending on how reliable one considers the native speaker's intuition as a source. Those who accept, for instance, the view outlined in Sampson (2007) and Sampson and Babarczy (2014), would consider the intuition of the native speaker an unreliable source and would assign the statement claiming the grammaticality of a sentence a very low plausibility value or even no plausibility value at all.[4] In con-

4 "If linguistics is indeed based on intuition, then it is not a science [...]. Science relies exclusively on the empirical." (Sampson, 1975, 60)

trast, generative linguists in the eighties and nineties would consider the intuition of the native speaker to be a highly reliable source which assures the high plausibility value of grammaticality judgments.[5] And vice versa: linguists trusting intuition assign statements based on statistical data stemming from corpora a low plausibility value,[6] whereas corpus linguists consider them highly plausible.

It may also happen that some sources support the negation of the given statement and make it *implausible*. Further, if several sources support a statement, then its plausibility value is higher on the basis of all sources together than its plausibility value on the basis of any of the sources alone.[7]

It is important to bear in mind that plausibility and (subjective or objective) probability are completely different concepts. For example, low plausibility values do not mean improbability. Even in such cases the source votes for the given hypothesis, although it provides only a weak argument for its acceptance. If a source is against a hypothesis, then it makes its negation plausible.[8]

To sum up, the plausibility value of the statement p on the basis of the source S is such that:

(a) $|p|_S = 1$, if p is true with certainty on the basis of S;
(b) $|p|_S = 0$, if p is of neutral plausibility on the basis of S, i. e., if it is neither plausible nor implausible on the basis of this source (we can accept neither p nor its negation on the basis of S);
(c) $0 < |p|_S < 1$, if p is plausible on the basis of S (i. e., we are ready to accept p on the basis of S);
(d) $0 < |{\sim}p|_S < 1$, if p is implausible on the basis of S (S provides support for the negation of p – we can accept p's negation on the basis of S).

5 For example, according to Jackendoff (1994, 48), introspection "is *so reliable* that, for a very good first approximation, linguists tend to trust their own judgments and those of their colleagues" (emphasis added).

6 For example, Chomsky declares that "corpus linguistics doesn't mean anything" (Andor, 2004, 97). Or, to quote Pullum, who in other respects is very critical of Chomsky's publications: "[...] corpus linguistics based on huge corpora has been gaining popularity as a methodology for syntax. But it will fail to have the effect it should on theoretical linguistics if its adherents fall into the extreme 'everything-is-correct' trap. [...] What, for example, is the point of the tendency seen in some grammars of the last ten years toward using corpus-derived examples for illustration? Why is it that some grammarians seem to think that every example in a reference grammar should come from a corpus? It seems to me it is not even generally desirable, let alone fully feasible." (Pullum, 2017, 284)

7 Rescher (1976) and Kertész and Rákosi (2012) represent plausibility values numerically. However, these numerical values do not have any exact meaning which could be applied in every context, but merely indicate different *relative strengths* of plausibility (reliability, supportedness, acceptance) within one theory (argumentation process). The minimum value of plausibility is indicated by 0 which means that the given hypothesis is of neutral plausibility on the basis of the source at issue. The maximal value is indicated by 1, meaning that the statement in question is true with certainty on the basis of the given source. For the sake of simplicity of exposition, in the present chapter we will not differentiate plausibility ratings numerically.

8 For details, see Kertész and Rákosi (2012, 66) and Rescher (1976, Chapter IV).

We distinguish between direct and indirect sources. In the case of *direct sources* the plausibility of the statement at issue is evaluated with respect to the reliability of the source. Such direct sources are, for example, the linguistic intuition of native speakers, corpus data, whole publications, or even the authority of one person. With *indirect sources*, however, the plausibility value of the given statement is determined with reference to the plausibility of other statements – the paradigm cases of indirect sources are inferences, to which we turn next.

2.1.2 Obtaining new information from uncertain information: plausible inferences

Only deductive inferences the premises of which are true with certainty are capable of guaranteeing the truth of the conclusion. However, in generative linguistic argumentation, very often either there is no logical consequence relation between the premises and the conclusion (that is, the inference is not deductively valid), or at least one of the premises, instead of being certainly true, is only plausible in the light of the given sources. Moreover, the combination of these two cases may also occur. Although such inferences are not capable of securing the truth of the conclusion, they may – under appropriate circumstances – make the conclusion *plausible*. The next sub-thesis is (Tb):

(T) b. The prototypical cases of the inferences applied in generative linguistic theorizing are plausible inferences.

Plausible inferences are mappings associating a set of plausible or true statements (premises) with a plausible statement (conclusion). In such inferences the connection between the premises and the conclusion cannot be reduced to the relationship between their logical structures, as is the case with deductive inferences. Plausible inferences take into consideration not only the logical structure of the premises and the conclusion but their plausibility values and semantic structure as well. They always rest on *a semantic relation*: for example, causality, analogy, similarity, sign, necessary or sufficient condition, part-whole relation, etc.

The first group of plausible inferences consists of cases in which there is a logical consequence relation between the premises and the conclusion, but at least one of the premises is only plausible and not true with certainty.[9] In such cases, the conclusion is plausible on the basis of this inference as a source.[10]

9 More precisely, a logical consequence relation and some semantic relation hold between the premises and the conclusion, the premises constitute a consistent set of statements, and all premises have a positive plausibility value (that is, they are either plausible or true with certainty) on the basis of some set of sources *S*, while at least one of them is not true with certainty.

10 As for the connection between the plausibility value of the premises and conclusion, see Kertész and Rákosi (2012, Section 9.5.5).

In *the second group* of plausible inferences there is no logical consequence relation between the premises and the conclusion, and the premises are certainly true or at least plausible with respect to some set of reliable sources.[11] Such inferences are called enthymematic. In enthymematic inferences, the set of the premises has to be complemented by *latent background assumptions* (Rescher, 1976, 60–63; Polya, 1948, 223). These background assumptions have to be true with certainty, plausible, or at least not known to be implausible or false with certainty according to some source so that they make it possible to transform the inference to a plausible inference belonging to the first group. Of course, the conclusion is only plausible and not true with certainty.

Plausible inferences are *fallible*. At the outset, their conclusion is not true with certainty but only plausible to some extent; that is, the conclusion gets only partial support from the premises. Plausible inferences with latent background assumptions are especially liable to mislead because if one of the latent background assumptions is implausible or false, then the inference may be insufficient to establish the plausibility of the conclusion even when the premises are true or plausible. Moreover, taking into consideration a wider scope of sources, one may come into possession of information that makes the premises, the latent background assumptions or the conclusion implausible or false. Therefore, it is always the case that the conclusion is plausible only *relative to* the premises, the latent background assumptions and the sources supporting them. Accordingly, since the supposed rules, principles, constraints and the like that have been put forward during the history of generative linguistics have been obtained as conclusions of plausible inferences, and plausible inferences are fallible, the p-model claims:

(T) c. The hypotheses of generative linguistics are fallible.

2.1.3 The p-context and its informational over- or underdetermination

The above characterization of plausible inferences suggests that the relation between the premises and the conclusion cannot be reduced to their formal properties. Beyond their logical structure, we have to take into consideration all pieces of information that may be relevant for judging the plausibility value of the premises and latent background assumptions. This motivates the introduction of the notion of *p-context*, which serves as the background against which plausible inferences can be put forward, used and evaluated. The p-context includes, first, a set of sources

11 More precisely, the statements playing the role of the premises are consistent, they are plausible or true with certainty according to a set of sources, and a semantic relation can be reconstructed that connects them with the conclusion – but there is no logical consequence relation between them and the conclusion.

in terms of which the plausibility value of statements can be judged. For example, the linguistic theory may make use of the following sources in order to determine the plausibility of statements pertaining to the acceptability of sentences: native speakers' linguistic intuitions collected in an experiment, the linguist's own linguistic intuition, written or spoken corpora, well-designed experiments, online corpora, a particular tree-bank etc. Second, it covers a set of statements together with their plausibility values with respect to the sources in the p-context, and their logical and semantic structure. This means that the plausibility of each statement (datum, hypothesis, conjecture, prediction, etc.) has to be made explicit. Third, the p-context also involves the accepted methodological norms related to its components.

It may happen that in a p-context the sources yield too much information, in the sense that there is a statement which is made plausible by some source while its negation is made plausible by another. In such cases, the p-context is informationally *overdetermined* (Rescher, 1976, 2; Rescher and Brandom, 1980, 3–4) and the set of the plausible statements in the p-context is *p-inconsistent*. A typical case of overdetermination occurs when linguists reject a hypothesis on the basis of their linguistic intuition, while psycholinguistic experiments provide support for it – as discussed, for example, in Featherston (2007).

Nevertheless, the p-context may be informationally *underdetermined* as well (Rescher and Brandom, 1980, 3–4). A typical case of the informational underdetermination of the p-context is its *p-incompleteness*, insofar as there are statements which are neither plausible (in the extreme case: true with certainty) nor implausible (in the extreme case: false with certainty) with respect to any source given. This is the case if we are not capable of finding a reliable source which could make it possible to assign a plausibility value to a statement or to its negation.

A p-context may be simultaneously informationally under- and overdetermined with respect to different statements. We call instances of p-inconsistency and p-incompleteness *p-problems*. If a p-context is characterised by over- and/or underdetermination in this sense, then it is called *p-problematic*.

2.1.4 'Data' and 'evidence' as plausible statements

The p-model suggests that data are not 'examples' such as *Wen*$_i$ *meint Lydia, liebt Jakob* t_i*?* but that their structure consists of two components: a statement capturing an information content and a plausibility value (see also Rescher, 1979, 69):

(D) A datum is a plausible statement with a positive plausibility value originating from some direct source.

For example:

(2) $0 <$ |The sentence *Wen$_i$ meint Lydia, liebt Jakob t_i?* is acceptable in German.|$_S < 1$[12]

Here *S* may be an experiment as a direct source. The first idea motivating the above definition is that the data sources used in linguistics are not completely reliable but have their own weaknesses and limitations.[13] Nonetheless, statements which are of neutral plausibility or implausible according to some source in the p-context, do not qualify as data. The second idea behind (D) is that data are 'given' in a specific sense: their initial plausibility is determined not with the help of plausible inferences, but directly on the basis of the reliability of their direct source. Thus, they function as *starting points*: plausibility values may enter the argumentation process through them – and these plausibility values can be transferred to other hypotheses of the theory by plausible inferences. Since data are defined as statements possessing a positive plausibility value, and the plausibility value of statements depends on the p-context, data are p-context-dependent.[14] Thus, our next thesis is:

(T) d. In generative linguistics, data are plausible statements with a positive plausibility value originating from some direct source.

The p-model defines three types of evidence in order to grasp the relationship between data and the hypotheses of the theory.

Weak evidence for a hypothesis *h* simply means that we can build plausible inference(s) making use of the given datum as a premise that make(s) *h* plausible (in the extreme case true with certainty). A datum can be weak evidence for a statement and its rival simultaneously, although the strength of the support that it provides them may differ. *Relative evidence* for a hypothesis *h* also requires that the datum provides stronger support to *h* than to its rivals. The third type is *strong evidence* which means that the datum makes only hypothesis *h* plausible and does not provide any support to its rivals.

The standard view of the philosophy of science treats evidence as a special subset of data that is assumed to be objective, is expected to justify hypotheses, is immediately given, and is held to be entirely reliable and primary to the theory. However, the concepts of weak, relative and strong evidence as we have just introduced them, differ from this view considerably. First, as opposed to the standard view of the philosophy of science, a datum is evidence not *per se*, but relative to a hypothesis. Accordingly, in the p-model evidence is p-context dependent and reliable only to a certain extent. Second, data which meet the criteria in the above three definitions, do not, as a rule, perfectly support or refute the given hypothesis. The connection between the datum and

12 [whom$_i$ thinks Lydia loves Jakob t$_i$], 'Who does Lydia think Jakob loves?'

13 This is the reason why in the first decade of our century a heated debate on the nature of linguistic data and evidence was initiated. See Kertész and Rákosi (2012) for the analysis of this discussion.

14 Accordingly, the p-model explicates the 'theory-ladenness' of data as 'p-context-dependence'.

the hypothesis is established by plausible inferences relying on plausible premises. Third, the function of evidence is not restricted by the p-model to the testing of hypotheses, but data and evidence play a role in every stage of the process of linguistic theorising.

This treatment of 'evidence' seems to reflect the way it has been used in generative linguistics. Although authors make ample use of the term 'evidence', virtually no hypothesis put forward within any version of generative linguistics could be supported with certainty by what a certain author called 'evidence'. Since virtually all hypotheses in the past frameworks had to be given up, what has been called 'evidence' is substantially uncertain as well.

To sum up what has just been said, the last claim that explicates the property (P1) mentioned in the Introduction says:

(T) e. In generative linguistics, evidence is uncertain and does not provide full support for the hypotheses of the theory.

2.2 On property (P2): The tolerance of inconsistency

From the late 1970s on, Chomsky regularly emphasised that his generative linguistics fits into what he calls the 'Galilean style of science'. One component of the Galilean style of science is the 'principle of epistemological tolerance' (Botha, 1983; Klausenburger, 1983; Riemer, 2009; Kertész, 2012; Kertész and Rákosi, 2013), which Chomsky characterizes as follows:

> [a]pparent counterexamples and unexplained phenomena should be carefully noted, but it is often rational to put them aside pending further study when principles of a certain degree of explanatory power are at stake. How to make such judgements is not at all obvious: there are no clear criteria for doing so. [...] But this contingency of rational inquiry should be no more disturbing in the study of language than it is in the natural sciences. (Chomsky, 1980, 2)

Applied to linguistic theories, this amounts to the claim that contradictions between the data and the hypotheses may be temporarily tolerated, in the hope that later, as inquiry progresses, more perfect versions of the theory will eliminate them. Thereby, the tolerance of the inconsistency serves the protection of the theory's explanatory principles.

Of course, inconsistency tolerance is highly problematic, because it is at variance both with the criteria of rationality advocated by the standard view of the philosophy of science and the principle of non-contradiction of classical two-valued logic:

> For it can easily be shown that if one were to accept contradictions, then one would have to give up any kind of scientific activity: it would mean *a complete breakdown of science*. This can be shown by proving that if two contradictory statements are admitted, any statement whatever must be admitted; for from a couple of contradictory statements any statement whatever can be validly inferred. (Popper, 1962, 313; emphasis added)

Nevertheless, new developments in the philosophy of science and logic shed new light on the role which contradictions play in scientific theories. First, it has been argued that most scientific theories are inconsistent and that thereby the crucial question is why they are workable in spite of their being inconsistent (see Meheus ed., 2002). Second, different kinds of paraconsistent logics have been developed which allow the simultaneous presence of contradicting statements but avoid logical chaos.

The logical consequence relation of classical two-valued logic is said to be *explosive*, which means that from a contradictory pair of premises any statement can be inferred. In contrast, a logical consequence relation is *paraconsistent* if and only if it is not explosive, and a logic is called paraconsistent if its logical consequence relation is not explosive. Accordingly, a paraconsistent logic allows for specific kinds of inconsistency without triggering logical chaos.[15]

Rescher and Brandom's 'logic of inconsistency' (Rescher and Brandom, 1980) is based on a Kripke-semantics, and they introduce, among other things, the operation of *superposition* on the set of possible worlds:

(3) The *superposition* of the possible worlds w_1 and w_2 is a possible world w in which a statement p is true if and only if it is true either in w_1 or in w_2.

Superposed worlds are *overdetermined* in the sense that in a superposed world w it can happen that both p and $\sim p$ are true (Rescher and Brandom, 1980, 10).

'$\uplus$' symbolises the operation of superposition. (3) says that p is true in $w_1 \uplus w_2$ if and only if it is true *at least* in one of the component worlds w_1 and w_2, respectively.

The semantic consequence relation in superposed worlds seems to differ radically from that of standard worlds because *a series of classical inference schemata are not valid*. Thus, principle (4) – which says that if the premises of a syntactically valid inference are true, then the conclusion must also be true – does not hold in superposed worlds:

(4) If
  a. $p_1, p_2, \ldots, p_n \vdash q$ is a valid inference principle of classical logic, and
  b. p_1 is true in w, p_2 is true in $w, \ldots, p_n$ is true in w,
  then
  c. q is true in w.

As (4b) indicates, the premises are true *distributively*. According to (3), however, it is possible that p_1, p_2 etc. are true in different component worlds, without being true in the superposed world w as well. Nonetheless, a second interpretation of the requirement that "the premises have to be true" is also possible: in this interpretation, the

15 However, in order to avoid misunderstandings, it is important to distinguish paraconsistency from dialetheism. See on this, e. g. Priest et al.'s (2016) and Kertész and Rákosi's (2013) concise overview.

premises must be true *collectively*, that is, within one component world. And this interpretation results in a principle that prevails in superposed worlds as well:

(5) If

a. $p_1, p_2, \ldots, p_n \vdash q$ is a valid inference principle of classical logic, and

b. $p_1 \,\&\, p_2 \,\&\, \ldots \,\&\, p_n$ is true in w,

then

c. q is true in w.

From these considerations it is clear that the classical and paraconsistent treatment of inconsistency are different. In classical logic, there is no difference between (4) and (5), and any arbitrary statement can be inferred from a contradiction. In contrast, in paraconsistent logic, that is, in superposed worlds, (4b) and (5b) do not coincide and only (5) holds. Thus, we can distinguish between two kinds of inconsistency.

We will speak of *weak* inconsistency if $w = w_1 \uplus w_2$, and p is true in w and $\sim p$ is true in w. What does this mean? According to (3), if p is true in w, then p has to be true in one of the two possible worlds. Let us suppose that p is true in w_1. Since in w_1 and w_2 the principles of standard logic prevail, $\sim p$ cannot be true in w_1, too. However, $\sim p$ is true in w as well. From this it follows that $\sim p$ can be true only in the possible world w_2, that is, the contradictory statements obtain in two distinct possible worlds. Therefore, logical chaos cannot emerge because the simultaneous use of p and $\sim p$ as the premises of inferences is, due to the failure of (4), forbidden; the two statements are *separated* into two different possible worlds. *Strong* inconsistency emerges if $p \,\&\, \sim p$ is true in w, and, as (5) shows, p and $\sim p$ emerge within the same component world. Weak inconsistency can be evaluated as harmless, because it does not lead to the collapse of the system, whereas strong inconsistency is harmful.[16]

To sum up, the p-model captures the inconsistent nature of generative grammars in that it fits into current views in the philosophy of science which attempt to account for inconsistencies in scientific inquiry and also, in that it integrates paraconsistent logic. The above considerations boil down to the following claim:

(T) f. Generative linguistic theorizing makes use of paraconsistent tools, tolerating weak inconsistency but forbidding strong inconsistency.

2.3 On property (P3): The cyclic and prismatic nature of generative linguistic theorizing

In order to solve a p-problem (cf. Section 2.1.3), we have to re-evaluate the p-context by revising its elements. A *solution of a problem* is achieved if a p-context has been arrived

16 See Kertész and Rákosi (2013) for examples in generative linguistics.

at in which either (a) the statement responsible for the incompleteness of the previous p-context-version is unanimously supported or opposed by the sources, that is, it becomes either plausible or implausible on the basis of all sources, or (b) a paraconsistent treatment can be elaborated, that is, the statements generating p-inconsistency can be separated systematically and in a well-motivated way.

It is possible, however, that *a p-problem has several solutions*. This necessitates the introduction of the notion of the *resolution of a p-problem*. We resolve a p-problem if we find a solution of the given p-problem which is, when compared with other solutions, the best according to a particular set of accepted criteria and according to the information available. It may be the case, however, that in an informational state one can only show that for the time being there is no resolution achievable.

To achieve the solutions or the resolution of a given p-problem, a heuristic process is needed that we will call *plausible argumentation*. In simple terms, plausible argumentation consists of chains of plausible inferences and amounts to the gradual transformation of a p-problematic p-context into one which is no longer (or at least, less) p-problematic. This involves the successive *re-evaluation* of a p-problematic p-context by the elaboration of possible solutions to the problems it has raised, as well as the comparison of the alternative solutions. Its aim is to detect all available solutions and to decide which of them is to be accepted as the resolution of the given p-problem.

Accordingly, since, as a rule, the re-evaluation of a p-problematic p-context does result directly in an unproblematic one, but may raise new problems, the argumentation process requires the revision of previous decisions, the assessment of other alternatives, etc. Therefore, throughout the argumentation process one returns to the problems at issue again and again, and retrospectively re-evaluates the earlier decisions about the acceptance or rejection of statements, the reliability of the sources, the plausibility values of the statements, the workability of methodological norms, the conclusions previously reached by inferences etc. (cf. also Rescher, 1976, 1987). This kind of retrospective re-evaluation is not linear, but *cyclic* and *prismatic*. The prismatic character means that the cycles continuously change the *perspective* from which the pieces of information constituting the p-context are evaluated (cf. Rescher, 1987).

An *argumentation cycle* is a phase of the plausible argumentation process which develops a new solution to the p-problems included in the starting p-context and examines whether with this solution their resolution has been achieved. It may have *sub-cycles* which revise certain decisions made within the given argumentation cycle, and continue the argumentation process by coming back to an earlier stage of the argumentation cycle.

In sum:

(T) g. Generative linguistic theorizing is of a cyclic and prismatic nature instead of being linear.

3 Case study

In order to explain the behaviour of modal verbs in Spanish and Italian, Zubizarreta (1982) raises the following hypothesis:

(6) Modal verbs behave in Spanish and Italian like main verbs (and unlike auxiliaries).

The starting p-context is p-incomplete because the plausibility/implausibility of (6) cannot be judged at the outset. She decides on the tenability of this hypothesis with the help of a 4-cycle process of plausible argumentation.

Argumentation cycle 1. First, she finds that (6) is supported by a series of syntactic tests:

(7) *Null-complement anaphora*: In Spanish, VP complements of main verbs can be dropped in such a way that the complement can be reconstructed from the previous clause or the context, while with auxiliaries, this is not the case.[17]

(8) *Placement of negation*: In Spanish, the negation is placed between main verbs and their verbal complements. Auxiliaries, however, are different: the negation cannot occur between the auxiliary and the verbal complement.

(9) *Cliticization*: In Spanish, main verbs – in contrast to auxiliaries – allow clitics to be attached to their verbal complements.

(7)–(9) work as parts of indirect sources (that is, chains of plausible inferences) with the help of which the initial plausibility value of (6) can be determined. The first plausible inference, based on (7), is (10):

(10) 0 < |If in Spanish, VP complements of modals can be dropped in such a way that the complement can be reconstructed from the previous clause or the context, then the sentences *Juan podría/debería visitar a María y Pedro también podría/debería* are grammatically correct.$|_Z < 1$[18]

0 < |The sentences *Juan podría/debería visitar a María y Pedro también podría/debería* are grammatically correct.$|_Z < 1$

---

0 < |In Spanish, VP complements of modals can be dropped in such a way that the complement can be reconstructed from the previous clause or the context.$|_{(10)} < 1$

17 For the concept of 'null-complement anaphor' see, for example, Huang (2000, 5).

18 [John could/should visit to Mary and Peter also could/should […]]
'John could/should visit Mary and Peter could/should also do so.'

Apparently, there is no logical consequence relation between the premises and the conclusion. This inference is enthymematic but capable of making its conclusion plausible because it belongs to the second group of plausible inferences we have introduced in Section 2.1. The plausibility value of the first premise of (10) originates from Zubizarreta's metalinguistic intuition as a direct source. It contains the result of linguistic analyses, because it presupposes that the two sentences mentioned in the consequent have the syntactic structure described in the antecedent. The second premise is a grammaticality judgement based on Zubizarreta's linguistic intuition as a native speaker. This means that both premises are plausible statements since their sources are not completely reliable.[19] Thus, they capture that (6) has been successfully tested on two pieces of linguistic data. Further, the premises can be completed by latent background assumptions such as the statement that all other sentences with a similar syntactic structure are also grammatical in Spanish etc. These background assumptions are plausible (as soon as they have been checked) or they are of neutral plausibility (if they have not been checked yet). Since all premises and latent background assumptions possess a plausibility value, the conclusion of (10) can be deemed plausible as well, on the basis of this inference as an indirect source.

The next member of this chain of inferences, (11), belongs to the first group of plausible inferences, because it is deductively valid and its premises are plausible statements:

(11) 0 < |If in Spanish, VP complements of auxiliaries can be dropped in such a way that the complement can be reconstructed from the previous clause or the context, then the sentence *Juan ha visitado a María y Pedro también ha* is grammatically correct.$|_Z$ < 1[20]

0 < |The sentence *Juan ha visitado a María y Pedro también ha* is grammatically not correct.$|_Z$ < 1

---

0 < |In Spanish, VP complements of auxiliaries cannot be dropped in such a way that the complement can be reconstructed from the previous clause or the context.$|_{(11)}$ < 1

Making use of the conclusions of (10) and (11), Zubizarreta draws the following plausible inference:

19 Cf., for example, Schütze (1996) on the unreliability of grammaticality judgements and our earlier remarks on the current discussion of linguistic data and evidence.

20 [John has visited to Mary and Peter also has […]]
'John has visited Mary and Peter has also done so.'

(12) 0 < |In Spanish, VP complements of modals can be dropped in such a way that the complement can be reconstructed from the previous clause or the context.$|_{(10)}$ < 1
0 < |In Spanish, VP complements of auxiliaries cannot be dropped in such a way that the complement can be reconstructed from the previous clause or the context.$|_{(11)}$ < 1
[0 < |In Spanish, VP complements of main verbs can be dropped in such a way that the complement can be reconstructed from the previous clause or the context.$|_{Z}$ < 1] (= (7))

---

0 < |Modals behave in Spanish like main verbs (and not as auxiliaries).$|_{(12)}$ < 1 (= (6))

The conclusion of (12) contains a generalisation because it declares a high degree of analogy based on one common feature. Accordingly, (12) is an indirect source making (6) plausible, or, to formulate the result of our reconstruction differently, the datum 'The sentences *Juan podría/debería visitar a María y Pedro también podría/debería* are grammatically correct' provides strong evidence for (6).

The second test made use of by Zubizarreta pertains to the place of negation, see (8). The related chain of inferences can be reconstructed as follows:

(13) 0 < |If in Spanish, the negation is placed between modals and their verbal complements, then the sentences *Pedro podría/debería* ***no*** *contestar la carta* are grammatically correct.$|_{Z}$ < 1[21]
0 < |The sentences *Pedro podría/debería* ***no*** *contestar la carta* are grammatically correct.$|_{Z}$ < 1

---

0 < |In Spanish, the negation is placed between modals and their verbal complements.$|_{(13)}$ < 1

(14) 0 < |If in Spanish, the negation is placed between auxiliaries and their verbal complements, then the sentence *Pedro ha* ***no*** *contestado la carta* is grammatically correct.$|_{Z}$ < 1[22]
0 < |The sentence *Pedro ha* ***no*** *contestado la carta* is grammatically not correct.$|_{Z}$ < 1

---

0 < |In Spanish, the negation cannot be placed between auxiliaries and their verbal complements.$|_{(14)}$ < 1

It is easy to see that (13) and (14) have the same structure as (10) and (11), respectively. From the conclusions of (13) and (14), Zubizarreta obtains (15), which is a counterpart of (12):

21 [Peter could/should not answer the letter]
'Peter could/should not answer the letter.'

22 [Peter has not answered the letter]
'Peter has not answered the letter.'

(15) [0 < |In Spanish, the negation is placed between main verbs and their verbal complements.$|_Z$ < 1]
0 < |In Spanish, the negation is placed between modals and their verbal complements.$|_{(13)}$ < 1
0 < |In Spanish, the negation cannot be placed between auxiliaries and their verbal complements.$|_{(14)}$ < 1

---

0 < |Modals behave in Spanish like main verbs (and not as auxiliaries).$|_{(15)}$ < 1 (= (6))

Since the inference (15) makes (6) plausible as an indirect source, the plausibility of this hypothesis increases.

The third test, appertaining to the cliticization in Spanish, was based on (9). The chain of inferences related to (9) is similar to (10)–(12) and (13)–(15), and makes use of the following data:

(16) 0 < |The sentences *Pedro puede/debe contestar**la*** are grammatically correct.$|_Z$ < 1[23]

(17) 0 < |The sentence *Pedro ha contestado**la*** is grammatically not correct.$|_Z$ < 1[24]

When we add these data and indirect sources (plausible inferences) to the p-context, the plausibility of (6) further increases. This value becomes even higher after extending the p-context with (18), because there is a strong analogy between the two hypotheses:

(18) 0 < |From several points of view, modals behave in Italian like main verbs.$|_{\mathit{Rizzi}\ (1978)}$ < 1

If one made a decision on the basis of this state of the p-context, one should evidently give up ~(6) and keep (6) as the solution of the starting p-problem. There are, however, further data that are relevant to this problem. Therefore, no decision can be made at this point, but a new argumentation cycle has to be started which will yield a new solution to the starting p-problem.

Argumentation cycle 2. There are also tests that lead to a result inconsistent with (6). The first of these is (19):

(19) *Cliticization*: In Spanish and Italian, auxiliaries allow clitics to be attached to them. Main verbs, in contrast, do not allow it.

In relation to (19), we obtain the following series of plausible inferences:

---

23 [Peter can/must answer-acc. cl.]
'Peter can/must answer it.'

24 [Peter has answered-acc. cl.]
'Peter has answered it.'

(20) 0 < |If in Spanish and Italian, modals allow clitics to be attached to them, then the sentences *Pedro* ***le*** *pudo/debió hablar personalmente* and *Gianni* ***gli*** *ha dovuto/potuto parlare personalmente* are grammatically correct.$|_Z$ < 1[25]

0 < |The sentences *Pedro* ***le*** *pudo/debió hablar personalmente* and *Gianni* ***gli*** *ha dovuto/potuto parlare personalmente* are grammatically correct.$|_Z$ < 1

---

0 < |In Spanish and Italian, modals allow clitics to be attached to them.$|_{(20)}$ < 1

(21) 0 < |If in Spanish and Italian, main verbs allow clitics to be attached to them, then the sentences *Pedro* ***le*** *prometió hablar personalmente* and *Gianni* ***gli*** *ha promesso di parlare personalmente* are grammatically correct.$|_Z$ < 1[26]

0 < |The sentences *Pedro* ***le*** *prometió hablar personalmente* and *Gianni* ***gli*** *ha promesso di parlare personalmente* are grammatically not correct.$|_Z$ < 1

---

0 < |In Spanish and Italian, main verbs do not allow clitics to be attached to them.$|_{(21)}$ < 1

(22) makes use of the comparison and the contrast between the conclusions of the previous two inferences:

(22) 0 < |In Spanish and Italian, modals allow clitics to be attached to them.$|_{(20)}$ < 1

0 < |In Spanish and Italian, main verbs do not allow clitics to be attached to them.$|_{(21)}$ < 1

0 < |If in Spanish and Italian, modals allow clitics to be attached to them, while main verbs do not, then modals behave in Spanish like non-main verbs.$|_Z$ < 1

---

0 < |Modals behave in Spanish like non-main verbs.$|_{(22)}$ < 1 (= (~(6))

The last test Zubizarreta refers to, can be formulated as follows:

(23) *Impersonal passive*: Auxiliaries, in contrast to main verbs, allow impersonal *se*-passive in Spanish and impersonal *si*-passive in Italian.

Plausible inferences related to (23) increase the plausibility value of ~(6), since they result in the same final conclusion. They have the same structure as (20)–(22), and rely on the following data:

25 [Peter dat. cl. could/had to speak personally]
'Peter could/had to speak to him/her personally.'

26 [Peter dat. cl. promised to speak personally]
'Peter promised to speak to him/her personally.'

(24) 0 < |The sentences *Estos libros **se** podrían/deberían comprar ya* and *Questi libri **si** potrebbero/dovrebbero comprare gia* are grammatically correct.$|_Z$ < 1[27]

(25) 0 < |The sentences *Estos libros **se** prometieron comprar* and *Questi libri **si** promissero di comprare* are grammatically not correct.$|_Z$ < 1[28]

The upshot of this argumentation cycle is that these two chains of inferences are indirect sources that make the hypothesis (6) implausible.

Obviously, the simultaneous presence of (6) and its negation make the p-context *p-inconsistent*. This p-inconsistency relies on the background assumption that modals do not constitute a third, autonomous category of verbs but behave either like main verbs or as auxiliaries, and on two analogies: Spanish and Italian modals behave like main verbs from certain points of view, and, at the same time, they are similar to auxiliaries in some other respects.

Argumentation cycle 3. As we have seen, both members of this p-inconsistency are supported by pieces of evidence with a similar origin and structure. Therefore, the strategy of treating (6) and its negation as rival alternatives and trying to make a decision between them does not seem to be viable. Thus, while the outcome of the tests points toward the stance that neither member of this inconsistency can be given up, it is also clear that the unrestricted simultaneous maintenance of (6) and its negation would lead to logical chaos. Zubizarreta (1982, 138), however, realises that (6) and its negation can be maintained simultaneously in such a way that they become separated from each other. The first step in this direction is a thorough description of the circumstances in which the non-main verb behaviour of modals appears:

(26) When modals behave in Spanish like non-main verbs in connection with cliticization and impersonal passive, or in connection with Auxiliary Change in Italian, then they also behave as non-main verbs with respect to the following phenomena: Cleft-formation, Right-node raising, Heavy-NP shift, *Wh*-movement, null-complement anaphora and placement of negation. Otherwise, they behave as main verbs.

(26) is made plausible by two plausible inferences as indirect sources:

27 [These books refl. cl. could/should buy by-now]
'These books could/should have been bought by now.'

28 [These books refl. cl. promised buy]
'These books were promised to be bought.'

(27) 0 < |If (26), then a modal behaves in respect to null-complement anaphors as a non-main verb when a clitic is attached to it.$|_Z = 1$
0 < |If a modal behaves in respect to null-complement anaphora as a non-main verb when a clitic is attached to it, then the sentences *Juan podría/debería visitar a Maria y Pedro también* ***la*** *podría/debería* are grammatically not correct.$|_Z < 1$[29]
0 < |The sentences *Juan podría/debería visitar a Maria y Pedro también* ***la*** *podría/debería* are grammatically not correct.$|_Z < 1$

---

0 < $|(26)|_{(27)} < 1$

(28) 0 < |If (26), then negation cannot be placed between the modal and the verb when a clitic is attached to the former.$|_Z = 1$
0 < |If negation cannot be placed between the modal and the verb when a clitic is attached to the former, then the sentences *Juan* ***la*** *podría/debería* ***no*** *contestar* are grammatically not correct.$|_Z < 1$[30]
0 < |The sentences *Juan* ***la*** *podría/debería* ***no*** *contestar* are grammatically not correct.$|_Z < 1$

---

0 < $|(26)|_{(28)} < 1$

(27) and (28) belong to the first group of plausible inferences, that is, they are enthymematic. Therefore, they have to be completed by latent background assumptions, whose elaboration, however, we omit for the sake of transparency of the structure of these inferences. These background assumptions refer to further successful tests of the predictions obtained from (26).

As a second step, Zubizarreta (1982, 139) investigates Strozer's (1976) solution of the p-inconsistency between (6) and its negation:

(29) Modals in Spanish and Italian are main verbs and also auxiliaries under the circumstances described in (26).

Hypothesis (29) is a paraconsistent solution. It keeps both members of the conflict, but it does not lead to logical chaos, because with the help of (26), it clearly separates the fields of application of the two conflicting statements.

Nonetheless, Zubizarreta (1982, 139–140) comes to the conclusion that this hypothesis is not acceptable, either, because it generates further p-inconsistencies. Namely, it is in conflict with several hypotheses of generative grammar with a high plausibility value. For example:

29 [John could/should visit to Mary and Peter also acc. cl. could/should]
'John could/should visit Mary and Peter could/should also do so.'

30 [John acc. cl. could/should not answer]
'John could/should not answer it.'

(30) 0 < |If the sentences *Pedro **la** debería poder visitar* and *Pedro **la** querría poder comenzar a escribir* are grammatically correct, and modals in Spanish are main verbs and also auxiliaries under the circumstances described in (26) (= (29)), then modals in Spanish can co-occur if they function as auxiliaries.$|_Z < 1$[31]

0 < |If modals can co-occur if they function as auxiliaries, then Aux is a recursive node.$|_Z < 1$

0 < |The sentences *Pedro **la** debería poder visitar* and *Pedro **la** querría poder comenzar a escribir* are grammatically correct.$|_Z < 1$

0 < |Aux is not a recursive node.$|_Z < 1$

---

0 < |It is not the case that modals in Spanish are main verbs and auxiliaries as well under the circumstances described in (26) (= ~(29)).$|_{(30)} < 1$

It is easy to identify the problematic point: while (29) seems to be on the right track in relation to the double-facedness of Spanish and Italian modals and stipulating the circumstances when they behave as main verbs and non-main verbs, it categorizes the non-main verb behaviour incorrectly.

Argumentation cycle 4. Zubizarreta re-evaluates this part of (29), and raises a modified version of it:

(31) Modals in Spanish and Italian have two parallel syntactic structures. They are main verbs (argument-taking predicates) as well as non-main verbs (that is, verbal affixes), but under different conditions.

That is, she represents Spanish and Italian sentences containing a modal verb with the help of the following double tree structure (cf. Zubizarreta, 1982, 161):

$[_{S1}\ NP_1\ [_{VP}\ \ V_1\ [_{S2}\ NP_2\ [_{VP}\ V_2\ NP_3]]]]$

Juan$_i$ puede e$_i$ visitar a María

$[_S\ NP_1\ [_{VP}\ [_V\ V_{Affix}\ +\ V]\ \ NP_3]]$

Argumentation cycle 4 results in a *modified* p-context version. This p-context version contains both members of the p-inconsistency but in such a way that with the help of (26), they are *systematically separated* from each other into two p-context versions. That is, if a clitic is attached to the modal, or if the modal is in *se*-passive, then the modal does not behave like a main verb but as a verbal affix; in contrast, in other cases it behaves like a main verb.

31 [Peter acc. cl. should can visit]; [Peter acc. cl. would-like-to can start to write]
'Peter should be able to visit her'; 'Peter would like to be able to start writing it'.

In this way, Zubizarreta considers (31) the best available solution, because on the basis of the pieces of information at her disposal, there is no counter-evidence against it. Accordingly, (31) is the *resolution* of the p-problem pertaining to (6) in the sense of the p-model (see Section 2.3).

That (31) does not result in logical chaos boils down to the claim that Zubizarreta's theory is *paraconsistent*:

(32)
a. Both (6) and its negation can be assigned a similarly high plausibility value on the basis of a series of indirect sources.
b. Therefore, the p-context containing (6) and its negation are p-inconsistent.
c. Let w_1 be the set of sentences in which cliticization, impersonal passive, or Auxiliary Change occur in connection with modals and w_2 sentences in which these structures do not occur in connection with modals.
d. Let $w = w_1 \cup w_2$. Since w_1 and w_2 are complementary, w is the whole of Zubizarreta's theory.
e. (6) will be true in w_1 and false in w_2, while its negation will be true in w_2 and false in w_1.
f. Both (6) and ~(6) are true in w, because each of them is true in one of the component worlds of w. However, the statement (6) & ~(6) is false in w_1, in w_2, and, therefore, in w too, because neither w_1 nor w_2 includes both of them.

4 Conclusions

The p-model has yielded the hypotheses (Ta)–(Tg) on the nature of generative linguistic theorizing (see also Kertész, 2017). By arguing for these tenets, we have shown, first, that the way in which generative linguistics proceeded in the past decades and is functioning today differs significantly from what either of the extreme positions mentioned in (i) and (ii) in Section 1 assumes. Proposals which in the generative linguistics literature have been put forward to fulfil the requirement of turning linguistics into a mature empirical theory are not based on workable and generally applied norms of natural sciences but rather, on outmoded and untenable tenets of the standard view of the analytical philosophy of science.

The p-model of plausible argumentation rejects – in harmony with current trends in the philosophy of science – a series of methodological prejudices which mainstream linguistics takes implicitly or explicitly for granted, but which are at variance with the research practice. Therefore, the p-model's foci are up-to-date from a more general point of view, too. The p-model's originality is due to the way it tries to integrate these themes to a comprehensive and full-fledged model of linguistic theorizing.

Finally, in the course of linguistic inquiry three closely interacting activities are needed: of course, *object-scientific* research in the sense that the linguist applies the framework chosen to the selected data in order to solve the problems raised; careful metatheoretical *self-reflection* on her own activity that may also include considering foundational questions of the field only indirectly related to her everyday problem solving activity; and *meta*theoretical insights gained by professional philosophers of science which may be used to furthering the *object*-scientific research methods of linguistics. Only if all three processes are simultaneously present, carefully comprehended and adjusted can the linguist hope to contribute to the effectiveness, fruitfulness and reliability of linguistic theorizing.

References

Andor, József. 2004. The master and his performance: An interview with Noam Chomsky. *Intercultural Pragmatics* 1. 93–111.

Behme, Christina. 2014. A 'Galilean' science of language. *Journal of Linguistics* 50. 671–704.

Botha, Rudolf P. 1983. On the 'Galilean style' of linguistic inquiry. *Lingua* 58. 1–50.

Chomsky, Noam. 1957. *Syntactic structures*. The Hague & Paris: Mouton.

Chomsky, Noam. 1980. On binding. *Linguistic Inquiry* 11. 1–46.

Featherston, Sam. 2007. Data in generative grammar: The stick and the carrot. *Theoretical Linguistics* 33(3). 269–318.

Huang, Yan. 2000. *Anaphora. A cross-linguistic study*. Oxford: Oxford University Press.

Hung, Edwin. 2014. *Philosophy of science complete. A text on traditional problems and schools of thought*. Wadsworth: Cengage Learning.

Jackendoff, Ray. 1994. *Patterns in the mind: Language and human nature*. New York, NY: BasicBooks.

Kertész, András. 2012. The 'Galilean style in science' and the inconsistency of linguistic theorising. *Foundations of Science* 17. 91–108.

Kertész, András. 2017. *The historiography of generative linguistics*. Tübingen: Narr Francke Attempto.

Kertész, András & Csilla Rákosi. 2012. *Data and evidence in linguistics: A plausible argumentation model*. Cambridge: Cambridge University Press.

Kertész, András & Csilla Rákosi. 2013. Paraconsistency and plausible argumentation in Generative Grammar: A case study. *Journal of Language, Logic and Information* 22. 195–230.

Kertész, András & Csilla Rákosi. 2014. The p-model of data and evidence in linguistics. In András Kertész & Csilla Rákosi (eds.), *The evidential basis of linguistic argumentation*, 15–48. Amsterdam & Philadelphia: Benjamins.

Klausenburger, Jürgen. 1983. Review of Botha, R., 'On the 'Galilean Style' of Linguistic Inquiry'. *Language* 59. 434.

Meheus, Joke (ed.). 2002. *Inconsistency in science*. Dordrecht: Kluwer.

Piattelli-Palmarini, Massimo. 1998. Foreword. In Juan Uriagereka (ed.), *Rhyme and reason: An introduction to minimalist syntax*, xxi–xxxvi. Cambridge, MA: The MIT Press.

Polya, George. 1948. *How to solve it*. Princeton: Princeton UP.

Popper, Karl R. 1962. *Conjectures and refutations*. London.

Priest, Graham, Koji Tanaka & Zach Weber. 2016. Paraconsistent Logic. In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/archives/win2016/entries/logic-paraconsistent/.

Pullum, Geoffrey K. 2017. Theory, data, and the epistemology of syntax. In Marek Konopka & Angelika Wöllstein (eds.), *Grammatische Variation. Empirische Zugänge und theoretische Modellierung*, 283–298. Berlin: De Gruyter.

Rescher, Nicholas. 1976. *Plausible reasoning*. Assen & Amsterdam: Van Gorcum.

Rescher, Nicholas. 1979. *Cognitive systematisation*. Oxford: Blackwell.

Rescher, Nicholas. 1987. How serious a fallacy is inconsistency? *Argumentation* 1. 303–316.

Rescher, Nicholas & Robert Brandom. 1980. *The logic of inconsistency*. Oxford: Blackwell.

Riemer, Nick. 2009. Grammaticality as evidence and as prediction in a Galilean linguistics. *Language Sciences* 31. 612–633.

Rizzi, Luigi. 1978. A restructuring rule in Italian syntax. In Samuel Jay Keyser (ed.), *Recent transformational studies in European languages* (Linguistics Inquiry Monograph Three), 113–158. Cambridge: MIT Press.

Sampson, Geoffrey. 1975. *The form of language*. London: Weidenfeld & Nicholson.

Sampson, Geoffrey. 2007. Reply. *Corpus Linguistics and Linguistic Theory* 3. 111–129.

Sampson, Geoffrey & Anna Babarczy. 2014. *Grammar without grammaticality. Growth and Limits of Grammatical Precision*. Berlin: Mouton de Gruyter.

Schütze, Carson T. 1996. *The empirical base of linguistics. Grammaticality judgments and linguistic methodology*. Chicago & London: The University of Chicago Press.

Strozer, Judith Reina. 1976. Clitics in Spanish. Doctoral Dissertation, UCLA.

Zubizarreta, Maria Luisa. 1982. On the Relationship of the Lexicon to Syntax. Ph.D. Diss., Cambridge, Mass: MIT.

Stephan Kornmesser

18 The multiparadigmatic structure of science and generative grammar

Abstract: In this article, I will analyze the linguistic paradigm of generative grammar from the point of view of Thomas Kuhn's philosophy of science. I will show that generative grammar satisfies all of Kuhn's paradigm criteria. However, according to Kuhn, science is always based on one single paradigm. In contrast to that monoparadigmatic view, I will argue that it is more adequate to assume a multiparadigmatic structure of science. To this end, I will work out the components of the paradigm of generative grammar with a special focus on its empirically immunized, non-falsifiable elements. Further, I will show that linguistics has a multiparadigmatic structure by contrasting generative grammar with the paradigm of construction grammar and describing both approaches as co-existing and competing paradigms. Finally, I will analyze the incommensurability of both paradigms.

1 Introduction

In this article, I will discuss the development of and the relation between linguistic paradigms from the point of view of Thomas Kuhn's philosophy of science. In the second section, I will introduce Kuhn's model of scientific development. It is a consequence of Kuhn's model that at a certain level of development a scientific discipline is always based on *one* single paradigm. I will challenge this model arguing that a *multiparadigmatic* structure of science is a more adequate description of the structure of a comprehensive class of scientific disciplines.

In section three, I will provide a refined characterization of a paradigm and apply this notion to the linguistic approach of generative grammar (hereafter: *GG*). The main aim of the third section is to show that GG satisfies all conditions for being a paradigm and hence constitutes a scientific paradigm in the sense of Kuhn.

According to Kuhn, each paradigm contains a set of theoretical assumptions that are completely immunized against contradicting evidence and are hence not falsifiable in the sense of Popper (Kuhn [1962] 1996, 77). In the fourth section, I will discuss two case studies of non-falsifiable statements of GG. I will show that even if the empirical predictions based on these assumptions fail, the assumptions are not given up within the paradigm.

However, if anomalies increase within a paradigm, new approaches coping better with the anomalies might appear. In Kuhn's model, there will be one new approach replacing the old paradigm and thereby constituting the new paradigm of the discipline. The new and the old paradigm are incommensurable to each other due to their

https://doi.org/10.1515/9783110540253-018

deeply grounded conceptual differences. In contrast to Kuhn's model, I will show in the fifth section that, first, the old paradigm does not need to be replaced, and second, that there can be several new approaches constituting a multiparadigmatic structure of science. To this end, I will introduce the paradigm of construction grammar (hereafter: *CG*), the elements of which aim to accomplish the same goals as GG by means of incommensurable foundations. In section six, I will summarize the main results.

2 Challenging Kuhn's model of scientific development

In Section 2.1, I will introduce Thomas Kuhn's model of scientific development. In Section 2.2, I will present a refined version of Kuhn's model that is on the one hand based on Kuhn's descriptive and historical analyses of scientific development, but on the other hand provides a more adequate view on the structure of science.

2.1 The monoparadigmatic structure of science

According to Kuhn's ([1962] 1996) historical analyses of the development of the natural sciences in *The Structure of Scientific Revolutions* (hereafter *SSR*), a scientific discipline develops in an alternating process of *normal science* and *revolutionary science*. Normal science is a phase of cumulative progress, in which the discipline is based on a certain *paradigm*. Such a paradigm consists of theoretical assumptions, methodological requirements, and exemplary problem solutions shared by a certain scientific community (for a more detailed analysis of the concept *paradigm*, see Section 3.1). In normal science, scientific research is led by a paradigm – it determines what counts as a scientific problem and how to treat it. Kuhn ([1962] 1996, 187) describes normal science as a process of *puzzle-solving*: the paradigm provides techniques and standard examples for solving scientific puzzles and suggests new puzzles leaving enough space for scientific exploration and a cumulative growth of knowledge.

In a discipline, in which a paradigm has not yet been established, there are several schools or approaches competing with each other. According to Kuhn ([1962] 1996) only a paradigm-based discipline is a mature science, otherwise it is non-mature and should rather be characterized as pre-scientific. It is a normative consequence of Kuhn's model that each discipline must seek to overcome the phase of multiple approaches in order to become a paradigm-based science. Kuhn, a physicist, states that only the natural sciences are paradigm-based disciplines, concluding that only the natural sciences are mature sciences.

If in normal science anomalies occur – for example conflicts between theoretical predictions and empirical evidence – the paradigm will not simply be given up by the scientific community. Instead, if possible, the paradigm will be adjusted in order to be able to cope with the conflicting evidence. There are several ways the scientific community of a paradigm may react to an anomaly, for example by ignoring or challenging the conflicting evidence, by introducing ad hoc hypotheses or by revising certain laws of the paradigm. That is, a paradigm is immune to a direct empirical falsification in the sense of Popper.

However, it might be the case that anomalies arise that, first, cannot be solved within the existing paradigm, and, second, are widely accepted by the scientific community as an important problem that actually should be solved within normal science. Now, the paradigm is in crisis and normal science is interrupted by the phase of *extraordinary science*, in which new approaches are developed in order to solve the anomalies and to replace the existing paradigm. If one of the alternative approaches becomes a promising candidate, it can be accepted as new paradigm in a phase of *revolutionary science*, replacing the old paradigm in a *scientific revolution*.

According to Kuhn ([1962] 1996), a scientific revolution is a *non*-cumulative process because the old paradigm and the new one are *incommensurable*. A revolutionary paradigm shift is incommensurable due to changes from the conceptual structure of the former to the conceptual structure of the latter paradigm (Kuhn, 2000). That is, scientific revolutions cause concept shifts between the languages of both paradigms, complicating their comparison or, in the strong interpretation of incommensurability, even making comparison impossible.

After the phase of revolutionary science, the new paradigm has been established and constitutes a new phase of normal science generating cumulative scientific progress. Hence, it follows from Kuhn's cyclic model of scientific development introduced above that there always is only one paradigm at a time within a scientific discipline. I call this assumption the *monoparadigmaticity of science* (*MS*). It is a consequence of MS that incommensurability can only occur between two paradigms P^1 and P^2 with P^2 having replaced P^1, i. e. between two paradigms that do not exist parallel to each other.

There are two ways to understand MS. First, the statement that there always is only one paradigm at a time in a mature science can be understood to be *analytic*. In this interpretation MS is a conceptual truth because being the only scientific approach in a certain state of development in a scientific discipline is a necessary condition for this approach to be a paradigm. In other words: the concept *paradigm* is inter alia defined by the property of monoparadigmaticity. In this interpretation, a multiparadigmatic structure of science is impossible on conceptual grounds since it would be a conceptual contradiction. However, although Kuhn states MS and associates normal science with monoparadigmaticity, the concept *paradigm* is not defined in this way. Instead it is more adequate to assume that it is defined by the *components* of a paradigm (that I will explicate in Section 3.1) as well as by the *acceptance* of a *scientific community*. However, the acceptance of a scientific community does not entail that there is only one single scientific community in a certain discipline accepting only one single paradigm at a time. Therefore it is more adequate to consider MS to be an *empirical* statement. Hence, MS is an inductive generalization based on Kuhn's diachronic investigation of scientific development – it is stated to be an empirical fact that there always is only one paradigm at a time because, up to now, there has always been only one particular paradigm in a mature science.

To sum up, Kuhn's descriptive and diachronic analyses of the history of science result in the following three main theses (Kornmesser, 2014a):

- First, none except the natural sciences are mature sciences, i. e. paradigm-based sciences.
- Second, in a scientific discipline, there is always only one single paradigm leading scientific exploration in a particular phase of development (MS).
- Third, incommensurability only occurs between an old paradigm and a new paradigm replacing the old one in a scientific revolution.

2.2 The multiparadigmatic structure of science

In contrast to Kuhn's model of the structure of science, there is a considerable amount of research showing that MS is wrong (Schurz and Weingartner, 1998; Kornmesser, 2014a, 2014c; Kornmesser and Schurz, 2014). It has been shown for a wide range of different scientific disciplines that they constitute a multiparadigmatic structure with competing paradigms co-existing for long periods of time and setting up different parallel normal sciences. The main line of argument for a multiparadigmatic structure of science runs as follows: A scientific approach is a paradigm if it consists of the components of a paradigm as introduced by Kuhn ([1962] 1996, postscript) and if it is accepted by a scientific community. Hence, scientific approaches satisfying these criteria are paradigms in the sense of Kuhn. However, it might be the case that there are several paradigms coexisting.[1] The multiplicity of approaches within a discipline x does not entail anymore that x is a non-mature discipline – it is not the number of approaches that is relevant for being a mature science, but the status of being paradigms in the sense of Kuhn. Since MS is not an analytic but an empirical statement, the coexistence of paradigms is not a conceptual contradiction but empirical evidence challenging MS. The coexistence of paradigms can be found in natural sciences but is primarily a phenomenon occurring in non-natural sciences. From this it follows that Kuhn's three main theses have to be revised and replaced by the following:

- First, not only natural sciences are paradigm-based sciences. Other disciplines also have phases of paradigm-based normal science. Hence these disciplines are mature sciences.

1 Of course, in this case the scientific community accepting a paradigm is a narrower circle than all researchers of a given research field. Otherwise there could not be two scientific communities accepting different paradigms at one and the same time.

- Second, there can be several coexisting paradigms within a scientific discipline, constituting a multiparadigmatic structure.
- Third, incommensurability can occur between coexisting paradigms.

The new theses are grounded on a reanalysis of the role of paradigms in scientific research about half a century after SSR using Kuhn's conceptual tools *paradigm*, *normal science*, and *incommensurability*, but refuting his normative claim that only disciplines that satisfy MS are mature sciences. Rather, the kind of empirical phenomena particular disciplines are concerned with (for instance, language, cognition, society, etc.) allow for the constitution of different and possibly incommensurable paradigms explaining these phenomena.

3 The structure of a paradigm

In this section, I will explicate the structure of a paradigm. In 3.1, I will introduce the components a paradigm consists of, and in 3.2, I will apply the notion of a paradigm to the theory of GG.

3.1 The components of a paradigm

Kuhn introduced the concept *paradigm* in SSR. However, there has been strong criticism of his vague use of this concept (Masterman 1970). Kuhn used the notion of a paradigm to refer to constellations of methods, laws, values, and exemplary problem solutions widely accepted by a scientific community. However, he also referred to influential scientific books like Newton's *Philosophiae Naturalis Principia Mathematica* as a paradigm and presented several metaphorical explications of a paradigm. In the second edition of SSR and in Kuhn (1977), he introduced the new notion of a *disciplinary matrix*, explicating his previously diffuse use of the term *paradigm*. According to Kuhn ([1962] 1996), a disciplinary matrix consists of the four components *symbolic generalizations*, *models*, *values*, and *exemplars*. The meanings of these components will be introduced below.

However, even if Kuhn introduced the new term *disciplinary matrix* in order to avoid ambiguities, I will still use the term *paradigm* in this article because of its prominence in the philosophy of science and beyond. Based on Kuhn's concept of a disciplinary matrix, Kornmesser and Schurz (2014), and Kornmesser (2014a, 2014c) defined a scientific approach to be a paradigm if and only if it consists of three components (a theoretical, an empirical, and a methodological component) and is accepted by a scientific community using the paradigm to generate a cumulative growth of knowledge in a phase of normal science. In the following, I will introduce the three components of a paradigm and relate them to Kuhn's components of a disciplinary matrix.

3.1.1 The theoretical component of a paradigm

The theoretical component of a paradigm consists of a *theory core* and a *theory periphery*. The theory core contains the *theoretical laws* and the *model ideas* of the paradigm. Theoretical laws are generalized statements that contain *theoretical terms*, i. e. scientific terms that refer to non-observable entities. Theoretical laws are used to predict and to explain empirical phenomena. They correspond to Kuhn's symbolic generalizations of a disciplinary matrix (Kuhn [1962] 1996, postscript). The model ideas contain ontological assumptions presupposed within the paradigm and analogies and metaphors for heuristic purposes. The model ideas correspond to Kuhn's models of a disciplinary matrix, which he sometimes describes to be the metaphysical part of paradigm (Kuhn [1962] 1996, postscript). The theory periphery consists of special laws that in conjunction with the theory core allow for the derivation of empirical statements. In case of a conflict with experience, it is the special laws of the theory periphery that are modified or even given up, while the theory core is preserved.[2]

3.1.2 The empirical component of a paradigm

The empirical component contains the *empirical applications* of the paradigm. It consists of three parts – a domain of *paradigmatic applications*, a domain of *successful applications*, and a domain of *programmatic applications*.

The paradigmatic applications correspond to Kuhn's exemplars of a disciplinary matrix and can be considered the central and most important part of Kuhn's conception of a paradigm for two reasons (Kuhn [1962] 1996, postscript): First, paradigmatic applications are solutions to scientific problems that motivated the development of the respective paradigm. That is, paradigmatic applications are standardized problem solutions that, in normal science, are transferred to new problems in order to solve them in the same way. Second, students and scientific offspring acquire a paradigm by means of paradigmatic applications: According to Kuhn, a paradigm is not acquired by learning its theoretical component in lectures but by learning to see the similarities of new problems and the paradigmatic applications one is familiar with.

The successful applications of a paradigm are the paradigm-based solutions to scientific problems at a certain point of time. The successful applications are a superset of the set of paradigmatic applications.

The programmatic applications contain further problems that are not yet solved but expected to be solved by the paradigm in the future. The intended and program-

2 The theory core of the theoretical component of a paradigm resembles the hard core of a research program according to Lakatos (1970), containing the central theoretical assumptions of a research program. The distinction between a theory core and a theory periphery takes into account Stegmüller's (1976) reconstruction of Kuhn's notion of a paradigm.

matic applications can be exemplified by Kuhn's metaphor of a paradigm being a map of a landscape, some parts of which are explored and already included in the map (successful applications) and other parts that have not yet been explored, leaving white areas on the map that have to be filled out by future efforts of normal science (programmatic applications).

3.1.3 The methodological component

The methodological component approximately corresponds to Kuhn's values of a disciplinary matrix. It contains methodical, epistemological, and normative elements shared by the scientific community, restricting the interplay between the theoretical and the empirical component in normal science. Preferring quantitative to qualitative research, integrating the paradigm into an epistemological tradition, or defending the point of view that science should have socially beneficial effects are examples of elements of the methodological component that could be accepted in a paradigm.

3.2 The paradigm of generative grammar

In this section, I will argue that Noam Chomsky's generative grammar (GG) is a prototypical paradigm. To this end, I will show that GG meets all of Kuhn's requirements to a paradigm as explicated by means of the paradigm components introduced in 3.1. That is, I will show that GG has all the components of a paradigm and is accepted by a wide spread scientific community constituting cumulative knowledge in a phase of normal science. For this purpose, I will focus on the normal scientific development of GG. I consider the normal scientific development of GG as a sequence of developmental steps each of which entails changes but has an outlasting common ground in the theoretical, methodological, or empirical components of the paradigm. I link the developmental steps to Chomsky's *Syntactic Structures* ([1957] 1971), *Aspects of the Theory of Syntax* (1965), *Government and Binding* (1981), *Knowledge of Language* (1986) and *The Minimalist Program* (1995), all of which are usually considered to be fundamental extensions or modifications of GG (for a brief overview, see Lasnik and Lohndal 2013).

3.2.1 Syntactic structures

Syntactic Structures can be regarded as the foundation of GG. Chomsky ([1957] 1971, ch. 2) introduces the requirement of the *independence of grammar* as a basic methodical requirement according to which the syntax of a language shall be analyzed independently of its semantics. As one of the fundamental model ideas of the theoretical component of GG, a grammar is assumed to be a finite set of (recursive) rules by which an infinite set of well-formed sentences can be generated (*assumption of generativity*). The grammar of a language L is identified with a theory from the laws of which all and only grammatical sentences of L can be derived. Hence, the empirical component contains *grammatical structures* of languages as the empirical evidence of the paradigm. In Syntactic Structures, the laws are *phrase structure rules* that, for instance, can be inductively generalized from a specific language by means of constituent analysis. However, GG is not a theory of a certain language but a more general theory of how particular theories of particular languages are to be constructed. In this sense, it is a basic law of GG that all languages can be reconstructed as an ordered pair $\langle \Sigma, F \rangle$ with a set of initial strings Σ and a set of phrase structure rules F such that the derivations from Σ by F terminate, i. e. cannot be rewritten any further by phrase structure rules. I call this law the $\langle \Sigma, F \rangle$ *termination law*.

In order to keep grammatical descriptions as simple as possible and to derive sentences that cannot be generated solely by phrase structure rules, Chomsky ([1957] 1971, ch. 4) suggests a *transformational component*. For example, the transformational component transforms active sentences into passive. The assumption of a transformational component is part of the theory core of GG. Chomsky ([1957] 1971, ch. 6) states that it is the aim of a grammar of a certain language L, i. e. of a theory of L, to explain linguistic observations of L and to predict new observable phenomena of L. Hence, sentences and transformations of particular languages are the successful applications of GG, with generating declarative sentences and passive-transformations or question-transformations being paradigmatic applications introducing and motivating the paradigm. Table 1 presents the emergence of GG in Syntactic Structures (hereafter GG^{SynS}) by summing up the main elements of the theoretical, the methodological, and the empirical component.

Table 1: Theoretical, methodological, and empirical components of GG^{SynS}.

Paradigm Components	GG^{SynS}
theoretical component	generativity $\langle \Sigma, F \rangle$ termination law transformational component
methodological component	independence of grammar
empirical component	grammatical structures

3.2.2 Aspects of the theory of syntax

I take Chomsky's (1965) *Aspects of the Theory of Syntax* (hereafter *Aspects*) as the next landmark of the normal scientific development of GG. I will refer to this developmental step as GG^{Asp}.

First, GG^{Asp} is explicitly located within cognitivism. In contrasts to behavioristic approaches of language and language acquisition proposed by Quine and Skinner (Chomsky 1959), GG^{Asp} is stated to be a theory about the knowledge of language represented in the mind of (idealized) speaker-hearers (Chomsky, 1965, ch. 1).

Second, the theoretical component is augmented by postulating a *lexicon* as a further element of the *deep structure* generating phrase-markers that are interpreted semantically and transformed by the transformational component into *surface structure* sentences (Chomsky, 1965, ch. 3). As opposed to GG^{SynS} the phrase structure rules generate not terminal but preterminal strings that are formed into terminal strings by inserting of lexical formatives from the lexicon (Chomsky, 1965, 84–88). Separating the lexicon from the phrase structure component has the advantage that idiosyncratic morpho-syntactic information is outsourced from the purely syntactic phrase structure rules (*assumption of the demarcation between lexicon and syntax*). For instance, there is no need for different phrase structure rules for transitive verbs that permit or do not permit object deletion. Instead, this information is represented in each lexical entry of a particular transitive verb. Hence, the phrase structure component with the $\langle \Sigma, \mathrm{F} \rangle$ termination law is rearranged to the $\langle \Sigma, F \rangle$ *pretermination law* and a lexicon added as an additional component. Furthermore, the theoretical component is augmented by the *assumption of nativism* stating that there is an innate universal grammar containing linguistic universals common to all languages (Chomsky, 1965, 27 ff.).

Third, the methodological component is sharpened, thereby further constraining the interplay between the theoretical and the empirical component. The requirement of the independence of grammar is maintained (Chomsky, 1965, 141). Chomsky (1965, 3 ff.) introduces the distinction between the *competence* and the *performance* of a speaker-hearer and requires the competence to be the only object of research (*requirement of competence investigation*). Chomsky (1965, 24 ff.) distinguishes between *descriptively adequate linguistic theories* correctly describing the competence of an idealized speaker-hearer of a particular language and *explanatorily adequate linguistic theories* additionally explaining language acquisition, i. e. the *selection* of the right descriptively adequate grammar on the basis of linguistic input compatible with several descriptively adequate grammars. Hence, explanatory adequacy of a linguistic theory presupposes descriptive adequacy. For GG^{Asp} explanatory adequacy is required (*requirement of explanatory adequacy*). In *Aspects*, GG^{Asp} is associated with the epistemological tradition of *rationalism* in contrast to empiricist theories of language (Chomsky, 1965, 51 ff., see also Chomsky, 1966, 59–72). The main argument to take a rationalist point of view is the assumption of nativism, according to which there are innate linguistic universals out of which a descriptively adequate grammar emerges on the basis of observed linguistic data.

Fourth, the empirical component is augmented by integrating the study of language acquisition into the successful applications of the paradigm. That is, beyond the scope of a descriptively adequate linguistic theory aiming to correctly describe the

Table 2: Theoretical, methodological, and empirical components of GG^{Asp}.

Paradigm Components	GG^{Asp}
theoretical component	generativity $\langle \Sigma, F \rangle$ pretermination law demarcation between lexicon and syntax transformational component nativism
methodological component	independence of grammar competence investigation explanatory adequacy rationalism
empirical component	grammatical structures language acquisition

observed linguistic data of particular languages, GG^{Asp} intends to explain the process of language acquisition. Hence, the data of language acquisition is an intended application of GG^{Asp}. Table 2 summarizes the paradigm components of GG^{Asp}.

3.2.3 Principles and parameters

The next developmental step in the normal science of GG is the *principles and parameters approach* introduced in Chomsky (1981, 1986) that I will hereafter refer to by GG^{PP}. From GG^{Asp} to GG^{PP}, the assumption of nativism became a more dominant factor of the theoretical component of the paradigm. Chomsky and Fodor (1980) showed in a number of examples that the acquisition of language is underdetermined by the language input and cannot be understood as an inductive process of developing and verifying hypotheses. Generalizing the examples leads to the well-known *poverty of the stimulus argument* justifying the assumption of nativism. The premises of the argument are, first, that a child acquires its native language in a certain amount of time, second, that the language input is not sufficient for a child to acquire this language, and third, that a language is either acquired completely from experience or innate linguistic knowledge exists.[3] The conclusion is that innate linguistic knowledge exists, which was already anticipated in the normal scientific phases before GG^{PP} (e. g. Chomsky, 1965, ch. 1; Chomsky, 1972, 27). The first two premises of the argument are also known as *Plato's Problem* (Chomsky, 1986, xxv–xxix).

It is a main model idea of the theoretical component of GG^{PP} that the innate linguistic knowledge is based on a set of universal grammatical *principles* common to all languages and a set of *parameters*. Parameters are binary variables switched on

3 Usually, the third premise remains implicit and unmentioned.

or off with respect to whether the language that is acquired has certain grammatical properties or not.

The account of principles and parameters is a theoretical progression with respect to the account of phrase structure rules in the following sense: Phrase structure rules are like *empirical laws* that can be inductively generalized from the observed sentences of a particular language. However, principles and parameters are like hypothetically introduced *theoretical laws* from which grammatical rules for particular languages can be derived deductively. For example, the X′-theory (Jackendoff, 1977) presupposed in GG^{PP} is a theoretization of the phrase structure rules of $GG^{SynS/Asp}$. X′-theory is a universal principle that in conjunction with a lexicon and other principles (e. g. the projection principle) generates syntactic structures.

GG^{PP} postulates a modular cognitive architecture according to which the mind is organized in different *modules* (*assumption of modularity*). Modules are separate cognitive devices that compute given inputs by means of module-specific algorithms. The language faculty is assumed to be an innate module that can be investigated independently of other cognitive modules. The language module is subdivided into different sub-modules (also called *theories*), like binding theory, case theory etc. That is, the modules are like ontological instantiations of the formal theories of GG^{PP}.

The methodological component is extended by the requirement of core-grammar investigation. Chomsky (1981, 7–8) distinguishes between the periphery and the core-grammar of a language. The periphery contains all irregular and idiomatic elements of a language that cannot be derived from principles and parameters but have to be learned and stored in the lexicon. The core-grammar of a language contains all structures that are derivable from the principles and parameters of the syntactic component. Thus, the requirement of core-grammar investigation strongly restricts the intended and programmatic applications of the empirical component to regular grammatical phenomena.

In the empirical component of GG^{PP}, *language typology* is highlighted in accordance with the notion of parameters in the theoretical component introduced to explain systematic grammatical differences between languages. Hence, the empirical component contains the research fields of *language acquisition*, *typology*, and *grammatical structures*. The element of grammatical structures can be considered to have both a universalist part, aiming to characterize a universal grammar for all possible human language, and a particularist part, aiming to provide grammars for particular languages (Newmeyer, 2013). However, the empirical data are only grammatical structures from particular languages, as only concrete structures are observable. In sum, the core-grammars of particular languages, language acquisition, especially with respect to Plato's Problem, and typological data of language comparison shall be explained by universal principles and language-specific parameter settings. Table 3 summarizes the paradigm components of GG^{PP}.

By this time, the following question arises: Why is the development from GG^{SynS} to GG^{PP} considered to be the normal science of one and the same paradigm, and not

Table 3: Theoretical, methodological, and empirical components of GG^{PP}.

Paradigm Components	**GG^{PP}**
theoretical component	generativity principles and parameters modularity demarcation between lexicon and syntax transformational component nativism
methodological component	independence of grammar competence investigation explanatory adequacy core-grammar investigation rationalism
empirical component	grammatical structures language acquisition language typology

rather as distinct paradigms, one replacing the other in a sequence of scientific revolutions? Despite the differences between the developmental steps of GG, there are fundamental elements of the theoretical and the methodological components that are maintained and that are essential for the identity of the GG paradigm – for example the theoretical assumption of generativity or the methodological requirement of the independence of grammar. Further, there is a high similarity between the paradigmatic examples of the different phases motivating the theoretical component of the paradigm. However, as exemplified by the minimalist program in the next section, not every developmental step introduced by Chomsky automatically needs to be considered progress in normal science.

3.2.4 Minimalist program

The next developmental step of GG is the minimalist program (*MP*) (Chomsky, 1995). However, MP exhibits some fundamental differences compared to GG developed so far. Hence the question arises whether MP is a further step in the normal science of GG or whether it leaves the paradigm and constitutes a new one. In the second case, it would be more adequate to characterize the development of MP as revolutionary science. In the following, I give a brief description of MP, followed by a discussion concerning its developmental status from the point of view of the philosophy of science.

MP shares the theoretical assumptions of generativity, of the demarcation between syntax and lexicon, and of nativism (Chomsky, 1995, 1–10 and 167–172). The transformational component is rejected since the distinction between a deep level and a surface level of syntactic structures is abandoned. Instead, syntactic structures are generated by what Chomsky (1995, 249) calls the *bare phrase structure*. The bare phrase structure is mainly based on the operation *merge*, combining lexical items to complex structures, and the operation *move*, restructuring the generated syntactic objects (Chomsky, 1995, 241–255). The operation *merge* avoids several problems of X'-theory, like vacuous projections or redundant distinctions between terminal nodes and lexical items (Nunes, 2013). Therefore, I take the bare phrase structure to be a fundamental theoretical assumption of MP.

MP also assumes a modular cognitive structure, but redefines the assumed modular architecture as follows: MP does not postulate independent sub-modules of the language faculty like binding theory, case theory, etc. (Chomsky, 1995, 170–171; Boeckx, 2015, 436; Hornstein, 2009). Instead, the operations of the language faculty are constrained by two performance systems interfacing with the language faculty: the articulatory-perceptual system and the conceptual-intentional system. The model idea of the restriction of syntactic derivations by the two interface systems motivates the methodological requirement of economy: linguistic expressions are formal objects *optimally* satisfying the conditions of the interface systems (Chomsky, 1995, 171). For example, postulating *merge* to take at least and at most two syntactic objects and hence to produce binary branching structures is considered to satisfy the requirement of economy (Nunes, 2013, 84). Except for focusing on the requirement of economy, the methodological component of MP equals that of GG^{PP}.

Is MP a normal scientific developmental step within the paradigm of GG or is it more adequately characterized as a discrete approach, based on a new and essentially different fundament? On the one hand, there are essential differences in the modular architecture and the assumed syntactic operations between MP and earlier stages of GG. On the other hand, central theoretical assumptions and methodological requirements are kept. I think that it is reasonable to argue that the detailed discussions comparing GG^{PP} and MP lead to the conclusion that they are not incommensurable. From this it follows that the development of MP is not a scientific revolution. However, theoretical changes anticipated for the future development of MP possibly deepen the demarcation between GG^{PP} and MP. For instance, Boeckx (2015, 436) predicts that the notion of parameters will be given up within the minimalist program due to economy reasons. In sum, it seems to be too early to characterize the developmental step from GG^{PP} to MP. From a future perspective it might even be the case that a third category has to be introduced: paradigm development between normal and revolutionary science.

To sum up, it has been shown that GG is a paradigm in the sense of Kuhn. First, it contains all components of a paradigm, and second GG constitutes a normal science carried out by an international scientific community accepting the paradigm as the basis for linguistic research (of course, this does not entail that there are no linguists not using GG). In the next section, I will discuss the non-falsifiability of paradigms – a property of paradigms pointed out by Kuhn's diachronic analyses – with respect to the paradigm of GG.

4 Normal science of GG: non-falsifiability and quasi-analytic statements

According to SSR, a paradigm is an instrument for scientific puzzle-solving used by a certain scientific community to accumulate scientific knowledge in the phase of normal science. For the paradigm of GG, there is an immense amount of research concerning the three elements of the empirical component that cannot be listed let alone discussed here. Instead, I will rather focus on the property of non-falsifiability of a paradigm in normal science. According to Kuhn ([1962] 1996), a paradigm is never given up because of an anomaly – a point of view objecting to Karl Popper's ([1934] 1992) falsificationism. Rather, a paradigm is considered to be a tool for scientific research and a tool would never be disposed of, no matter how deficient it works – unless a better working tool is at hand.

In the following, I will discuss two case studies of the immunization of theory core elements of GG against contradicting experience, preventing a falsification of the theory core. Both case studies lie within the phase of GG^{PP} – probably the most prominent version of GG. In Section 4.1, I will discuss the non-falsifiability of the *modular matching model* of Crain and Thornton (1998) and in Section 4.2, research of the pro-drop parameter and its consequences for the notion of parameters in general.

The aim of the analysis is twofold: First, it will confirm Kuhn's general assumption that paradigms are held on to by a scientific community in spite of anomalies. Second, the epistemological status of the laws and model ideas of the theory core is an important prerequisite to understand the incommensurability of paradigms (sec. 5).

4.1 The modular matching model

The first case study is concerned with language acquisition. Crain and Thornton (1998) developed the *modular matching model* (hereafter *MMM*) presupposing the theoretical component of GG^{PP}, especially the assumptions of modularity and nativism (Crain and Thornton, 1998, ch. 1–4). MMM adds a further theoretical assumption that I will call the *assumption of full processing* referring to Drozd (2004). According to the assumption of full processing, "children and adults share a common language processing system [and] appeal to the same processing mechanisms of language production and understanding" (Crain and Thornton, 1998, 6). Hence, the assumption of full processing rules out any language processing differences between children and adults with respect to the modular organisation of innate linguistic knowledge. In sum, MMM is a specialization of GG^{PP} used to focus on language acquisition as one particular element of the empirical component of GG^{PP}. However, as Crain and Thornton (1998, ch. 14) themselves report, MMM fails in experiments on scopus interpretations of ambiguous sentences like (4.i) with two possible interpretations (4.ii) and (4.iii).

(4.i) *The big elephant is the only one playing the guitar*.

(4.ii) *The only thing playing the guitar is the big elephant*.

(4.iii) *The only elephant playing the guitar is the big elephant*.

(4.ii) is called the *subset reading* and (4.iii) the *superset reading* because (4.ii) entails (4.iii) but not the other way around. According to MMM, both children and adults should prefer the same interpretation. However, experiments show that children prefer the subset reading and adults prefer the superset reading (Crain and Thornton, 1998, ch. 14). Hence, the experiment is a falsification of MMM. However, instead of giving up MMM or revising at least one of its components, it is upheld by introducing a new ad hoc hypothesis called the *semantic subset principle*. According to the semantic subset principle, children prefer the subset reading of ambiguous sentences because of certain properties of language acquisition. The semantic subset principle is assumed to be an element of the *language acquisition device*, which is a cognitive module supporting language acquisition. Crain and Thornton (1998, 115) do not assess the experimental results on ambiguous sentences as falsifications of MMM, but, on the contrary, present the results as “empirical support for the kind of modular architecture that underlies the Modularity Matching Model”. Furthermore, they present four additional methodical factors explaining differences in language processing between children and adults:

> We discuss four sources of children’s nonadult behavior: (a) artifacts in experimental design, (b) extralinguistic knowledge that develops over time, (c) the complexity of nonlinguistic processes that are needed to provide correct answers in a psycholinguistic task, and (d) performance factors that influence the linguistic behavior of both children and adults. (Crain and Thornton, 1998, 121)

The factors (a)–(d) are not explicitly defined. For instance, referring to factor (c) it is unclear at which level of complexity of nonlinguistic processes the experiment fails. Rather, (c) seems to be intended to be interpreted as follows: If an experiment leads to non-predicted results, then the complexity of nonlinguistic processes was too high to get valid results. Crain and Thornton (1998, 123) themselves state that “even consistent nonadult responses in such circumstances cannot be construed as evidence of children’s nonadult language processing”. Hence MMM, which entails the theory core of GG^{PP}, seems to be completely immunized against falsifying experiences with respect to the study of language acquisition (Akhtar, 2004, 460; Kornmesser, 2014a).

4.2 The pro-drop parameter

The second case study is concerned with typology. In GG^{PP} systematic grammatical differences between languages are explained by parameters. A parameter is considered to be a cognitive switch that is switched on or off with respect to the grammatical properties of the language in question. However, there is not one parameter for each grammatical property – this would be a mere increase of ontology without any empirical or explanatory value added. Instead, there can be several grammatical properties that are explained by one parameter being switched on or off. This is what Chomsky (1981, 241) calls a *clustering of properties*. In GG^{PP}, the *pro-drop parameter* (also called *null subject parameter*)[4] is the prototypical parameter used in most introductory textbooks in order to introduce the notion of parameters (e. g. Haegeman, 1994; Cook and Newson, 1996).[5]

Kayne (1980) correlated the grammatical properties of *null thematic subject* (*NTS*), *null non-thematic subject* (NNTS), *subordinate subject extraction* (SSE) and *subject inversion* (*SI*) for a number of languages and explained the correlation by a common cause that in GG^{PP} is referred to as the pro-drop parameter.[6]

In sentences with null thematic subjects, the referring subjects are omitted without the sentences becoming ungrammatical. Null thematic subjects are, for example, possible in a pro drop language like Spanish, but not in standard English.

(4.iv) *We have worked all day*.

(4.v) **Have worked all day*.

(4.vi) *Hemos trabajado todo el día*.
‘(We) *have worked all day.’

In sentences with null non-thematic subjects, non-referring subjects are omitted without the sentences becoming ungrammatical. Null non-thematic subjects are, for example, possible in pro-drop languages like Spanish, but not in standard English.

(4.vii) *It’s raining*.

(4.viii) **raining*.

(4.ix) *Llueve*.
‘(It’s) *raining.’

In sentences with subordinate subject extraction,[7] the subjects of subordinate clauses introduced by “that” can be questioned without the sentences becoming ungram-

4 For a brief discussion about which terminology is adequate with respect to the presupposed theoretical apparatus see D’Alessandro (2015, 201–203).

5 In the following, I will discuss the classical notion of the pro-drop parameter (see D’Alessandro, 2015) and not more recent approaches like the distinction between macro and micro parameters (see Baker 2008).

6 Research on the pro-drop parameter can be traced back to Perlmutter (1971) who correlated the properties NTS, NNTS, and SSE. He did not call the explaining principle a *parameter*, but in the generative literature, he is stated to be one of the initiators of parameter research as Gilligan (1987, 76) states: “[T]he data Perlmutter noted is the first cross-linguistic generalization, i. e., parameter, in GG. For better or worse, it has served as the basis for all subsequent work on Pro-drop phenomena.”

7 Since the publication of Chomsky and Lasnik (1977), constructions of this kind have usually been referred to as *that-t filter violations*. However, in this paper we use the original conception of Perlmutter matical. This is possible in pro-drop languages like Spanish, but not in standard English.

(4.x) **What did you say that happened?*

(4.xi) *Qué dijiste que pasó?*
'What did you say (*that) happened?'

In sentences with subject inversion, the subject is moved to a postverbal position without the sentence becoming ungrammatical. This is possible in pro-drop languages like Italian, but not in standard English.

(4.xii) *Sono arrivati molti amici.*
'*Have arrived many friends.'

The grammatical properties NTS, NNTS, SSE, and SI correlate for the languages English, French, Italian and Spanish. That is, for each of those languages either all of the properties are grammatical or all of them are ungrammatical. The correlations are explained by the pro-drop parameter assumed to be the common cause of the correlations. Since parameters are innate and universal, all languages can be distinguished into *pro-drop languages*, in which the properties are grammatical, and non-pro-drop languages, in which they are not.

Logically, the concepts *pro-drop language* and *non-pro-drop language* are both empirically *operationalized* with respect to the properties NTS, NNTS, SSE and SI, resulting in grammatical or ungrammatical sentences in particular languages. An operationalized concept is determined by a certain reaction to a certain test condition. The logical structure of operationalizations can be explicated by *bilateral reduction sentences* (Carnap, 1936). Bilateral reduction sentences require first order predicate logic and – for the operationalizations of the concepts *pro-drop language* and *non-pro-drop language* – the following predicates (Kornmesser, 2016, 2017):[8] "NL(y)" (y is a natural language), "G(x,y)" (x is grammatical in y), "NTS(x,y)", "NNTS(x,y)", "SSE(x,y)", and "SI(x,y)" (x is a sentence with the grammatical property NTS, NNTS, SSE or SI, respectively, in y); "PDL(y)" (y is a pro-drop language). The determination of the concept *pro-drop language* (PDL) by Kayne (1980) can be expressed by the operationalizations $\mathrm{OP1}^{\mathrm{Kayne}}$–$\mathrm{OP4}^{\mathrm{Kayne}}$.

($\mathrm{OP1}^{\mathrm{Kayne}}$) $\forall x, y\, (NL\,(y) \wedge NTS\,(x, y) \rightarrow (PDL\,(y) \leftrightarrow G\,(x, y)))$

($\mathrm{OP2}^{\mathrm{Kayne}}$) $\forall x, y\, (NL\,(y) \wedge NNTS\,(x, y) \rightarrow (PDL\,(y) \leftrightarrow G\,(x, y)))$

($\mathrm{OP3}^{\mathrm{Kayne}}$) $\forall x, y\, (NL\,(y) \wedge SSE\,(x, y) \rightarrow (PDL\,(y) \leftrightarrow G\,(x, y)))$

($\mathrm{OP4}^{\mathrm{Kayne}}$) $\forall x, y\, (NL\,(y) \wedge SI\,(x, y) \rightarrow (PDL\,(y) \leftrightarrow G\,(x, y)))$

(1971) which in accordance with Gilligan (1987, 105) is assumed to address almost the same grammatical phenomena.

8 For a more detailed reconstruction, see Kornmesser (2014b).

As mentioned above, operationalized concepts are determined by means of a certain reaction with respect to a certain test condition. The test condition is the antecedent of the implication expressing that y is a natural language and x a sentence in y with a certain grammatical property (NTS, NNTS, SSE or SI). The reaction is whether x is judged to be grammatical or not for y (G(x,y)) – for example by a native speaker – determining whether y is a pro-drop language (PDL) or not.

The operationalization of a concept is distinct from the definition of a concept (Kornmesser, 2016, 2017, 2018). An important difference is the *empirical creativity* of operationalizations. Definitions are analytic statements that do not have any empirical content. In contrast, the multiple operationalization of a concept entails empirical consequences that can be confirmed or falsified empirically. From $OP1^{Kayne}$–$OP4^{Kayne}$ the following empirical laws entailed by the multiple operationalization of PDL can be derived:

EmpLaw 1: $\forall w, x, y\,(NL\,(y) \wedge NTS\,(x, y) \wedge NNTS\,(w, y) \rightarrow (G\,(w, y) \leftrightarrow G\,(x, y)))$

EmpLaw 2: $\forall w, x, y\,(NL\,(y) \wedge NTS\,(x, y) \wedge SSE\,(w, y) \rightarrow (G\,(w, y) \leftrightarrow G\,(x, y)))$

EmpLaw 3: $\forall w, x, y\,(NL\,(y) \wedge NTS\,(x, y) \wedge SI\,(w, y) \rightarrow (G\,(w, y) \leftrightarrow G\,(x, y)))$

EmpLaw 4: $\forall w, x, y\,(NL\,(y) \wedge NNTS\,(x, y) \wedge SSE\,(w, y) \rightarrow (G\,(w, y) \leftrightarrow G\,(x, y)))$

EmpLaw 5: $\forall w, x, y\,(NL\,(y) \wedge NNTS\,(x, y) \wedge SI\,(w, y) \rightarrow (G\,(w, y) \leftrightarrow G\,(x, y)))$

EmpLaw 6: $\forall w, x, y\,(NL\,(y) \wedge SSE\,(x, y) \wedge SI\,(w, y) \rightarrow (G\,(w, y) \leftrightarrow G\,(x, y)))$

For example, EmpLaw 1 states that if in a natural language y a sentence x with the grammatical property NTS is constructed and a sentence w with the property NNTS, then w is grammatical in y if and only if x is grammatical in y. None of the empirical laws continue to contain the operationalized concept PDL anymore and thus they are purely empirical statements. However, all of the empirical laws are derived from the multiple operationalization of the concept PDL in $OP1^{Kayne}$–$OP4^{Kayne}$. This relation logically explicates what is meant by the pro-drop parameter being the common cause for the empirical correlations stated by Kayne (1980) for a certain sample of languages. The languages that satisfy EmpLaw 1–6 are successful applications of the pro-drop parameter research. The languages French, English, Italian and Spanish can be considered to be the paradigmatic applications. Both paradigmatic and successful applications are part of the typological element of the empirical component of GG^{PP}.

Rizzi (1982) and Safir (1985) have weakened the operationalizations of Kayne (1980) due to the integration of new languages into the set of successful applications that did not satisfy all of the empirical consequences of $OP1^{Kayne}$–$OP4^{Kayne}$. However, the set of successful applications was still comparatively small. According to the theoretical assumption of nativism, parameters must be universally applicable. Hence in order to empirically justify the assumption of a parameter, more comprehensive sets of successful applications are needed. Auwera (1984) applied the operationalizations of PDL to a set of 13 languages, and Gilligan (1987) applied PDL to a set of hundred languages systematically chosen from the world's language families (Gilligan, 1987, 102–104). The combination of the surveys of Auwera (1984) and Gilligan (1987) rules out almost all empirical consequences of Kayne's operationalization of PDL, except for the following:

$$\text{EmpLaw } 1^{\text{Au-Gil}}\!: \quad \forall w, x, y\, (NL\,(y) \wedge NTS\,(x, y) \wedge NNTS\,(w, y) \rightarrow (G\,(w, y) \leftarrow G\,(x, y)))$$

Consequently, the operationalizations of PDL have to be weakened extensively, with only the following operationalizations remaining empirically adequate:

$$(\text{OP1}^{\text{Au-Gil}}) \quad \forall x, y\, (NL\,(y) \wedge NTS\,(x, y) \rightarrow (PDL\,(y) \leftarrow G\,(x, y)))$$

$$(\text{OP2}^{\text{Au-Gil}}) \quad \forall x, y\, (NL\,(y) \wedge NNTS\,(x, y) \rightarrow (PDL\,(y) \leftrightarrow G\,(x, y)))$$

In sum, the beginning of pro-drop parameter research seemed to be a successful project due to the empirical applicability to a set of successful applications and a multiple operationalization of PDL by means of four independent grammatical properties. However, with respect to $\text{OP1}^{\text{Au-Gil}}$ and $\text{OP2}^{\text{Au-Gil}}$, the content of PDL is nearly empty. It could even be argued that EmpLaw $1^{\text{Au-Gil}}$ is not an empirical law at all but a purely analytic statement because all it says is that if subjects can be omitted in a language, then a special subclass of subjects (non-thematic subjects) can be omitted, too. In this interpretation, there is no evidence for an innate pro-drop parameter left.

What does this mean for GG^{PP}? Is the central theoretical assumption of principles and parameters refuted? This does not seem to be the case, because even if the pro-drop parameter is given up, there are other parameters postulated to exist. However, as Newmeyer (2004, 2005) argues, there is no convincing evidence that parameters exist at all. I call this problem the *parameter anomaly*. Does the parameter anomaly force the scientific community to give up GG^{PP}? The answer is no again, in full accordance with Kuhn's assumption of the non-falsifiability of paradigms.

There are several ways to react to the parameter anomaly. First, even the falsification of all parameters assumed to exist so far does not falsify the theoretical statement that there are parameters. For logically, this is an existential statement the negation of which cannot be empirically verified (Popper [1934] 1992, ch. III, § 15). Second, the outcome of Gilligan's (1987) cross-linguistic survey can be reinterpreted to be a success. For example, Roberts and Holmberg (2005, 544) argue with respect to Gilligan's (1987) negative results that the "fact that any coherent patterns survived is telling and a sign that Rizzi's observations were clearly on the right track". Third, new theoretical entities and relations can be introduced in order to principally explain counterexamples as, for instance, parameter interaction (Biberauer, 2008, 15–16) or a hierarchy of parameters (Baker, 2001). Fourth, it is possible to keep the overall paradigm of GG but to reject the existence of parameters (Newmeyer, 2004, 2005). However, fifth, it might be the case that the parameter anomaly is accepted to be a serious anomaly leading GG^{PP} into a paradigm crisis. In this case, a new paradigm may emerge focusing on typology – and eventually also on the other elements of the empirical component of GG – from a totally different point of view (see sec. 5).

To sum up: The case studies show that scientific behaviour in normal science does not follow Popper's verdict of falsificationism. Rather, a paradigm is maintained by a scientific community even in case of anomalies in order to proceed with normal science. The fundamental theoretical assumptions can always be immunized against falsifying experiences. For this reason, Kuhn characterizes the elements of the theory core as *tautologies* or *quasi-analytic statements* (Kuhn, 1977, 468–487).[9] He often refers to Ohm's law or to Newton's second law that "behaves for those committed to Newton's theory very much like a purely logical statement that no amount of observation could refute" (Kuhn, 1996, 78). As I tried to exemplify in this section, the same holds for GG: The statements of the theory core are quasi-analytic statements that are empirically not falsifiable and behave for the scientific community "committed to [GG] very much like a purely logical statement that no amount of observation could refute". This is not to say that it is *in principle* not an empirical question whether, for example, parameters or innate cognitive modules exists or not, but *within* GG^{PP} these assumptions become quasi-analytic because of their central status within the theoretical component.

From a normative point of view, it may be argued that this is inadequate scientific behaviour. However, from a descriptive point of view, that is how normal science seems to work – as long as the paradigm can be used as a tool for the explanation and exploration of empirical phenomena. The instruments of immunization correspond to what Lakatos (1970) calls a *protection belt* and a *negative heuristic* of a research program protecting the hard core of the program. However, if anomalies accumulate, a new paradigm may be established by a new scientific community (see sec. 5).

5 Coexisting competing paradigms and incommensurability

In a scientific revolution a new paradigm P^2 comes into play in which the quasi-analytic statements of the already existing paradigm P^1 become false. This causes incommensurability of P^1 and P^2 with respect to the central theoretical concepts because the quasi-analytic statements of P^1 are constitutive for the meanings of those concepts of P^1 occurring in the quasi-analytic statements due to their empirical non-falsifiability (Hoyningen-Huene, 1993, 210–211).[10] In Kuhn's account of scientific rev-

9 A *quasi-analytic* statement behaves like an analytic statement although it is neither a logical truth nor a definition.

10 Beside the theoretical concepts, in a scientific revolution the pre-theoretical categorization of the phenomenal world also changes. However, in this article I focus on incommensurability caused by changes on the theoretical level. For an extensive discussion, see Kornmesser (2014a, 2014c).

olutions, P^2 replaces P^1 maintaining the monoparadigmatic structure of science. However, in a number of case studies it has been shown that for several scientific disciplines, it is a more adequate view to assume a multiparadigmatic structure of science (Kornmesser and Schurz, 2014). This book being a linguistic handbook, I will discuss the case of two linguistic paradigms that aim to answer nearly the same research questions. As we will see, there is no replacement of one of the paradigms by the other one. Hence, we have coexisting incommensurable paradigms competing with respect to shared programmatic goals.

In Section 5.1, I will briefly introduce the paradigm of construction grammar which developed approximately parallel to GG^{PP}. In Section 5.2, I will contrast both paradigms and discuss examples of semantic concept shifts causing *semantic incommensurability* of both paradigms as well as methodological disagreements causing *methodological incommensurability*.

5.1 The paradigm of Construction Grammar

The paradigm of construction grammar (CG) has been constituted in several approaches, some of the most important of which are Fillmore et al. (1988), Goldberg (1995, 2006), Lakoff (1987), Langacker (1987), Croft (2001), and Tomasello (2003). Distilling the common ground of the (not exhaustive list of) construction grammar approaches results in theoretical, methodological and empirical components constituting a paradigm in the sense of Kuhn, shared by a stable scientific community.

In CG, the basic units of grammatical analysis are *constructions*. A construction is a sign consisting of a certain form and a certain meaning. The paradigmatic applications of CG are idiosyncratic elements of a language the form or the meaning of which cannot be derived from more general semantic (5.i) or syntactic rules (5.ii) (Goldberg, 1995, 4).

(5.i) *kick the bucket*

(5.ii) *all of a sudden*

The expressions (5.i) and (5.ii) both have a certain form and a certain meaning and are hence constructions. In GG idioms like (5.i) and (5.ii) would be assumed to be part of the lexicon containing all idiosyncratic elements of a language. CG, however, claims that there are also productive constructions, i. e. constructions that are not specific with respect to their semantics or their syntax like (5.i) and (5.ii) but that are schematic form-meaning-pairs that can be instantiated in concrete expressions. For example, the ditransitive construction with the schematic form [subj verb obj_1 obj_2] can be instantiated in the concrete sentence (5.iii) (Goldberg, 1995, 141).

(5.iii) *Sally baked her sister a cake.*

The sentence (5.iii) means that Sally bakes a cake with the intention of transferring the cake to her sister. According to CG, the intended-transfer-meaning is not part of the verb *to bake*, but of the schematic ditransitive construction inducing this meaning to its instantiations like in (5.iv).

(5.iv) *Sally paints her sister a picture*.

Thus constructions can be schematic and productive, taking on the task of phrase structure rules or principles and parameters in GG (*assumption of schematic constructions*).

Constructions are form-meaning-pairs. This is why grammatical analysis is always inherently semantic analysis, contrary to the requirement of the independence of grammar in GG. For example, the ditransitive construction has an idiosyncratic meaning that is entailed in the meaning of each of its instantiations. Therefore, on the one hand it cannot be part of a mental lexicon *in spite of* its idiosyncrasy since it is productive. On the other hand, the ditransitive construction cannot be part of what GG calls the core grammar of a language *because of* its idiosyncrasy. This is why CG refuses the demarcation between lexicon and syntax as it is postulated by GG and assumes a *lexicon-syntax-continuum* from specific to schematic constructions. Words are on the one end of the continuum since they are by definition the most specific kind of constructions because of their idiosyncratic form as well as their idiosyncratic meaning. Highly schematic constructions like the ditransitive construction are at the other end of the continuum. However, specific and schematic constructions are principally of the same kind, unlike in GG which principally distinguishes between words on the one hand and a productive syntax containing rules or principles on the other hand. From this point of view, the sentences of a language are not *generated* by a formal component based on phrase structure rules or innate principles, but are *instantiations* of certain patterns of form and meaning.

In the following, I call this point of view concerning the production and understanding of sentences the *assumption of productivity by instantiation*, contrasting with the assumption of generativity referring to the formal derivation of sentences in GG. For the same reason, CG refuses to accept the existence of a transformational component (*assumption of monostratality*). For instance, there are no rules transforming an active sentence into a passive sentence, but there are different constructions for active and passive differing semantically in which participant of an action is brought into focus.

Furthermore, CG does not postulate a modular architecture of the mind including language specific modules, but presupposes domain-general cognitive mechanisms for learning and processing languages. In what follows, I call this assumption the *assumption of cognitive holism*.

Constructions are learned from experience: specific constructions are learned directly from perceived utterances, and schematic constructions are inductively generalized from perceived instantiations of schematic constructions (*assumption of in-*

ductivism). Consequently, CG denies the existence of innate linguistic knowledge and shares the goals of John Locke's ([1690] 1961, 9) empiricism who aims to explain "how men, barely by the use of their natural faculties, may attain to all the knowledge they have, without the help of any innate impressions, and may arrive at certainty without any such original notions or principles". In this sense, CG is part of the epistemological *tradition of empiricism*, as opposed to GG that is part of the rationalist tradition.

In accordance with the assumption of inductivism and the empiricist tradition, methodologically CG requires that the explanation of the acquisition and the cognitive representation of linguistic knowledge is based on the actual *language use* (*requirement of usage basedness*). The methodological requirement of usage basedness also contrasts with fundamental methodological requirements of GG. First, in usage-based theories, the distinction between linguistic competence and performance becomes inadequate since linguistic knowledge (competence) is reconstructed from language use (performance). Second, constructions identified in the actual language use are the central units of language acquisition as well as of the grammatical description of a language. Hence, the distinction between descriptive and explanatory adequacy becomes needless. Third, the distinction between core and periphery of a language is inadequate because neither specific constructions nor instantiations of schematic constructions are in any sense prior in the analysis of language use.

In sum, CG can be reconstructed as a paradigm in the sense of Kuhn containing a theoretical, a methodological, and an empirical component (see Table 4).

5.2 The multiparadigmatic structure of linguistics

In this section, I will contrast the paradigm of CG with GG^{PP}. The comparison of both paradigms will show that linguistics has a multiparadigmatic structure containing competing paradigms coexisting.[11] Table 4 contains the theoretical, methodological and empirical components of both paradigms.

As the empirical components of Table 4 show, GG^{PP} and CG aim to explain the same research questions.[12] However, the empirical research is based on fundamentally different methodological and theoretical components.

11 However, this is of course not to say that CG and GG are the only paradigms in linguistics. If more paradigms are taken into account, the question arises whether there are fundamental differences between the paradigms as it is the case for GG and CG, or whether there is a family resemblance between them, making them members of the same *paradigm family* (Kornmesser, 2014a). For example, GPSG, HPSG and Chomsky's generative grammar can be considered members of the same paradigm family, with generative grammar being the prototype of the family to which the other members are more or less similar.

12 GG and CG accounting for the same research questions is the main argument for considering GG and CG to be two paradigms in one and the same science (linguistics) and not to be distinct sciences.

Table 4: Structure of the paradigms GG^{PP} and CG.

Paradigm Components	**GG^{PP}**	**CG**
theoretical component	generativity	productivity by instantiation
	principles and parameters	schematic constructions
	modularity	cognitive holism
	demarcation between lexicon and syntax	lexicon-syntax-continuum
	transformational component	monostratality
	nativism	inductivism
methodological component	independence of grammar	constructions as form-meaning-pairs
	competence investigation	
	explanatory adequacy	usage-based
	core-grammar investigation	
	rationalism	empiricism
empirical component	grammatical structures	grammatical structures
	language acquisition	language acquisition
	language typology	language typology

The different methodological components already discussed in Section 5.1 cause *methodological incommensurability*. For example, both paradigms disagree about which kind of empirical phenomena is relevant for theory building and theory testing.[13] Grammatical analyses within CG start with idiosyncratic expressions of languages, considering more abstract form-meaning-pairs also to be constructions and hence to be principally of the same kind as words and idioms. GG, by contrast, rules out all idiosyncratic expressions by the methodological requirements of core-grammar investigation and independence of syntax. Hence, the *paradigmatic* applications of CG (idiosyncratic expressions) are not included in the successful applications of GG because they belong to the periphery of a language. This difference in the empirical applications is deeply related to the different methodologies of both paradigms and leads to incommensurable theoretical assumptions, like the demarcation of lexicon and syntax in GG and the syntax-lexicon-continuum in CG.

Fundamental differences between the theoretical components of two paradigms lead to what Sankey and Hoyningen-Huene (2001) call *semantic incommensurability*. In Section 4, I analyzed in two case studies examples of quasi-analytic statements of the theory core of GG^{PP} that are immune to empirical falsification. Because of their non-falsifiability, quasi-analytic statements are constitutive for the meaning of the concepts they contain (Hoyningen-Huene, 1993, 210–211). For example, within GG the feature *has a modular structure* is part of the intension of the concept *mind*, and the features *is partially innate* and *contains parameters* are part of the intension of the

13 See for example Croft (2013) and Baker (2013) on hypothesis formation and hypothesis testing in CG and GG.

concept *linguistic knowledge*. That is to say, within GG it is not an empirical question whether statements like *The mind has a modular structure* or *Linguistic knowledge is partially innate and contains parameters* are true: it is more reasonable to consider these statements to be linguistic conventions within the paradigm because of their immunization against contradicting evidence. Semantic incommensurability of two paradigms P^1 and P^2 arises if quasi-analytic statements of P^1 become invalid in P^2. As Table 4 shows, this is the case for the elements of the theoretical components of GG^{PP} and CG. Further, the theoretical assumptions of CG can also be considered quasi-analytic statements within CG. For instance, in CG, it is not an empirical question whether all linguistic knowledge is learned from experience by means of domain-general cognitive mechanisms, rather it is presupposed (Goldberg, 2009, 203). In sum, the differences between the theoretical components of GG and CG constitute semantic incommensurability of both paradigms.

6 Final conclusions

In Section 3, I showed that GG satisfies all conditions to be a paradigm, establishing a normal scientific development within linguistics. According to Kuhn, a paradigm is a tool for scientific exploration containing a theory core, parts of which are immunized against empirical anomalies. In Section 4, I discussed two case studies of empirical immunization in GG resulting in non-falsifiable and hence quasi-analytic statements. In contrast to Kuhn's monoparadigmatic view of science, in Section 5 I introduced the paradigm of CG, arguing that GG and CG are co-existing paradigms that compete with each other in answering almost the same research questions. However, none of the paradigms is replaced by the other one. Thus, it is more adequate to view linguistics as a multiparadigmatic science instead of a pre-scientific, non-mature discipline which is the consequence that Kuhn would draw. Finally, I discussed the methodological and semantic incommensurability of both paradigms. The methodological incommensurability can be directly read from the methodological components of both paradigms as presented in Table 4. The semantic incommensurability follows from the quasi-analytic statements of one paradigm that are not accepted in the other one, leading to a semantic concept shift between both paradigms. The methodological and semantic incommensurability causes difficulties in the comparison of both paradigms, resulting in a long-lasting constellation of co-existence.

References

Akhtar, Nameera. 2004. Nativist versus constructivist goals in studying child language. *Journal of Child Language* 31. 459–462.

van der Auwera, Johan. 1984. Subject vs. non-subject asymmetries in the relativization of embedded NP's. In Wim de Geest & Yvan Putseys (eds.), *Sentential Complementation. Proceedings of the International Conference held at UFSAL, Brussels, June 1983*, 257–269. Dordrecht: Foris Publications.

Baker, Mark C. 2001. *The Atoms of Language*. New York: Basic Books.

Baker, Mark C. 2008. The macroparameter in a microparametric world. In Theresa Biberauer (ed.), *The Limits of Syntactic Variation*, 351–374. Amsterdam & Philadelphia: John Benjamins Publishing Company.

Baker, Mark C. 2013. Hypothesis Testing in the Minimalist Program. In Silvia Luraghi & Claudia Parodi (eds.), *The Bloomsbury Companion to Syntax*, 22–31. London: Bloomsbury.

Biberauer, Theresa. 2008. Introduction. In Theresa Biberauer (ed.), *The Limits of Syntactic Variation*, 1–72. Amsterdam & Philadelphia: John Benjamins Publishing Company.

Boeckx, Cedric. 2015. Linguistic Minimalism. In Bernd Heine & Heiko Narrog (eds.), *The Oxford Handbook of Linguistic Analysis*, 2nd edn., 429–448. Oxford University Press.

Carnap, Rudolf. 1936. Testability and Meaning. *Philosophy of Science* 3. 419–471.

Chomsky, Noam. 1959. Review of *'Verbal Behavior'*. By B. F. Skinner. (The Century Psychology Series.) Pp. viii, 478. New York: Appleton-Century-Crofts, Inc. 1957. *Language* 35(1). 26–58.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, Mass.: The MIT Press.

Chomsky, Noam. 1966. *Cartesian Linguistics*. New York & London: Harper & Row.

Chomsky, Noam. 1971. *Syntactic Structures*. Ninth Printing. Paris: Mouton. First published 1957.

Chomsky, Noam. 1972. *Language and Mind*, Enlarged Edition. New York, Chicago, San Francisco & Atlanta: Harcourt Brace Jovanovich.

Chomsky, Noam. 1981. *Lectures on Government and Binding*. Dordrecht: Foris Publications.

Chomsky, Noam. 1986. *Knowledge of Language*. New York: Praeger.

Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, Massachusetts: The MIT Press.

Chomsky, Noam & Jerry A. Fodor. 1980. The Inductivist Fallacy. In Massimo Piatelli-Palmarini (ed.), *Language and Learning. The Debate between Jean Piaget and Noam Chomsk*, 255–275. Cambridge, Mass.: Harvard University Press.

Chomsky, Noam & Howard Lasnik. 1977. Filters and Control. *Linguistic Inquiry* 8. 425–504.

Cook, Vivian J. & Mark Newson. 1996. *Chomsky's Universal Grammar. An Introduction*, 2nd edn. Oxford & Cambridge: Blackwell.

Crain, Stephen & Rosalind Thornton. 1998. *Investigations in Universal Grammar. A Guide to Experiments on the Acquisition of Syntax and Semantics*. Cambridge, Mass. & London: The MIT Press.

Croft, William. 2001. *Radical Construction Grammar. Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press.

Croft, William. 2013. Hypothesis Formation. In Silvia Luraghi & Claudia Parodi (eds.), *The Bloomsbury Companion to Syntax*, 11–21. London: Bloomsbury.

D'Alessandro, Roberta. 2015. Null subjects. In Antonio Fabregas, Jaume Mateu & Michael T. Putnam (eds.), *Contemporary Linguistic Parameters*, 201–226. Bloomsbury Academic.

Drozd, Kenneth F. 2004. Learnability and Linguistic Performance. *Journal of Child Language* 31. 431–457.

Fillmore, Charles J., Paul Kay & Mary C. O'Connor. 1988. Regularity and idiomacity in grammatical constructions: The case of *let alone*. *Language* 64(3). 501–538.

Gilligan, Gary M. 1987. A Cross-Linguistic Approach to the Pro-Drop Parameter. University of California. Ph. D. Dissertation.

Goldberg, Adele E. 1995. *A Construction Grammar Approach to Argument Structure*. Chicago & London: The University of Chicago Press.

Goldberg, Adele E. 2006. *Constructions at work. The nature of generalization in language*. Oxford: Oxford University Press.

Goldberg, Adele E. 2009. Constructions work. *Cognitive Linguistics* 20(1). 201–224.

Haegeman, Liliane. 1994. *Introduction to Government & Binding Theory*, 2nd edn. Oxford & Cambridge: Blackwell.

Hornstein, Norbert. 2009. *A Theory of Syntax*. Cambridge University Press.

Hoyningen-Huene, Paul. 1993. *Reconstructing Scientific Revolutions*. Chicago & London: The University of Chicago Press.

Jackendoff, Ray. 1977. *X-bar syntax: A study of phrase structure* (Linguistic Inquiry Monograph 2). Cambridge, MA: MIT Press.

Kayne, Richard S. 1980. Extensions of Binding and Case-Marking. *Linguistic Inquiry* 11. 75–96.

Kornmesser, Stephan. 2014a. Rivalisierende Paradigmen in der Linguistik: Generative Grammatik und Konstruktionsgrammatik. In S. Kornmesser & G. Schurz (eds.), *Die multiparadigmatische Struktur der Wissenschaften*, 229–270. Wiesbaden: Springer VS.

Kornmesser, Stephan. 2014b. Model-Based Research Programs. *Conceptus* 41. 135–187.

Kornmesser, Stephan. 2014c. Scientific Revolutions without Paradigm-Replacement and the Coexistence of Competing Paradigms: The Case of GG and CG. *Journal for General Philosophy of Science* 45. 91–118.

Kornmesser, Stephan. 2016. A frame-based approach for theoretical concepts. *Synthese* 193. 145–166.

Kornmesser, Stephan. 2017. A Frame-Based Approach for Operationalized Concepts. In Michela Massimi, Jan-Willem Romeijn & Gerhard Schurz (eds.), *EPSA 15 Selected Paper*, 205–220. Cham: Springer.

Kornmesser, Stephan. 2018. Frames and concepts in the philosophy of science. *European Journal for Philosophy of Science* 8. 225–251.

Kornmesser, Stephan & Gerhard Schurz (eds.). 2014. *Die multiparadigmatische Struktur der Wissenschaften*. Wiesbaden: Springer VS.

Kuhn, Thomas S. 1977. Second Thoughts on Paradigms. In Frederick Suppe (ed.), *The Structure of Scientific Theories*, 2nd edn., 459–482. Urbana, Chicago & London: University of Illinois Press.

Kuhn, Thomas S. 1996 [1962]. *The Structure of Scientific Revolutions*, 3rd edn. Chicago & London: The University of Chicago Press.

Kuhn, Thomas S. 2000. What are Scientific Revolutions. In James Conant & John Haugeland (eds.), *The Road since Structure*, 13–32. Chicago & London: The University of Chicago Press.

Lakoff, George. 1987. *Women, Fire, and Dangerous Things. What Categories Reveal about the Mind*. Chicago & London: The University of Chicago Press.

Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar. Volume I. Theoretical Prerequisites*. Stanford, California: Stanford University Press.

Lakatos, Imre. 1970. Falsification and the Methodology of Scientific Research Programmes. In Imre Lakatos & Alan Musgrave (eds.), *Criticism and the Growth of Knowledge*, 91–196. Cambridge: Cambridge University Press.

Lasnik, Howard & Terje Lohndal. 2013. Brief overview of the history of generative syntax. In Marel den Dikken (ed.), *The Cambridge handbook of Generative Syntax*, 26–60. Cambridge University Press.

Locke, John. 1961 [1690]. *An Essay Concerning Human Understanding*. London: J. M. Dent & Sons LTD.

Masterman, Margaret. 1970. The Nature of a Paradigm. In Imre Lakatos & Alan Musgrave (eds.), *Criticism and the Growth of Knowledge*, 59–89. Cambridge: At the University Press.

Newmeyer, Frederick J. 2004. Against a parameter-setting approach to typological variation. In Pierre Pica, Johan Rooryck & Jeroen van Craenenbroeck (eds.), *Linguistic Variation Yearbook, Volume 4*, 181–234. Amsterdam: Benjamins.
Newmeyer, Frederick J. 2005. *Possible and Probable Languages. A Generative perspective on Linguistic Typology*. Oxford: Oxford University Press.
Newmeyer, Frederick J. 2013. Goals and methods of generative syntax. In Marel den Dikken (ed.), *The Cambridge Handbook of Generative Syntax*, 61–92. Cambridge University Press.
Nunes, Jairo. 2013. X′-Structure and Minimalism. In Silvia Luraghi & Claudia Parodi (eds.), *The Bloomsbury Companion to Syntax*, 76–87. London: Bloomsbury.
Perlmutter, David M. 1971. *Deep and Surface Structure Constraints in Syntax*. New York: Holt, Reinhart and Winston, Inc.
Popper, Karl R. 1992 [1934]. *The Logic of Scientific Discovery*. London & New York: Routledge.
Rizzi, Luigi. 1982. *Issues in Italian Syntax*. Dordrecht: Foris Publications.
Roberts, Ian & Anders Holmberg. 2005. On the role of parameters in Universal Grammar: a reply to Newmeyer. In Hans Broekhuis, Norbert Corver, Riny Huybregts, Ursula Kleinhenz & Jan Koster (eds.), *Organizing Grammar*, 538–553. Berlin & New York: Mouton de Gruyter.
Safir, Ken J. 1985. *Syntactic Chains*. Cambridge: Cambridge University Press.
Sankey, Howard & Paul Hoyningen-Huene. 2001. Introduction. In Paul Hoyningen-Huene & Howard Sankey (eds.), *Incommensurability and Related Matters*, vii–xxxiv. Dordrecht, Boston & London: Kluwer Academic Publishers.
Schurz, Gerhard & Paul Weingartner (eds.). 1998. *Koexistenz rivalisierender Paradigmen*. Oplanden & Wiesbaden: Westdeutscher Verlag.
Stegmüller, Wolfgang. 1976. *The Structure and Dynamics of Theories*. Heidelberg/Berlin: Springer.
Tomasello, Michael. 2003. *Constructing a Language. A Usage-Based Theory of Language Acquisition*. Cambridge, Mass. & London: Harvard University Press.

Peter Ludlow

19 The philosophy of generative linguistics: best theory criteria

Abstract: Given the broad range of frameworks for syntactic theorizing within generative linguistics the question sometimes arises as to whether philosophical, or best theory, criteria could help us choose between those frameworks. For example, some frameworks are claimed to be more conceptually simple, others are claimed to be more formally rigorous, and some are claimed to be better grounded empirically or conceptually. Such cross-theoretical considerations need to be handled with care.

Consider the question of whether some theories handle more data or better data or are more "driven by data" and are thus more empirically sound. This quickly leads us to the question of what data is important and what relation it bears to theory choice. I argue that there is no atheoretical notion of best or most important data – whether data is interesting or important depends upon the theoretical questions that we are investigating and what the ultimate goal of our work is. Furthermore, I argue that data only really exists in the context of some theoretical framework, and thus claims of data-driven theorizing and atheoretical or framework-free data gathering needs to be viewed with some skepticism.

In addition, I consider the claims made about simplicity and formal rigor, and then offer a positive proposal for cross-theoretical evaluation: the best theory is the one that enables the theorist to tackle problems of interest with the most efficiency, and I argue further that the choice of best theory is a decision that is best left up to the individual theorist. Cross-theoretical superiority claims involving simplicity and formal rigor are vacuous unless understood as claims about ease of use, or alternatively as tools to facilitate reduction to a more basic science or scientific unification generally.

1 Data

1.1 On the alleged priority of data

Some linguists have argued that the traditional descriptive method in linguistics is preferable because it begins with the data and only then proceeds to theory construction. For example, Absalom and Hajek (1997, 177) claim that "the data needs to drive theoretical analysis and not the other way around." In still other cases linguists have claimed to dispense with the theory-construction portion altogether, arguing that the best way to proceed is to describe a language without any theoretical framework driving our investigation.

https://doi.org/10.1515/9783110540253-019

To illustrate this point, let's consider a view offered in Haspelmath (2010, 293) in his defense of what he calls "framework-free" linguistics.

> Since the advent of the Boasian approach in ethnography and structuralism (both European and American) in linguistics, it has been the goal of descriptivists to approach a language without prejudice and to do justice to its system, regardless of what systems other languages might have. We want to describe each language in its own terms.

The idea advanced by Haspelmath is that investigation should proceed in a bottom-up manner, beginning with theory neutral data collection and only then proceeding to the construction of theory, if at all. As Haspelmath notes, this approach was advocated by a number of Structuralist linguists. We might add that the approach is not unknown in the history of science, as it was advocated by the British philosopher Francis Bacon and later by John Stuart Mill in his *System of Logic* (1843). The problem is that the method, despite its intuitive appeal was quickly rejected by key natural philosophers (notably Pascal, Newton, Darwin, Pasteur, and Mendel).

For example, Darwin offered the following critique of bottom-up approaches to natural philosophy in a letter to Henry Fawcett:

> About 30 years ago there was much talk that geologists ought to observe and not to theorize; and I well remember someone saying that at this rate a man might as well go into a gravel pit and count the pebbles and describe the colors. How odd it is that anyone should not see that all observation must be for or against some view if it is to be of any service! (Darwin and Seward, 1903).

Darwin's point, on my understanding, is that you could describe every pebble in a gravel pit, but that would not be a particularly illuminating enterprise because it wouldn't necessarily lead to a deeper understanding of the world. Furthermore, no description is complete, and the so-called atheoretical observer consciously or unconsciously makes decisions about what to record and/or observe. For example, if one went to Darwin's gravel pit with a notebook and began recording observations, what would one choose to record? You could spend the entire day describing microfeatures of one pebble. And what detail would you go into on describing the color of a pebble? How many variations in hue and tone would you need to record before your effort was complete? Those decisions are hardly atheoretical.

Famously, Darwin spent 10 years studying Barnacles in fine detail, but that was no mere bottom up enterprise. His manuscript on the *Origen of the Species* was sitting in his desk as he studied those barnacles and it provided guidance for him in understanding what properties were important.[1] As Stephen Jay Gould (1990) put it,

> Darwin's view on the need for theory both to suggest and to coordinate observations has been widely acknowledged by scientists as both desirable and inevitable (despite the semiofficial persistence of a public myth about absolutely objective impartiality). This interplay of theory and

1 See Stott (2003) for discussion.

> empirical documentation has both positive and negative implications for the elusive notion of scientific "progress." Theory can prod, suggest, integrate, and direct in fruitful ways; I doubt that Darwin would ever have been able to formulate the theme of natural selection without the available context of Adam Smith's nearly identical causal system for economics (Darwin, in any case, surely did not "see" natural selection in the finches and tortoises of the Galapagos).

None of this is to say that theory can't steer us wrong. It surely can. Theory is a *necessary* condition for scientific success; it is not a *sufficient* condition. Theory may also be impossible to avoid. Sometimes theory masquerades as a purely inductive project.

The first problem is coherently formulating the thesis that one is beginning solely with naked data. This thesis immediately raises a question: What *is* theory-free data? This point is not appreciated in some of the linguistics meta-theoretical discussion, but it is a longstanding view in the philosophy of science. Since the work of Hanson (1958), Sellars (1949), Kuhn (1970) it has been widely held that there is no such thing as raw data or raw sense perception. Everything is viewed, as it were, through the lens of some theory or other, whether that theory be learned, tacitly held and cultural, or a feature of our cognitive architecture.

For example, in the linguistic case simply positing the existence of word boundaries assumes a lot of theory. Morphemes are not things that we nakedly observe; they are the product of scientific theorizing. Indeed, Claude Lévi-Strauss (1953, 350–351) compared the discovery of the morpheme and phoneme to the Newtonian revolution in physics. In other words, descriptions couched in terms of morphological elements, are not simply raw data; they presuppose one of the deepest theoretical discoveries in the history of Western Science. Likewise, even assuming that what one is perceiving is in fact part of human language (as opposed to noises like coughs and sneezes and sniffles) is a heavily theory-laden assumption.

Haspelmath argues that the descriptivist or framework-free approach to linguistics "is not only possible and widely practiced," and is "the best approach to the scientific study of language structure." I am dubious, and Haspelmath's examples of framework-free linguistics are hardly reassuring.

For example, Haspelmath argues that a framework-free approach has been applied to Tagalog.

> Schachter and Otanes (1972, 59–85), still under the influence of American structuralism, describe Tagalog basic sentence structure in its own terms, and the result is a picture that is rather different from what is found in English (with which the authors contrast Tagalog). The basic pattern of Tagalog is not [$_{\text{sentence}}$ NP VP], but [$_{\text{sentence}}$ Predicate Topic]. (Haspelmath 2010, 292)

A specific example offered by Haspelmath is the following, drawn from Schachter and Otanes.

(1) [Gumising]$_{\text{PRED}}$ [ang bata]$_{\text{TOP}}$
awoke TOP child
'The child awoke.'

Haspelmath notes that this syntactic description is borrowed from classical linguistics, but insists that this is fine because it is assigned a different meaning in this context. But even if this description isn't representing the notions of subject and predicate, it hardly matters, since assumed similarity can color observation as much as assumed identity and in this case the structural description is the same – subject-predicate *form* – whether this is really predication or merely something very much like it. One wonders why the notions of sentence, predicate, and topic (or their structural equivalents) are often considered atheoretical descriptors. While they are attested by Plato and Aristotle, surely some genius grammarian discovered them and the fact that today they appear to be theory-neutral descriptors is merely an indicator of how deeply they are sedimented into our observations of and understanding of language.

The allegedly framework-free approach doesn't merely carry theoretical prejudices of its own, but it can also make theorists blind to data as well. Haspelmath unwittingly provides an example.

> Next, what about the position of wh-phrases in a clause and other word order properties of the language (§ 4.4)? Hawkins (2002, § 4.3; 2004, § 7.3) argues that wh-movement creates filler-gap relationships that cause processing difficulty, and that the processing difficulty is greater if the verb (to which most wh-phrases are connected semantically) is further away. (Haspelmath 2010, 301)

There are many concerns here. Hawkins himself is working within a robust theoretical framework with complex hierarchical structures. Stripped of that framework the claim does not stand. Haspelmath enlists the claim to support his case against structural explanations in linguistics, but in this case the anti-theoretical prejudice blinds Haspelmath to oceans of data showing that filler-gap distance has nothing to do with the facts. These are just standard facts that date to Ross (1967).

(2) a. *Who did you hear the story that Bill hit?
    b. Who did you hear that Bill said that John threatened to hit?

(3) a. *Who did you see Bill and?
    b. Who did you say that Bill and Fred saw?

There are a broad range of facts that have been central to a half century of research in generative grammar, surviving various iterations of the theory, so that they have been successively called island constraints, or I-within-I constraints, or complex-NP constraints, or subjacency violations, or bounding node violations. That is, the facts have played a role in linguistics until now, when they are ignored because they conflict with the allegedly non-theoretical goal of doing linguistics in a theory neutral, bottom-up way.

Allegedly atheoretical linguistics can not only make you blind to certain facts, but it can smuggle in robust theoretical assumptions of its own. An example of this is the recent work on Pirahã, by Dan Everett, who reported numerous examples of recursive structures in a (1986) paper. For example, he pointed out that the Pirahã word for *ladder* was originally a noun-noun compound built from the words for foot and handle. The word for bowstring was from a noun-noun compound built from the words for vine and bow.

More interesting though, is example (210a) from Everett (1986):

ti xibıʹib-i-hiab-iig-aʹ kahaıʹ kai-sai

for which Everett gives two glosses.

(i) 'I am not ordering you to make an arrow.' or
(ii) 'I will not let you make an arrow.'

What Everett (1986) argued is that this must involve a sentential clause embedded within a sentential clause. The negation can take scope over the whole thing (yielding Everett's gloss 1) and it can take scope over the inner clause (yielding Everett's gloss 2) but it cannot take scope over just the first clause and not the second (because the second is tucked inside the first).

But as we know, Everett had a kind of conversion experience and in his (2005) paper began insisting that there is no recursion in Pirahã. In the more recent work Everett suggests that this clause is not embedded but is just tacked on as another sentence, which should give us pause, because the utterance does not mean this:

(iii) I'm not ordering you. Make an arrow!

However, Everett did not "unsee" all the data reported in his 1986 paper. His claim is that the linguistic material will ultimately be assembled correctly, but not at the level of syntax. Now, I have issues with this, because I assume that the language faculty is in the business of constructing form-meaning pairs, and I assume that ultimately one of the pairs is going to be a robust logical representation that is legible to the semantics and which thus must include logical connectives. My minor point here is that even if logical operators and recursion are not visible on the surface they are still syntax for all that. But for current purposes the really important point is this: Everett's claim that there is no recursion in Pirahã is not an atheoretical observation, but is a highly theoretical claim because it must make sense of the availability of (i) and (ii) but not (iii) by appeal to assumptions about where the recursive structures are ultimately constructed – not in the core grammar, it seems, but perhaps at a discourse structure level. But that claim is just as theory laden as it can be.

My point here is not to say that Everett is right or that he is wrong on the particulars;[2] my point is that so-called anti-theoretical frameworks are theoretical frameworks too, and that there is no reason to think they are less apt to color our collection

2 For discussion of this question, see Nevins et al. (2009), Sauerland (2015), and Salles (2015).

and interpretation of data. The danger is that suppressing hidden theoretical frameworks, while designed to signal impartiality, may actually work to disguise agendas and/or impediments to productive data gathering. (This also raises the question of whether theoretical linguists ought to farm out so much data gathering to "atheoretical" field linguists.)

Returning to Haspelmath, and his quote from above, it is worth noting that his appeal to "processing difficulty" is puzzling, since you can't simply *see* processing difficulty and it is certainly not an atheoretical concept. Indeed, it only makes sense to talk about it once you have a theory of how language is processed, including, for example whether it is processing language in parallel, how the parser(s) is (are) designed etc.

One gets the impression that inductive approaches to language are not merely carrying water for fossilized grammatical theory, but that there is a high level empirical hypothesis in the mix as well – typically a thesis to the effect that the underlying biological faculty of language is doing minimal work (perhaps it is just generalized cognition) and that the work is being done by a confluence of environmental and historical factors. This is not some theory-neutral position. It is a controversial view in cognitive science that needs to be defended.

Haspelmath, offers what he takes to be a platitude in support of his approach, perhaps not recognizing that he is offering a theoretical defense of his implicit background theory:

> The narrow range of actually existing organisms is primarily determined by survival (i. e. the chance of successful replication), not by constraints on what the genetic code allows. To study the nature of the cognitive code, we should study the acquisition of unattested language types under natural or artificial conditions, but we should not hope to derive much insight from constraints on attested languages. (Haspelmath 2010, 301)

My principle point here is that this approach to language is nothing if not deeply theoretical, and it could not help but drive his empirical work. But beyond that, Haspelmath's claim about evolution is no truism. It is certainly not how I understand evolutionary theory: Evolutionary possibilities are radically constrained by biological, physical, and mathematical constraints on possibly evolutionary outcomes.

Here is one example that has influenced Chomsky's recent work in the Minimalist Program. Near the end of his life, Alan Turing began exploring the physical constraints on biological phenotypes, and argued that, for example, the spot and stripe patterns on animals are not wide open from an evolutionary perspective, but are constrained by low level biophysical principles. Turing's (1952) speculative thesis was that much of the range of patterns on animals could be a function of the interplay in a so-called "reaction-diffusion system." Such a system consists of an "activator," that can make more of itself and an "inhibitor," that slows production of the activator; and a mechanism for diffusing the cells (or chemicals). The system allows a broad range of patterns, but it also radically delimits the class of possible patterns to what are now called "Turing Patterns."

The proposal was nothing if not driven top-down, and it was obviously based on Turing's limited observation of animal patterns, and to this day it has not been possible to empirically test the hypothesis except in simpler animals, but it remains a healthy working hypothesis nonetheless, and there are more speculative versions of the hypothesis in which it is extended to lung, limb, and tooth development.

But this story from biology just recapitulates stories about the success of science in general. Scientists don't just gather data and seek to generalize from that data. They propose bold theories with the goal of deepening our understanding of not only how things are, but why they are that way. They seek to explore the deep underlying explanation for what is going on.

Even today, with the advent of big data, it is just false to claim that the work being done is bottom up inductive data gathering. The biologist Craig Ventor (2008), who is often considered a primary advocate for bottom-up big data science is explicit about this.

> Despite my urging that we always look for those big questions, data generation for its own sake continues to be a major impediment to real scientific breakthroughs in genomics. It is not hard to understand why investigators, particularly young scientists, are satisfied being data generators, as government agencies and some foundations continue to pay out hundreds of millions of dollars for just DNA sequencing or, even worse, microarrays, creating huge datasets but seldom any real scientific insight. (Ventor 2008, 2)

The minute one begins gathering data one is already fully immersed in theory – and it may well obstruct more interesting and valuable theoretical work. This is why claims that research is being driven by the data are so pernicious.

The claims are pernicious because by claiming their proposals are "frame-work" free they are falsely claiming a position of neutrality, and more importantly, they are giving their proposals an aura of unchallengability. (If a proposal is framework-free, then of course any challenge would have to be theoretically prejudiced.) Beyond that, the approach is a way to smuggle in theoretical commitments without making them explicit or even acknowledging they are there, much less subjecting those theoretical commitments to critical scrutiny.

But there is more to it than that, for the very methodological insistence that one proceed in a bottom-up way can be theoretically stifling. We often forget that even Chomsky, during his tenure in the Harvard Society of Fellows attempted to utilize Structuralist "discovery procedures". Not successfully, as he reported.

> By 1953, I came to the same conclusion [as Morris Halle]: if the discovery procedures did not work, it was not because I had failed to formulate them correctly, but because the entire approach was wrong … [S]everal years of intense effort devoted to improving discovery procedures had come to naught, while work I had been doing during the same period on generative grammars and explanatory theory, in almost complete isolation, seemed to be consistently yielding interesting results. (Chomsky 1975, 131)

Earlier in this section I mentioned Darwin's monumental work on barnacles and how during that entire period of research he kept his theory of evolution a secret (almost costing him credit for the discovery). A standard view is that he kept the theory of evolution under wraps because he feared criticism from religious authorities. But Francisco Ayala (2009) has suggested a different reason.

On Ayala's view Darwin felt compelled to keep his theory under wraps because he was concerned about the reaction of the framework-free naturalists, who were under the theoretical spell of Bacon and Mill. As a consequence, while his personal writing (as in the gravel pit example above) showed him working in a top-down manner, in his public writing he claimed he was proceeding inductively from the bottom-up. Not only did he carry on this ruse himself, but in 1863 he wrote to a young naturalist, advising him to do likewise.

> I would suggest to you the advantage, at present, of being very sparing in introducing theory in your papers (I formerly erred much in Geology in that way); let theory guide your observations, but till your reputation is well established, be sparing of publishing theory. It makes persons doubt your observations. (Darwin and Seward, 1903, 323)

The moral is that purely inductive approaches to empirical investigation, whether they label themselves as framework-free or something else can be just as hegemonic as the most dogmatically advocated theory. Or religion, for that matter.

1.2 Popper

Another issue that comes to the fore is whether, once we settle on what counts as data (within a theory), failures to account for all the data should become disqualifying for a theory. For example, Loporcaro (1989, 343) contends that "although descriptive simplicity is a desirable goal, it should not be attained at the cost of contradicting actual linguistic data." This certainly seems to be a widely shared view, and when the view is shared the name of Karl Popper is often intoned.

The appeal to Popper is puzzling, because his position is certainly a marginal one in the philosophy of science today, and it has never been faithful to actual scientific practice. The general consensus (in theory and in practice) is that you don't torch your ship when it springs a leak. You patch the leak and wait for a better ship to come along. And of course, even Popper came to agree that you don't junk a theoretical paradigm because of a counterexample. Thus Popper (1985) pointed out that we did not reject the theory of planetary motion because of failed predictions in the motions of the planets. We posited the existence of a hypothetical body that was causing the perturbations. And that amended hypothesis not only accounted for the facts, but it actually led to the discovery of Neptune.

But what is most puzzling about these claims about falsification is that they belie a deep misunderstanding of Popper's (1959) point. His point was not in any way shape or form a data-first approach to theorizing. Indeed, he argued that the inductive method would get you nowhere. His point about falsifiability was a call to develop theories in a top-down manner for which there are bold empirical predictions.

More to the point, the kind of science that results from a "framework-free" approach to science is precisely the sort of thing that Popper was warning against. It offers no bold empirical hypotheses – indeed it is effectively banned from doing so. Further (since it is actually itself a theory in disguise) it is the worst possible kind of scientific theory – a gerrymandered and unprincipled theory that is crafted specifically to comport with an arbitrary class of data. That suggests a kind of vacuity of theorizing that Popper was warning against.

1.3 Data coverage

A related concern is that it is far from clear what data a theory is actually responsible for. Sometimes a class of data is assumed to be a target for a theory but this assumption turns out to be false. A concrete example of this idea is the discovery of the viral cause of Polio. At the outset, 'polio' was a covering term for a broad range of symptoms and conditions for which we had no clear explanation. The subsequent viral theory of Polio did not cover all the phenomena that we took to be symptomatic of Polio. As reported in Assad and Cockburn (1972), Sabin (1981), and Gear (1981), we now know that there are dozens of other non-polio enteroviruses (NPEVs) that are capable of causing paralytic polio-like syndromes. It is generally agreed among medical researchers that many of the cases that fell under the rubric of 'Polio' originally were in fact caused by NPEVs.

In response to the apparently narrow data coverage of the theory of Polio, we *could* have said that the viral theory of polio was a failure because it did not "cover the data," but what scientists actually said was that the viral theory accounts for Polio and the other cases were conditions that we mistook for Polio.

To apply this point to a recent case in linguistics, the Minimalist Program has taken criticism for shedding responsibility for the explanation of a number of empirical phenomena – some of which have featured prominently as targets of explanation in linguistic theorizing for the past half century. But given the Minimalist Program's goal of achieving a reduction to (or integration with) low level biophysical principles, this is the sort of outcome one would expect. In this vein then, one can think of the shedding of responsibility for explaining classes of data to be preparatory work for the reductive project. Whether the reductive project works of course remains to be seen. My point here is that the loss of apparent explanations for empirical phenomena does not by itself suggest any limitation in the Minimalist Program or any other framework proposed in this volume. We simply can't know in advance what data a theory is responsible for. Of course it follows that perhaps Haspelmath is correct in ignoring data concerning island constraints; my point here is that we should recognize that move as consequence of theory-choice, to be defended on theoretical grounds, and we should not pretend it is an observation that extrudes from fidelity to raw data – assuming there could be such a thing. Limiting the scope of data relevant to Polio only made sense in the context of an explanatory theory of the underlying causes of the condition.

If it is difficult to judge the health of an empirical theory based on assumptions about data coverage, is there another way? Could we appeal to best theory criteria like simplicity and formal rigor?

2 Simplicity

Questions about simplicity have been part of generative linguistics since the very beginning; they played a large role in the so-called "generative semantics wars,"[3] and they have made a comeback in part due to discussion (some of it informal) surrounding the advantages of Chomsky's (1995, 2000, 2001a, 2001b) "Minimalist Program."

The problem is that there has been lots of confusion about what is meant by 'simplicity' when applied to linguistics. The problem is not entirely linguistics-internal, however, as a number of uses of 'simplicity' have been deployed in the sciences generally, and often without clarity.

For example, what does the term 'simplicity' apply to? As Barrios (2016) asks, does it apply to our vehicles of theorizing (formal methods and conceptual elements of those theories) or does it apply to that about which we theorize? Or does it apply to the way in which our theory fits that about which we theorize? If it is about our theories then what aspect of our theories is it about? For example, is it about the *conceptual* complexity of our theories or is it about the *formal* complexity of our theories? Alternatively, sometimes when linguists talk about simplicity they are talking about the simplicity of acquiring a grammar, in which case they are not talking about simplicity of the theory, but rather the simplicity of the task for the language learner.

Even when we talk about the simplicity of the vehicle of theorizing there are a variety of notions of simplicity in play. In Ludlow (2011) I distinguished between two versions of simplicity, based on the work of Charles Sander Peirce and Elliot Sober – the former notion I called P-simplicity (P for Peirce) and the latter I called S-Simplicity (S for Sober). P-simplicity applied to the mental cost of theorizing, and is the idea that simplicity is best understood as a problem of mental economy for the theorist – the theory that is easier to use is the simpler. This leaves open certain questions – for example are we talking about the economy of our calculations or our ability to grasp theoretical features of the theory. S-simplicity involves the preparation of our theory for reduction to more basic science.

3 Sometimes also known as the "linguistics wars". See Newmeyer (1986) and Harris (1993) for discussion.

Other notions of simplicity are certainly possible. Barrios suggests that we need a notion that we might call "unification simplicity" or U-simplicity (my terminology, not his) – a notion of simplicity that pertains to whether the proposal successfully unifies various other phenomena. A classic example of this would be the theory of gravity, which unified the motion of the planets, the tides, the motion of pendulums, etc.

In Ludlow (2011) I argued that regarding the formalization of linguistic theories, there are cases in which it makes sense to talk about simplicity and those in which the notion needs to be handled with care. In particular, within a particular linguistic theory it is possible to formally define simplicity in a way that distinguishes between rival proposals. But I argued that this sort of move is not possible for theory-external judgments of simplicity – for example between two rival theories. To illustrate this point, I drew on an argument made by Paul Postal in his (1972) paper "The Best Theory". Some review of that discussion might help illuminate this issue.

Postal (1972, 137–138) put the idea this way: "[w]ith everything held constant, one must always pick as the preferable theory that proposal which is most restricted conceptually and most constrained in the theoretical machinery it offers." That may sound sensible, but closer scrutiny raises some questions. The first issue concerns what counts as "less theoretical machinery." As I said earlier, within a particular linguistic theory it is possible to stipulate what counts as more or less machinery. For example, Chomsky and Halle (1968) define 'simplicity' in terms of the number of symbols contained in a phonological rule. But what happens when we consider theoretical simplicity across frameworks? What counts as more theoretical machinery? And for that matter, what counts as machinery?

One problem is that the use of 'machinery' is metaphorical and otherwise undefined cross-theoretically. To put the problem in terms of resources drawn on in generative linguistics we can ask the following questions: Are transformations machinery? Are levels of representation machinery? Are constraints on movement machinery? Are principles machinery? Are parameters? Are operations like merge? Legibility constraints? Features? Functions? Type-shifting? Mapping algorithms? If these elements count as machinery, when do they count as *more* machinery? For example, does one big complex and opaque algorithm count as less machinery than a handful of rules or does it count as more?

Hempel (1966, 42) noted that this question arises even in the context of the nonmetaphorical notion of axiom within geometry, illustrating the problem with the example, "for any two points there is a straight line containing them." Is this one axiom or more? As Hempel noted, you can think of it as a conjunction of two axioms – that there is at least one such line, and that there is at most one such line. How does one go about counting these kinds of formal elements? Again, there is no problem with establishing the criterion *within* a particular theory, but how do we establish the counting metric for *cross*-theoretical comparisons?

Even if you had an absolute way of counting axioms there are still issues to consider *vis-a-vis* simplicity. *Prima facie*, we might suppose that the simpler geometrical theory is the theory with the fewest axioms, but caution is necessary here. For example, a geometry with fewer axioms might end up with more complicated theorem proving (at least initially). Is that still a gain in simplicity?

Even if difficulty in theorem proving is not an issue, is a system with fewer axioms guaranteed to be the simplest? Not according to Sober (1975). On his view, it depends on the "naturalness" of the axioms (or basic elements of the theory), here taking "naturalness" to be a function of perspicuous theoretical reduction. If we trade in five natural axioms for four unintuitive axioms then according to Sober we have achieved no great gain in simplicity (S-simplicity). So, by some accounts, even if we have a definition of machinery and a way of quantifying the amount of machinery, we do not have a simpler theory simply by having fewer theoretical elements.

We can make the general problem more vivid by using an example that Postal offered. His basic idea was that if one could eliminate a level of representation (in this case the Deep Structure of Chomsky, 1965) and still cover the same data, one would have a *prima facie* reason to accept the "simpler" theory. Specifically, Postal held that the proposal in Katz and Fodor (1963) offered exactly this advantage. So, for example, the model of grammar represented in (4), might be replaced with the allegedly simpler (5).

(4) semantic representations

(projection rules)

deep structure representations

(transformations)

surface structure

(5) semantic representations

(transformations)

surface structure

Postal (1972) put the case as follows:

> What I wish to suggest briefly is that because of its *a priori* logical and conceptual properties, this theory of grammar […] is the basic one which generative linguists should operate from as an investigatory framework, and it should be abandoned, if at all, only under the strongest pressures of empirical disconfirmation. (Postal, 1972, 135)

While it appears to be true that a level of representation is eliminated by replacing (4) with (5), there remain questions about the nature of the transformations mapping between levels. Do they become simpler as well or must they become more complex? It is hard to see what is gained if a so-called level of representation is eliminated at the cost of more complex mapping rules.

But another consideration is the perspicuity of the representations for the theorist. Theories are designed in part to aid theorists. Is there some advantage to the theorist in eliminating a level of representation? Perhaps for some theorists, but it is hard to imagine that this could be true for all theorists.

If we are working with the notion of P-simplicity discussed above, simplicity is determined by a community of theorists with some shared technical background and abilities and a shared set of research goals regarding some reductive project. To such a community, given two proposals with roughly the same empirical coverage and explanatory power, the simplest theory is that theory which they find the easiest to use for constructing and evaluating hypotheses (the one that is P-simple). Alternatively, however, if we are talking about S-simplicity, then we are looking for the theory (or version of the theory) that more naturally affords integration with the target reducing science (the one that is S-simple). If we are concerned with U-simplicity, we are interested in whether the theory (or version of the theory) enables a more natural unification with other sciences.

Barrios (2016) makes the excellent point that these notions of simplicity do not merely enter into the picture when we encounter a “tie” with respect to data coverage. They play an important role in the construction of the theory. S-simplicity and U-simplicity guide our research even when there are no competing theories.

There are questions about how independent these notions of simplicity are, however. For example, you could think of all versions of S-simplicity as being a special case of U-simplicity – a case of unification that involves integration of a science with a more basic science. For that matter, you could think of S-simplicity as being a special case of P-simplicity, since a successful reduction might be understood as one in which the components of the theories dovetail in a way that is perspicuous to theorists. And of course, the same might be said of U-simplicity.

This then raises the question of whether S-simplicity and U-simplicity are actually proper notions of simplicity, or whether they package together other theoretical desiderata with a notion of P-simplicity. For example, we presumably prize unification in science whether or not it yields a simpler theory. The same can be said for a successful reduction of one science to another.

Barrios discusses a number of examples in linguistics in which simplicity considerations have played a role, most notably in certain moves within the Government-Binding framework of Chomsky (1980) and more recently the minimalist framework of Chomsky (1995) and subsequent work. I’m inclined to think that the appeals to simplicity in both of these frameworks are best understood as appeals to S-simplicity, or at the very least to U-simplicity.

In the Government-Binding framework, and for much of the history of generative linguistics, the task has been to construct a theory that unifies our explanation for the variety of grammars we witness. Although controversial (see the discussion of Haspelmath above) the approach is not driven by an abstract notion of simplicity, but rather by the hypothesis that there is a faculty of language common to all humans and thus that any attempt to probe the nature of that faculty will involve understanding what it is that different human grammars have in common – or rather, what underlying mechanism they are grounded in. The simplicity desiderata, if it is that, is better understood as a preparation of the theory of grammar for a perspicuous reduction to (or unification with) more basic sciences – in this case cognitive psychology.

Similarly, in the minimalist program, the concerns for economy and simplicity are driven in large measure by the working hypothesis that our language faculty came to be by virtue of certain low-level bio-physical and mathematical principles. The simplicity desiderata, if it is that, is in this case better understood as a preparation of the theory for a perspicuous reduction to or unification with the science of those low-level bio-physical and mathematical properties.

Notice that in each case I used the term 'perspicuous' to describe the reduction/unification, and I believe that here is where our notion of P-simplicity enters the game once again. Of course, P-simplicity, when it is a factor, needs to be understood as being in some sense "forward looking". We aren't interested in the perspicuity of the theory today, as we are in the perspicuity of the theory over the long haul. It would be foolish to evaluate one theory as being P-simpler simply because it is more familiar in the moment. Lindsay (1937, 166) put the point this way: "If, for example, a person familiar with classical mechanics can become equally well acquainted with another physical theory in a time of the same order of magnitude as that which he took to learn mechanics, he should consider this new theory as simple as mechanics, no matter how complicated it may seem at first examination to one unfamiliar with it."

The quote from Lindsay suggests that some things may seem complex to us at first, but reveal themselves as being P-simple after a reasonable period of time. This can happen for the field as a whole. So, theories that may seem complex at the outset, may turn out to be viewed as utterly P-simple as the field progresses. As Lindsay (167) further notes, "we shall ultimately consider [theories] simple when we have grown sufficiently familiar with them to forget that they ever seemed difficult to understand."

This leads to the question of when it is ok to say "no" to a new framework because it is excessively complicated and when it is right to make the jump. This is a question that is integrated with the issue of formal rigor as well, so I will take it up in Section 4, after we address the topic of formal rigor.

3 Formal rigor

In the 1980s it was sometimes argued that Government-Binding (GB) theory was defective because it was not presented in a fixed and rigorously worked out formalism

(This was before formalizations by Stabler (1992) and Johnson (1989, 1991)). For example, Gazdar et al. (1985) argued that the formal rigor of Generalized Phrase Structure Grammar (GPSG) weighed in its favor. Similarly, Bresnan and Kaplan (1982) argued that Lexical Functional Grammar (LFG) was a better framework for theorizing because it is more formally rigorous.

I contend that whether or not such claims have merit depends on our interests as theorists. For example, if the goal is the development of natural language processing systems (or linguistic theories that inform such systems) then a certain kind of formal rigor can be very important. Thus, I have no quarrel with the above-sited authors if their theoretical goals are the construction of theories amendable to natural language processing. Formal rigor is not, however, in and of itself, a desideratum for linguistic theorizing.

In Ludlow (2011, Section 7.2) I suggested that there are actually several different versions of formal rigor being used in the literature and in informal discussions of the topic. We might, for example, think of formal rigor as recursive specifiability. This notion of formal rigor is *among* those that are advocated by Gazdar, Klein, Pullum, and Sag.

> There may not even be algorithmic ways of confirming the consequences of some theories of grammar, of course: if the theory allows grammars for non-recursive sets, then we run the risk that the claim that some string is not generated by some grammar cannot be verified in principle. Familiar statements of the type 'Thus our grammar excludes examples like (158)', in other words, may simply be untestable conjectures. (Gazdar et al., 1985, 2)

Similarly, Bresnan and Kaplan argued that it is important that the set of well-formed sentences generated by the theory be recursive.

> The reliability constraint implies that the subset of data in the domain of the mapping for which there are well-formed grammatical relations is a recursive set (for the mapping must effectively compute whether an arbitrary string is grammatically well formed or not). (Bresnan and Kaplan, 1982, xl)

Pullum (1989) also clearly advocated this view of formal rigor, proposing the following three conditions on formal theories of grammar and suggesting that they were "non-negotiable".

> (I) The notion of 'structural representation' must be effective. That is, there must be an algorithm for determining whether some arbitrary string, graph, or diagram counts as a structural representation according to the theory.
> (II) The notion of 'rule' (or 'principle' or 'law' or 'condition' or 'constraint' or 'filter' or whatever) must be effective. That is, there must be an algorithm for determining whether some arbitrary string, graph, or diagram is a rule (or 'principle' or 'law'...) according to the theory.
> (III) The notion 'generates' (or 'admits' or 'licenses' or whatever) must be effective. That is, there must be an algorithm for determining whether some arbitrary structural representation is generated (or admitted or licensed...) by a given set of rules (or 'principles' or 'laws'...). (Pullum, 1989, 138)

Let's call this notion of rigor *EP-rigor* (EP for effective procedures, in the technical sense of 'effective'). Whatever the merits of formal rigor in linguistic theorizing (and there can be many), EP-rigor is too strong a demand for any theory that I, at least, can imagine. As Church (1936) proved, there isn't even an effective procedure for deciding if an arbitrary formula of first order logic is a theorem. Do we really expect good theories to be more "rigorous" than our procedures for determining the theorems of first order logic?

More generally, and apart from logic, no science to date has suggested that the consequences of a theory could be effectively determined. While there was an attempt to axiomatize biology – Woodger (1937, 1939) – even Woodger did not make a demand for effective procedures.

One might argue that while physics and other basic sciences do not construct theories with recursive sets of predictions, there is no reason why linguistics shouldn't endorse EP-rigor. Suppose, for example, we thought of the mind as a Turing Machine and thus expected any theory that is a chapter of cognitive science to be Turing computable.

One problem with this line of thinking is that even if we held such a view of the mind, it would only put constraints on a final version of linguistic theory (assuming the idea of a completed science even makes sense) – it wouldn't put constraints on the theory while it is being developed. There is also a question of whether the Turing Machine theory of the mind would put such constraints on the finished linguistic theory. Chomsky (1980, 127) asked that we imagine a language is technically non-recursive, but in which there is a Turing Machine that has a decision procedure for recognizing well-formed sentences that is effective for sentences with lengths of (let's say) less than one million words, but is not effective (in the technical sense of 'effective') for sentences of greater length. Such languages would be entirely compatible with the thesis that the mind is a Turing Machine – at least it would satisfy the original motivations for such a theory without being EP-rigorous.

Alternatively, one might weaken the demands for rigor here and propose that the idea is not to have *effective* procedures for generating the theorems of the theory, but merely *mechanical* procedures for doing so. Here a mechanical procedure would be one in which it is possible to automate the generation of theorems of the theory. Let's call this view *MP-rigor* (MP for mechanical procedure). The virtue of MP-rigor would be that one could use the procedure to generate theorems that might otherwise escape the attention of investigators, thus providing more opportunities for testing and possibly providing evidence against the theory (cf the discussion of Popper above). In addition, the availability of such a mechanical procedure would potentially sharpen claims about what counted as a prediction of the theory.

There is surely merit to the claim that MP-rigor is hypothetically useful for linguists under certain circumstances. It is certainly valuable to theories that are intended to be deployed in natural language processing. You might think that MP-rigor would also be a valuable tool for exploring the consequences of a theory. It is not obvious, however, that MP-rigor should be a global desideratum for theorizing. Mechanical procedures are certainly useful when it is difficult to see what the consequences of a theory might be, or when counter- evidence is hard to come by. But arguably this is not the state of affairs in current linguistic theory – or certainly not in all branches of it. Counterexamples to existing theories are not in short supply. In the current state of the field it is a challenge to come up with theories that fit the data in hand. When theories *do* cover the existing data (or at least the salient data), or even some fragment of it, counterexamples are quickly discovered.

Another way to think of formal rigor is that it is a pre-established logical/mathematical space in which linguistic proposals can be evaluated. Something like this idea seems to be supported by Gazdar et al.

> This approach to language is characterized by its goal of investigating natural language through the construction of fully explicit descriptions of particular languages and a formalized general framework for defining the space within which to locate such descriptions. (Gazdar et al., 1985, 1)

Based on the following passage, I believe that Kaplan and Bresnan endorse this idea as well.

> [...] this formalism, called *lexical-functional grammar* (LFG), has been developed to serve as a medium for expressing and explaining important generalizations about the syntax of human languages and thus to serve as a vehicle for independent linguistic research. (Bresnan and Kaplan, 1982: 173–174)

Let's call this notion of rigor PreF-rigor (where the PreF indicates the desideratum of a formal *pre-established framework* in which to theorize). In Ludlow (2011, Section 7.2) I was critical of this approach to theorizing, and I remain critical if it is offered as a demand for proper theorizing (but if some researchers prefer to work within an established formal framework there are worse strategies for theorizing). It becomes a problem when we make this a *global* demand or desideratum for theorizing and/or an argument of the superiority of one framework over another.

The problem with thinking of formal rigor in terms of a pre-established logical/mathematical space or language in which to conduct enquiry is simply that this is typically not how successful science proceeds. There is a myth that one often hears, for example, that the Calculus provided such a formal framework for the development of modern physics. But the Calculus was not fully formalized until classical physics had been eclipsed by more contemporary theories, and for the most part the development of the Calculus followed in the wake of developments in physical theory. By some accounts (e. g. Kitcher, 1981), attempts to rigorize the calculus in advance led to the stagnation of mathematics in Great Britain, even while it flourished on the continent.

For example, after Newton published his *Quadrature of Curves*, George Berkeley argued that the calculus involved the logical fallacy of shift in hypothesis (Newton had assumed in one argument both that *o* is nonvanishing and that it is zero). As noted by

Eves and Newsom (1965), the problems with the logical integrity of the calculus were overlooked because applications of the calculus were successful.

> It was natural that this wide and amazing applicability of [the calculus] should attract mathematical researchers of the day, and that papers should be turned out in great profusion with seemingly little concern regarding the very unsatisfactory foundations of the subject. It was much more exciting to apply the marvelous new tool than to examine its logical soundness, for, after all, the processes employed justified themselves to the researchers in view of the fact that they worked.
> Although for almost a hundred years after the invention of the calculus by Newton and Leibniz little serious work was done to strengthen logically the underpinning of the rapidly growing superstructure of the calculus, it must not be supposed that there was no criticism of the existing weak base. Long controversies were carried on by some mathematicians, and even the two founders themselves were unsatisfied with their accounts of the fundamental concepts of the subject. (Eves and Newsom, 1965, 198)

Kitcher (1981, 486–7) noted that Leibniz and his followers, by postponing "rigorization," were able to make greater advances than the Newtonians—who were hung up on making the calculus rigorous before moving on. Leibniz, on the other hand, held that the new theory was so revolutionary that the usual techniques for making the theory rigorous would be unsuccessful. Again, from Kitcher:

> Leibniz' successors produced so many apparently successful reasonings, which could not be readily reinterpreted using the methods favored by the eighteenth century Newtonians, that, by 1750, the Newtonian claim that all the unrigorous reasonings of the calculus could be reconstructed according to Newtonian proposals was no longer defensible. Similarly, the achievements of the Bernoulli, Euler, and other continental mathematicians, undercut the thesis that the question of rigorizing the calculus was an urgent one. (Kitcher, 1981, 487)

The moral I draw from this is that if one has a theory that is rigorous enough to conduct inquiry, and yet one devotes energies to rigorization under these circumstances, then it comes at a cost – the cost being successful application of theory to open problems. Should we insist, for example, that generative linguists provide PreF-rigorous foundations for their new theories before they present them to us? Would this be the most productive use of time for theoretical linguists? Would it help in application of the framework to the class of problems currently of interest to linguists?

The final version of formal rigor – and the one that I consider to be globally viable – is simply a desideratum that a statement of the theory (or theoretical framework) be clear enough to be understood, evaluated, and debated. Let's call this CE-rigor (CE for clear enough).

The first thing to understand is that CE-rigor does not require the use of a mathematical language. Mathematical languages need not be CE-rigorous, and CE-rigorous theories need not be mathematical. Further, what counts as CE-rigor depends on needs; it is context dependent. Indeed, we might say that the most helpful version of rigor is "rigorous enough for our current purposes".

The history of the development of the calculus can again be informative here. For example, Kitcher (1981) asks the question "why was the problem of rigorizing the use of infinite series urgent for Cauchy (and Abel) but not for Euler, and why did Cauchy adopt and Euler reject the same proposed rigorization?" Or, as we might put it, why was Euler's solution not CE-rigorous in the eyes of Cauchy? The answer, according to Kitcher, was that Euler and Cauchy were interested in different questions. Euler was interested in questions that concerned computing the sums of an infinite series. For Cauchy, on the other hand, important questions in the theory of real numbers could not be resolved until the infinite series techniques were *more* rigorous (see Kitcher (1981, 489)). As we might put this, given Cauchy's interests, the work of Euler was not CE-rigorous.

The moral of this for the linguist is that what counts as CE-rigor will change as the interests of linguistics research changes. Different linguistic research programs will have different interests and therefore different thresholds for CE-rigorization. One might also expect different *components* of a single theory within generative linguistics to have different thresholds for CE-rigorization at different times.

As noted, rigorization comes at a cost. Rigorization requires cognitive labor. Thus, our discussion of rigor converges with that of the notion of P-simplicity introduced in Section 2. Does the task of rigorization collide with the demands of P-simplicity? The answer, as you may have guessed, is that this depends. Rigorization might lead to a cognitive cost on the front end, but to a gain in simplicity or explanatory adequacy on the back end (cf our discussion of Lindsay above). This leads to an important question: When should we buy into a claim that more rigor is demanded (that we haven't achieved our CE-threshold)? When is it worth the cost? When do we make the switch from one theoretical framework to another?

4 Minimal effort and optimal switching points

In Sections 2 and 3 I looked at two features of best theory criteria – simplicity and formal rigor. The approaches I suggested for these criteria had a kind of unifying theme: make our lives as investigators as easy as possible, given our goals, interests, and abilities. Or to put it another way: How can we achieve our goals most fruitfully with the least expenditure of intellectual labor.

In the case of formal rigor, we want to deploy as much rigor as is necessary to resolve the questions that we are interested in (CE-rigor). In the case of P-simplicity, we want the theory to be P-simple because P-simple theories are just theories that make it easier for us to conduct the business of linguistics. In the case of S-simplicity we want our theory to line up naturally (and cost-effectively) with our goals for reduction to basic science. In the case of U-simplicity we want our theories to dovetail in a way that is perspicuous to investigators. Thus, even the goals of S-simplicity and U-simplicity are integrated with the quest for P-simplicity.

If I am right, then the choices linguists make in theory construction should not be driven by assumptions about some external, aesthetic notion of simplicity or rigor for rigor's sake, but rather should be driven by whatever makes it easier to construct theories that satisfy our explanatory, reductive and unificatory goals.

If we were really interested in investigating best theory criteria, then the idea would be to find criteria that best measured and predicted the amount of cognitive labor that would go into opposing theoretical approaches, given our reductive and other goals.

In the case of engineering problems (for example designing a new aircraft) these questions can actually be quantified, and indeed they should be quantified given the costs (and potential cost savings) involved. In linguistics, however, costs are not always that easy to quantify. However, if we are serious about cross-theoretical comparisons, we may want to think in terms of the utilizing our intellectual energy for optimum effect (again given reductive and other goals).

For the moment let's just focus on the goal of understanding the nature of the language faculty. Even once we fix this goal, we still have a range of conflicting methodologies, and through the years we have seen numerous theoretical frameworks – even within Chomsky's own project, from the Standard Theory of Chomsky (1965) to the Extended Standard Theory, to Government-Binding Theory, the Principles and Parameters Framework and the Minimalist Program. Each of these shifts in framework was triggered by the perception of investigators that a more efficient and effective theoretical framework was now available (more efficient in the sense of getting closer to the goal of explaining the language faculty in the least expensive way possible).

Shifts of this sort involve a significant cognitive burden – researchers who want to participate in the new framework need to retool. To illustrate, some linguists that were not near MIT during the 1970s–1990s may have balked at the cognitive burden of adopting the successive new frameworks. Some linguists trained up on old methodologies may not have seen the payoff down the road. Alternatively, others took on new research goals (for example developing linguistic theory so that it integrates better with natural language processing) which offered cognitive burdens of their own. There was, for sure, a steep learning curve with each theoretical shift. Were the shifts worth the cognitive labor? Not to everyone. Some found the existing paradigm fruitful enough so that a shift was not worth the bother. Others found that the results they prized were no longer the target of the new research paradigm.

The interesting thing about an academic field like linguistics is that there is no corporate research officer to decide when the optimal time to switch methods is. Of course, there are funding agencies that dictate research directions to some extent, and they do partially control the flow of money in the profession, but their control of research trajectories is limited. For the most part linguists are driven by their own abilities and interests. Some are gifted in technical methods; others are gifted in observing surprising new facts. Still others can see how changes to the theory can result in more explanatory approaches to the nature of language. My point in this chapter is that we should see this for what it is – each of us works with the methods that make our work as theorists easier, given our goals. It really accomplishes nothing to proclaim that an invented notion of simplicity or formal rigor is somehow necessary to do linguistics properly. What does make sense is for us to discover new ways to more efficiently tackle the problems that interest us, and convince those who share our interests that the new methods will help them too – that the learning curve will be worth it.

Of course, some research paradigms will be successful and others will not. This does not entail that the failed or unsuccessful projects should not have been undertaken. To the contrary, from the perspective of the research community, the diversity of approaches is preferable. You don't want everyone in the field pursuing a promising new technical research program, it is better to have technically able "scouts" to examine the results of the framework. The rest of the field (or much of it) may want to follow these scouts later.

Linguistics, like many other fields, has leading figures that can drive research projects that might otherwise not be undertaken. Chomsky's work in developing the Minimalist Program comes to mind here. At least in the oral tradition there have been objections to the ability of figures like Chomsky to drive research in this way. However, while one wouldn't want the entire field to pursue a high-risk path, it is valuable to the field that such research programs be undertaken from time to time, and it is difficult to see how they could be robustly undertaken without the example and direction of leading figures. Even if such research programs failed more often than not, the fact that they sometimes strike gold, or at least spin off valuable alternative research programs should not be discounted.

The point is that from the philosophy of linguistics perspective, the ecology of the research field is much healthier if there is a diversity of research projects and methods. When a research project fails or is in any case exhausted, the right response is not that the project should not have been undertaken. The field as a whole needs a broad range of methods and research goals, and this in turn demands that failed programs be accepted as part of the bargain.[4]

5 Data redux: linguistic judgments

In Section 1, I discussed the role of data in linguistic theory, but made no assumptions about the nature or source of the data; we were primarily interested in the relationship between data gathering and theorizing and whether the idea of data-before-theorizing

4 Pius ten Hacken (pc) has suggested that the same field might accommodate several Kuhnian paradigms simultaneously. In his view, and I agree, this only becomes problematic when theorists become so involved in cross-paradigm warfare that their theory-internal puzzle-solving stalls out.

made any sense. However, one source of linguistic data – linguistic intuitions/judgments – has proved to be contentious, and I turn to the topic here, armed with considerations about our efficiency as investigators. Efficiency considerations can help us to see the value of linguistic judgments as tools for efficiently making scientific progress. First, some background about linguistic judgments is necessary.

A number of philosophers and linguists interpret the practice of using linguistic intuitions as involving some inner "voice of competence." Devitt is one example: "I need a word for such special access to facts. I shall call it 'Cartesian"' (Devitt, 2006, 96).

In Ludlow (2011, Section 3.2) I argued it is doubtful that linguists actually adhere to such a view regarding intuitions. I also argued they *shouldn't* adhere to such a view. I am going to take a leaf from Williamson (2004), talking about so-called intuitions in other areas, and endorse his view that we ought to abandon this talk of 'introspection' and 'intuition' altogether:

> "What are called 'intuitions' [...] are just applications of our ordinary capacities for judgement. We think of them as intuitions when a special kind of scepticism about those capacities is salient. Like scepticism about perception, scepticism about judgement pressures us into conceiving our evidence as facts about our internal psychological states: here, facts about our conscious inclinations to make judgements about some topic rather than facts about the topic itself. But the pressure should be resisted, for it rests on bad epistemology: specifically, on an impossible ideal of unproblematically identifiable evidence." (Williamson, 2004, 109)

Let's follow this advice in the case of linguistics and think of so-called linguistic intuitions as being judgments, not so very different from the judgments that might be used in judging whether the meniscus of a liquid in a graduated cylinder is above or below a certain line. This still leaves open the question of what linguistic judgments are judgments about.

I would argue that linguistic judgments are not judgments about rules (or principles or whatever), or even rule compliance (understood in the sense that we judge that we are in compliance with a particular rule or set of rules that is transparent to us). They are simply judgments about linguistic facts or phenomena (these facts are determined by the linguistic rules).

Following Bogen and Woodward (1988) we can illustrate the relation between theory, phenomena, and data as follows.

Theory $\Longrightarrow$ explains/predicts $\Longrightarrow$ **phenomena** $\Longleftarrow$ is evidence for $\Longleftarrow$ **data**
$\Longleftarrow$ are evidence for $\Longleftarrow$

Phenomena (or facts) are stable and replicable effects or processes that are potential objects of explanation and prediction for scientific theories. In the case of generative linguistics, the phenomena will be linguistic/language-related facts – for example, the fact that certain linguistic forms are unacceptable. As noted earlier, there is no way to establish *a priori* what that domain of facts consists in. Our understanding of the domain may shift over time (as in the case of Polio, discussed above). If they are part of the relevant domain, some of these facts will be predicted to hold given the rules and representations (or principles and parameters or whatever explanatory mechanisms) of the grammar. Such facts will thus *provide evidence* for the theory of grammar, and the theory of grammar will in turn play a role in the *explanation and prediction* of at least some of these facts. So, facts provide evidence for the theory and the theory explains and predicts the facts.

I take *data* to be observational evidence for claims about phenomena. In the case of linguistics, linguistic *data* provide *evidence for* phenomena (like binding facts or "island effects") that are in turn *explained by* the theory of grammar.

In Ludlow (2011, Section 3.3) I argued that linguistic judgments are reliable sources of data, understood as playing the role in theorizing outlined above, but here I want to stress something further. They are not merely reliable sources of data, but they are *inexpensive* sources of data. If we think of the task of forming linguistic judgments as that of constructing scientific experiments we can see why this is so. Care must be taken to control for pragmatic and other effects, but even after this is done the experiment remains a low-cost source of data.

Another way of putting this is that linguistic judgments have the advantage of being "efficient" scientific experiments, and this is a good thing. Theorists want to pursue research methods that provide them the *relevant* data at the least cost. High energy physicists would love to be able to get the data they need without waiting fifteen years for the Large Hadron Collider to come online at the expense of 7.5 billion Euros, but their field of study has identified and pursued questions that cannot be investigated without great cost. While linguistic judgments are not illuminating of every question in linguistics, experience show us that they can be targeted onto a number of crucial questions, they are relatively cost free, and they are sufficient for answering many of these questions. The result has been a rich trove of illuminating data. Under such circumstances, when efficiency of method is a desideratum, there are questions about the need for pursuing more expensive sources of data. This is, I believe, the point being made in Chomsky (1965).

> The critical problem for grammatical theory today is not a paucity of evidence but rather the inadequacy of present theories of language to account for masses of evidence that are hardly open to serious question .... It seems to me that sharpening of the data by more objective tests is a matter of small importance for the problems at hand ... Perhaps the day will come when the kinds of data that we now can obtain in abundance will be insufficient to resolve deeper questions concerning the structure of language. (Chomsky, 1965, 20–21)

Of course, that was the state of affairs, as Chomsky saw it, in 1965. Schütze (1996) has made the case that while this may have been true in earlier days of linguistic theorizing, it is no longer true, and he argues that we ought to critically examine linguistic judgments. Schütze is quite correct to assert the current value of linguistic judgments is an open empirical question, but I am dubious as to his conclusion. It certainly seems false for areas of linguistics in which I work. In any case, we may find that linguistic judgments run out of steam, not in general, but for investigating certain classes of problems. The question is, when do we move on to more expensive modes of gathering data (psycholinguistic experiments, for example).

In my view, the important question is not whether linguistic judgments are still reliable in general. It is rather, how to know when they have been exhausted as a source of data for a particular problem space in linguistics (or any other domain). When do we have to move to more expensive modes of data gathering. As Thomke (2001) puts the question, what is the *optimal switching point*?

Thomke suggests that the switching point should be determined by the *efficiency of the experiment*, which is understood as the economic value of information gained from experiment minus the costs of doing the experiment. In business, unlike in many academic contexts, it is possible to quantify the economic value of the information gained by the experiment. It is also possible to quantify the costs of conducting the experiments (including opportunity costs – that is, the cost of missing out on doing something else). In linguistics, the value of the experiment (and cost of conducting it) is more subjective, obviously.

What does efficiency mean in a domain like linguistics where it is difficult to quantify the value of knowledge? Obviously, different researchers may value discoveries differently, depending upon their goals. Some may be looking for results in linguistics that can aid in natural language processing. Others may be looking for results that help us to understand the nature of the language faculty and how it evolved. It may not be exciting to hear, but arguably the best test of efficiency is for the theorist to ask, "is this worth my time?" But there is more to it than that, because there is also an admonition to consider the value of other research methods once the learning curve for them has been conquered. The correct stance it to always be on the lookout for more productive and efficient methods, where that efficiency is understood in the long term.

6 Conclusion

The considerations raised in the sections above might be quickly glossed as "let a thousand flowers bloom", but this characterization would be incorrect. Multiple formal frameworks do not entail an open-ended approach to theorizing – not every framework is viable. And it is not true that a thousand approaches are viable for each task and theoretical interest or for every theorist – once the research goals and interests are fixed, there may be only one species of flower that gets to bloom. The point is this: The wide range of theoretical approaches in the field (including those discussed in this volume) are not indicative of a science in crisis, but rather of a broad range of research goals and tools and abilities for addressing those goals. They are signs of a healthy research ecology within a science. If we wish to pick winners from among these frameworks, cross-theoretical claims of superiority are unlikely to hold up to serious scrutiny; research frameworks tend to die when they are no longer the most fruitful frameworks for investigators given their goals, not because they have been "refuted". This is as it should be. The only "best theory" criterion that matters is whether the framework is efficient and fruitful for the researcher, given her interests and goals.

References

Absalom, Matthew & John Hajek. 1997. Raddopiomento sintatico: what happens when the theory is too tight. In Pier Marco Bertinetto, Livio Gaeta, Georgi Jetchev & David Michaels (eds.), *Certamen Phonologicum II*, 159–179. Torino: Rosenberg and Sellier.

Assad, Fakhry & W. Charles Cockburn. 1972. Four-year study of WHO virus reports on enterovirus other than poliovirus. *Bulletin of The World Health Organization* 46. 329–336.

Ayala, Francisco. 2009. Darwin and the scientific method. *PNAS [Proceedings of the National Academy of Sciences]* 106(suppl. 1). 10033–10039.

Barrios, Edison. 2016. Simple is not easy. *Synthese* 193. 2261–2305.

Bogen, Jim & James Woodward. 1988. Saving the phenomena. *Philosophical Review* 97. 3–25.

Bresnan, Joan & Ronald M. Kaplan. 1982. Introduction: grammars as mental representations of language. In J. Bresnan (ed.), *The mental representation of grammatical relations*, xvii–lii. Cambridge: MIT Press.

Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge: MIT Press.

Chomsky, Noam. 1975. *The Logical structure of linguistic theory*. Chicago: University of Chicago Press.

Chomsky, Noam. 1980. *Rules and representations*. New York: Columbia University Press.

Chomsky, Noam. 1995. *The minimalist program*. Cambridge: MIT Press.

Chomsky, Noam. 2000. Minimalist inquiries: the framework. In R. Martin, D. Michaels & J. Uriagereka (eds.), *Step by step: Essays in honor of Howard Lasnik*, 89–155. Cambridge: MIT Press.

Chomsky, Noam. 2001a. Derivation by phase. In Michael Kenstowicz (ed.), *Ken Hale: A life in language*, 1–52. Cambridge: MIT Press.

Chomsky, Noam. 2001b. Beyond Explanatory Adequacy. Manuscript, Dept. of Linguistics and Philosophy, MIT.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper and Rowe.

Church, Alonzo. 1936. A note on the Entscheidungsproblem. *The Journal of Symbolic Logic* 1. 40–41.

Darwin, Francis & Albert Seward (eds.). 1903. *More letters of Charles Darwin: A record of his work in a series of hitherto unpublished letters*. London: John Murray.

Devitt, Michael. 2006. *Ignorance of language*. Oxford: Oxford University Press.

Everett, Daniel L. 1986. Pirahã. In Desmond C. Derbyshire & Geoffrey K. Pullum (eds.), *Handbook of Amazonian languages. Vol. 1*, 200–325. Berlin: Mouton de Gruyter.

Everett, Daniel L. 2005. Cultural constraints on grammar and cognition in Pirahã. *Current Anthropology* 46. 621–646.

Eves, Howard & Carroll Newsom. 1965. *An introduction to the foundations and fundamental concepts of mathematics*. New York: Holt, Rinehart and Winston.

Gazdar, Gerald, Ewan Klein, Geoffrey Pullum & Ivan A. Sag. 1985. *Generalized phrase structure grammar*. Cambridge: Harvard University Press.

Gear, James. 1981. Non-polio causes of polio-like paralytic syndromes. *Review of Infectious Diseases* 6. 5379–5384.
Gould, Stephen Jay. 1990. Down on the farm. *The New York Review of Books*. January 18, 1990. http://www.nybooks.com/articles/1990/01/18/down-on-the-farm/ (accessed February 15, 2018).
Hanson, N. Russell. 1958. *Patterns of discovery*. Cambridge: Cambridge University Press.
Harris, Randy. 1993. *The linguistics wars*. Oxford: Oxford University Press.
Haspelmath, Martin. 2010. Framework-free grammatical theory. In Bernd Heine & Heiko Narrog (eds.), *The Oxford handbook of linguistic analysis*, 341–365. Oxford: Oxford University Press.
Hawkins, John A. 2002. Symmetries and asymmetries: their grammar, typology and parsing. *Theoretical Linguistics* 28(2). 95–149.
Hawkins, John A. 2004. *Efficiency and complexity in grammars*. Oxford: Oxford University Press.
Hempel, Carl G. 1966. *Philosophy of natural science*. Engelwood Cliffs: Prentice Hall.
Katz, Jerrold J. & Jerry A. Fodor. 1963. The structure of a semantic theory. *Language* 39. 170–210.
Kitcher, Phillip. 1981. Mathematical rigor—who needs it? *Nous* 15. 469–494.
Johnson, Mark. 1989. Grammar as logic, parsing as deduction. In Steve Abney & Robert C. Berwick (eds.), *The MIT parsing workshop papers*. Distributed by MIT AI Lab, Cambridge, Massachusetts.
Johnson, Mark. 1991. Deductive parsing: the use of knowledge of language. In Robert C. Berwick, Steve Abney & Carol Tenny (eds.), *Principle-based parsing: Computation and psycholinguistics*, 39–64. Dordrecht: Kluwer.
Kuhn, Thomas S. 1970. *The structure of scientific revolutions*, 2nd edn. Chicago: University of Chicago Press.
Lévi-Strauss, Claude. 1953. Remarks. In Sol Tax & Loren Eiseley (eds.), *An appraisal of anthropology today*, 349–352. Chicago: University of Chicago Press.
Lopocaro, Michele. 1989. History and geography of radoppiomento sintattico: Remarks on the evolution of a phonological rule. In Pier Marco Bertinetto & Michele Lopocaro (eds.), *Certamen phonologicum: papers from the 1987 Cortona Phonology Meeting*. Torino: Rosenberg and Sellier.
Lindsay, Robert Bruce. 1937. The meaning of simplicity in physics. *Philosophy of Science* 4. 151–167.
Ludlow, Peter. 2011. *The philosophy of generative linguistics*. Oxford: Oxford University Press.
Nevins, Andrew, David Pesetsky & Cilene Rodrigues. 2009. Pirahã exceptionality: A reassessment. *Language* 85. 355–404.
Newmeyer, Frederick J. 1986. *Linguistic theory in America*, 2nd edn. San Diego: Academic Press.
Popper, Karl R. 1959. *The logic of scientific discovery*. London: Hutchinson.
Popper, Karl R. 1985. *Unended quest: An intellectual autobiography*. La Salle: Open Court Publishing.
Postal, Paul. 1972. The best theory. In Stanley Peters (ed.), *Goals of linguistic theory*, 131–179. Englewood Cliffs: Prentice-Hall.
Pullum, Geoffrey. 1989. Topic…comment: Formal linguistics meets the boojum. *Natural Language and Linguistic Theory* 7. 137–143.
Ross, John. 1967. Constraints on variables in syntax. PhD Dissertation, MIT.
Sabin, Albert. 1981. Paralytic poliomyelitis: Old dogmas and new perspectives. *Review of Infectious Diseases* 3. 343–364.
Salles, Raiane Oliveira. 2015. Understanding recursion and looking for self-embedding in Pirahã: The case of possessive constructions. Dissertação de Mestrado, Potífica Universidade Cattólica do Rio de Janiero.
Sauerland, Uli. 2015. False speech reports in Pirahã: A comprehension experiment. http://ling.auf.

net/lingbuzz/002850.
Schachter, Paul & Fe Otanes. 1972. *Tagalog reference grammar*. Berkeley: University of California Press.
Schütze, Carson. 1996. *The empirical basis of linguistics: Grammaticality judgments and linguistic methodology*. Chicago: University of Chicago Press.
Sellars, Wilfred. 1949. Acquaintance and description again. *The Journal of Philosophy* 46. 496–504.
Sober, Elliot. 1975. *Simplicity*. Oxford: Oxford University Press.
Stabler, Edward. 1992. *The logical approach to syntax*. Cambridge: MIT Press.
Stott, Rebecca. 2003. *Darwin and the barnacle: The story of one tiny creature and history's most spectacular scientific breakthrough*. New York: W.W. Norton.
Thomke, Stefan. 2001. The impact of technology on knowledge creation: A study of experimentation in integrated circuit design. In Ikujiro Nonaka & Toshihiro Nishiguchi (eds.), *Knowledge Emergence: Social, Technical, and Evolutionary Dimensions of Knowledge Creation*, 76–92. Oxford: Oxford University Press.
Turing, Alan. 1952. The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society of London B* 237. 37–72.
Ventor, Craig. 2008. Bigger Faster Better. *Seed magazine*. Nov. 20 2008. http://seedmagazine.com/stateofscience/sos_feature_venter_p1.html (last accessed 24/12/2018).
Williamson, Timothy. 2004. Philosophical "intuitions" and scepticism about judgement. *Dialectica* 58. 109–153.
Woodger, John. 1937. *The axiomatic method in biology*. Cambridge: Cambridge University Press.
Woodger, John. 1939. *The technique of theory construction* (International Encyclopedia of Unified Science Vol. II No. 5). Chicago: University of Chicago.

Pius ten Hacken

20 The research programme of Chomskyan linguistics

Abstract: A research programme is the set of usually implicit assumptions in a scientific community that makes its members work as if they followed a pure version of the empirical cycle. In the case of linguistic theory, the research programme specifies the nature of language, the nature of the data about language and the criteria for evaluating a theory of language. The concept of research programme is first illustrated with Chomskyan linguistics. Then it is shown how the analysis of the differences in the assumptions in Generalized Phrase Structure Grammar, Lexical-Functional Grammar and Parallel Architecture can be used to explain how and to what extent results obtained in one theory are reusable in another. Such differences can also explain incommensurability effects, where successful communication is not possible.

1 Research programmes

Kuhn (1962) was one of the most important publications in the philosophy of science of the 20th century. The central concept Thomas Kuhn (1922–1996) introduces in this work is that of the *paradigm*. The accessible style of Kuhn's writing has given rise to a range of different interpretations of this concept, which his later works attempted to make more specific. Many of the issues are presented in a coherent way in Hoyningen-Huene (1989). Kuhn's preface to Hoyningen-Huene (1989) indicates that the way these issues are settled there corresponds to his intentions. Therefore, I will take this presentation as authoritative.

Kuhn's theory of science is often contrasted with the one presented by Karl Popper (1902–1994), e. g. by Fuller (2003). Both Popper (1959) and Kuhn (1962) can be considered as reactions against the logical positivism of the *Wiener Kreis*. Wiener Kreis (1929) presented a view of science in which scientific theories would follow from logically correct operations on basic observations. The aim is to ensure that scientific theories represent truth.[1] Popper's objections to this model of science focus on the fact that the leap from basic observations to scientific theories cannot be bridged by logically correct operations. His solution is to admit that scientific theories have the status of hypotheses. It is less important how the hypothesis is arrived at than that it can be falsified. A theory must make predictions. If experiments yield results that contradict these predictions, the theory must be discarded.

1 This does not mean that Wiener Kreis (1929) proposes a naive inductionist view of science. The nature of the correct operations was a matter of debate, but this debate will not concern us here.

https://doi.org/10.1515/9783110540253-020

Kuhn (1962) has a different orientation. Whereas for Wiener Kreis (1929) and for Popper (1959), the central issue is how to distinguish proper science from pseudoscience, Kuhn is more interested in explaining how science operates in practice. Popper's falsifiability criterion does not correspond to scientific practice for two reasons. First, if an experiment yields results falsifying a theory T, T is not discarded in favour of nothing. T can only be discarded if there is an alternative theory U to replace it. Second, counterevidence does not have the immediate and absolute effect Popper predicts. The most common reaction is not to replace T with a completely new theory U, but to devise a variant of T, say T', which avoids the contradicting evidence. If there is no obvious candidate for T' available, or if there are several options, it is quite common that T remains the accepted theory even though it has counterevidence.

A common perception of scientific practice is based on what Nagel (1961) calls the *empirical cycle*. This cycle relates data, generalizations, theories and predictions. It does not have a clear starting point, but it can be understood most easily when starting from an existing theory. From a prospective viewpoint, the theory gives rise to predictions. These predictions are tested, giving rise to new data. These new data lead to generalizations which can be used to modify the theory. From a retrospective viewpoint, the existing theory was the result of incorporating generalizations based on data. However, there is no logically coherent way to get this process started. Theories and data are connected in a chicken-and-egg fashion. Wiener Kreis (1929) tried to solve this problem by returning to basic observations that cannot be doubted as a firm foundation for scientific theories.[2] However, for sophisticated theories, it is difficult to derive them from basic observations. Kuhn (1962) introduces the concept of *normal science* instead.

In normal science, there is a theory that is accepted as a starting point and a paradigm which can be used as a basis for the decisions to be taken at each stage of the empirical cycle. The paradigm indicates which observations are interesting, what kind of generalizations are promising, how a theory can be evaluated, etc. The paradigm is usually not discussed or explained explicitly, but it typically emerges in individual researchers as a side effect of being introduced to the field of study.

The central phenomenon to which Kuhn (1962) is devoted is the issue of *scientific revolutions*. Kuhn (1962) proposes that a scientific revolution is the replacement of one paradigm by another. The emergence of a new, competing paradigm only occurs when the field enters a crisis. There is no objective measure for when the problems in a field amount to a crisis. Rather, the crisis is a matter of perception. The initiators of the new paradigm apparently perceived the counterevidence against the existing theories as overriding the promise of progress in the extension of their coverage. The discussion between proponents of competing paradigms is problematic because of

2 This is not to imply that Wiener Kreis (1929) holds the view that theories should be *deduced* from basic observations. The operations for *deriving* theories are more complex than simple deduction.

what Kuhn (1962) calls their *incommensurability*. The evaluation criteria are paradigm-specific, so that a theory T_1 in paradigm P_1 cannot be compared objectively to a theory T_2 in paradigm P_2. Inevitably, T_1 is better according to the evaluation criteria in P_1, but worse according to the evaluation criteria in P_2. There are no evaluation criteria that are paradigm-independent and strong enough to evaluate T_1 in comparison to T_2.

In ten Hacken (2007), I compare this situation to determining whether a football team or a handball team is better. If they play a match but one team plays football and the other handball, they violate each other's rules. If both play football, the football team will win, but this is because the others have an unfair disadvantage. If they play another type of sport, say hockey, the match is fairer, but the result is in a sense irrelevant. This explains why determining the national championship for a particular sport is a very different matter to choosing the sports team of the year.

In the absence of a fully rational decision procedure in the competition between scientific paradigms, the choice of one paradigm over another is determined in part by issues of power, funding, attractivity to students, etc. This has resulted in some tendencies to minimize the rational component in scientific discussions. An example is Woolgar (1988). In the discussion of the application of Kuhn's theory of science to the field of linguistics, such tendencies are so strongly represented that it is hardly possible to mention *paradigm* without evoking a dominance of non-rational factors in their choice.

In ten Hacken (2007), I introduce the concept of *research programme* as a way of emphasizing the purely rational components of Kuhn's theory. It is not meant as a contention that the non-rational components are irrelevant, but by focusing on the rational aspects of the discussion, the aspects of choice that are not fully determined by the internal rationality of the scientific enterprise can be identified more precisely. What counts as rational is of course itself a controversial question. As a practical criterion, I take the research programme to contain those aspects of Kuhn's paradigms that are necessary to make the empirical cycle work in the way as perceived by typical scientists in the field.

The model of a research programme can be supported by an analysis of the discussions in the field. When proponents of different theories within the same research programme have a controversial discussion, they will be able to base their discussion on a general agreement on background assumptions. This means that the shared background provides criteria that are sufficiently powerful for a comparative evaluation of the theories. When proponents of different theories developed in different research programmes discuss the relative merits of their theories, we can often observe *incommensurability effects*. Such effects can range from mutual allegations of misinterpretation of the data to explicit statements that the opponent's point is not understandable. In ten Hacken (2007), I give several detailed analyses of such discussions.

In the domain of linguistics, at least the three issues listed in (1) have to be addressed in a research programme.

(1) a. What is the nature of language as an object of scientific study?
b. What are relevant data for the scientific study of language?
c. What are the criteria for evaluating competing theories?

In Section 2, I will explain how each of the questions in (1) is approached in Chomskyan linguistics. Section 3 gives some examples of alternative answers that have been given and the implications they have for the study of language.

2 Answers to the central questions in Chomskyan Linguistics

In his long career in linguistics, Noam Chomsky has developed many theoretical ideas. Unsurprisingly, later ideas were not always compatible with earlier ones. Four theoretical frameworks are usually distinguished, each with their own assumptions about the organization of syntax and its place in grammar. For the oldest work, Chomsky (1957) is emblematic. The stage of Transformational-Generative Grammar (TGG) is represented by Chomsky (1965). The Principles and Parameters (P&P) framework is presented in its earliest full form in Chomsky (1981). The Minimalist Program (MP) is presented in Chomsky (1993), but significantly modified in Chomsky (2000) with the introduction of phases. Compared to these developments in questions of theory, Chomsky's assumptions about the issues in (1) remain remarkably stable over this entire period. Therefore, it is justified to speak about Chomskyan Linguistics as one research programme. In ten Hacken (2007, 94–124), I discuss the transitions between the various theoretical frameworks and argue that they do not constitute new research programmes, but new theories within the same research programme. Here I will discuss each of the three questions in (1) separately, referring to the different theoretical stages where necessary.

2.1 The nature of language

One of the central ideas of Chomskyan linguistics is the opposition between competence and performance. Chomsky (1965, 4) formulates this distinction as in (2).

(2) "We thus make a fundamental distinction between *competence* (the speaker-hearer's knowledge of his language) and *performance* (the actual use of language in concrete situations)."

There are several other places in his early work where Chomsky makes this distinction in similar terms, but (2) is no doubt the most famous one. In later work, there are fewer references to the distinction, but we also find it, for instance, in Chomsky (1997, 9).

The reference to knowledge in the description of competence in (2) should not be understood in the same way as the knowledge of the capitals of the countries of the EU. Chomsky (1980) distinguishes the sense of *knowledge* intended in (2) from justified belief, capacity, and skill. Chomsky (1980, 92–93) observes that competence is a mental state and there is no necessity to believe the rules of one's language. These rules characterize the mental state. This implies that conscious, explicit knowledge of the rules, in the way we may know the capitals of the EU countries, is irrelevant for the existence of competence. Chomsky (1980, 51–53) discusses the possibility of temporary conditions making the use of language impossible. One could think of someone being unconscious. When these inhibiting conditions are no longer there, the person regains the capacity to use their competence. If we assume that the competence disappeared and reappeared, we would have to explain why it is the same competence that reappeared rather than, for instance, competence in a different language. Chomsky (1980, 100–103) argues against the view that competence equals skill, because the type of learning involved is very different. Whereas linguistic competence arises when a child grows up, as an automatic side effect of being exposed to linguistic performance, language skills (e. g. writing speeches or poetry) can be trained and are subject to a wide variation in the degree of talent individual speakers may have.

Competence is a state of mind that constitutes a necessary but not sufficient condition for the use of language. We need various other types of knowledge to use language in practice, including what Chomsky (1980, 59) calls *pragmatic competence*. We also need various capacities for realizing and understanding language, e. g. speech or writing and hearing or reading abilities. By contrast to pragmatic competence, Chomsky (1980, 59) characterizes *grammatical competence* as in (3).

(3) "By 'grammatical competence' I mean the cognitive state that encompasses all those aspects of form and meaning and their relation, including underlying structures that enter into that relation, which are properly assigned to the specific subsystem of the human mind that relates representations of form and meaning."

It is important to see that competence is an empirical entity. It exists in the mind/brain of a speaker.

For performance, it has never been doubted that it is also an empirical entity, but the formulation in (2) leaves the possibility of different interpretations. For *use*, we can adopt a process reading or a result reading. In the process reading, *performance* is interpreted as equivalent to *language processing*, the activity of the brain involved in the production of linguistic output. In the result reading, *performance* means the output itself, as it can be collected in corpora or sound files. Chomsky introduced the distinction between competence and performance in the discussion with representatives of Post-Bloomfieldian linguistics. As indicated by, for instance, Harris (1951), such linguists started research by collecting a corpus. Therefore it is plausible that

Chomsky originally intended *performance* to mean 'linguistic output' rather than 'linguistic processing'.

Chomsky (1986) introduces a different opposition pair, *I-language* and *E-language*. The *I* and *E* stand for 'internalized' and 'externalized', respectively. Chomsky (1986, 22) describes *I-language* as "[a] notion of structure" which is "[an] element of the mind of the person who knows the language". It is safe to assume that *I-language* is equivalent to *competence*. The meaning of *E-language* is somewhat more elusive. It is definitely not correct to consider it as an equivalent of *performance*, although the two terms have sometimes been confused. Chomsky (1986, 19) describes E-language as in (4).

(4) "[. . .] a collection of actions, or utterances, or linguistic forms (words, sentences) paired with meanings, or [. . .] a system of linguistic forms or events."

The core of (4) seems to be the idea of a language as a collection of linguistic forms paired with meanings. The scope of the occurrences of *or* in (4) is ambiguous, but the interpretation of the two first occurrences seems to be that "linguistic forms" can be replaced by "actions" or "utterances" that are paired with meanings. The last *or* can be interpreted as giving the option of replacing "collection" by "system". This interpretation corresponds to the ones given by Hornstein (2005, 157–158) and Smith and Allott (2016, 33–34).

Chomsky (1986, 26) qualifies E-languages as "not real-world objects but [. . .] artificial, somewhat arbitrary, and perhaps not very interesting concepts". Matthews (1993, 237–240) sees this as a change of mind. In fact, in earlier work, we find such statements as (5), taken from Chomsky and Halle (1968, 3).

(5) "We may think of a language as a set of sentences, each with an ideal phonetic form and an associated intrinsic semantic interpretation."

What is described in (5) is clearly an E-language in the sense of (4). On the same page, Chomsky and Halle describe the grammar of the language as the rule system specifying the sound-meaning correspondence and refer to the systematic ambiguity of *grammar* as the theoretical description of a speaker's competence and the rule system inherent in the competence.

It is not until Chomsky (1976) that we find an explicit discussion of the relation of *language* as described in (5) to linguistic theory. In a discussion of Lewis (1975), Chomsky (1980, 81–85) argues that a language cannot be identified on the basis of a population of speakers. Instead, a language can only be derived from a grammar. This insight has wide-ranging consequences. Uriagereka (1998, 27) formulates them somewhat provocatively in (6).

(6) "English doesn't really exist."

The obvious objection to (6) is that it is written in English. The reply to this objection is that we do not have to know about English as a language in order to write or understand (6). In (6), *exist* is used in the terminologically precise sense of 'being an empirical object'. This can be illustrated by means of the German contrast in (7).

(7) a. Wann kommst du nach Hause?
'when come$_{2SG}$ you to home', i. e. When do you come home?

b. *Wann kommst nach Hause?
'when come$_{2SG}$ to home'

The contrast in (7) can be used to show that German is not a pro-drop language. Even though the inflection of the verb indicates that the subject is second person singular, it is not possible to omit the subject pronoun *du* ('you$_{SG}$'). The point is how the examples in (7) are related to *German*. The contrast reflects the competence of particular speakers. Saying that (7) pertains to German is just a shorthand. Theoretical statements cannot be made about German, but only about the competence of individual speakers.

As summarized in ten Hacken (2007, 274–281), named languages such as English and German are not empirical objects and cannot be used as the basis of empirical claims. We can classify speakers or texts to arrive at a named language, but this makes English dependent on a decision about competence or performance. These decisions are not linguistic, but social and political in nature. In a linguistic context, the names can only be used pretheoretically or prescriptively.

In sum, Chomskyan linguistics distinguishes three senses of *language*. Competence or I-language is realized in the mind/brain of individual speakers and constitutes the basis of linguistic research. Performance is language as utterances or text. It arises from the interaction of competence with other factors. Named languages are not empirical objects. They cannot be used as the basis for linguistic theorizing.

2.2 Language data

The approach to the selection and use of data in Chomskyan linguistics is determined entirely by general considerations. Chomsky (1986, 37) states this in (8), but various similar statements can be found in Chomsky's work from all periods.

(8) "As in the case of any inquiry into some aspect of the physical world, there is no way of delimiting the kinds of evidence that might, in principle, prove relevant."

Despite the liberal statement in (8), Chomskyan linguistics is generally associated with an emphasis on grammaticality judgements, a rejection of corpus use, and an at best lukewarm attitude towards psycho- and neurolinguistic experiments. In fact, to the extent that these impressions are correct, they can be explained on the basis of the assumption that competence is the research object.

The explicit use of grammaticality judgements was a new idea when Chomsky (1957, 15) introduced it. He illustrates the distinction between grammaticality and acceptability with the examples in (9).

(9) a. Colorless green ideas sleep furiously.
 b. * Furiously sleep ideas green colorless.
 c. Have you a book on modern music?
 d. * Read you a book on modern music?

In (9a), the semantic incompatibility of the words tends to complicate the interpretation to the point of making the sentence unacceptable, but it is entirely grammatical. In fact, (9a) has become famous and various interpretations have been proposed in which, for instance, *green* refers to a political movement. This provides a further indication that the sentence is not ungrammatical. The reverse word order in (9b) is ungrammatical.

The contrast of (9c) and (9d) is one where acceptability in the sense of understanding is not a problem. Both sentences can be interpreted easily. Nevertheless, there is a clear distinction in grammaticality. Whereas (9c) may appear rather old-fashioned without *do*-support, it is grammatical. With *read* instead of *have* in (9d), *do*-support is required.

A grammaticality judgement is an introspective judgement which takes into account not only whether and to what extent the expression is correct, but also what causes the degree of correctness. As it is based on an introspective judgement, grammaticality is not directly related to being admitted by a standard grammar of the language. It reflects the speaker's competence, not a normative grammar. As a consequence, grammaticality judgements may differ from speaker to speaker.

A grammaticality judgement can be seen as a small experiment. It is easy to carry out. A problem is what is often called *judgement fatigue*. Considering a sentence of dubious grammaticality long enough will make it so familiar that one tends to judge it as grammatical. Considering too many similar sentences in one session may result in a general blur of grammaticality judgements.

Another property of grammaticality is that in the sense it is used here, it is not a dichotomy but a cline. Whereas formal grammars will either generate a particular sentence or not, intuitive judgements are often a matter of degree. Lasnik and Fiengo (1974, 64) give the examples in (10).

(10) a. Advantage was taken of Mary.
 b. * Advantage was easy to take of Bill.
 c. ? Headway is easy to make on problems like these.

Whereas (10a) is grammatical and (10b) not, the status of (10c) is somewhere in between the two. This illustrates that with grammaticality judgements we do not aim to collect a set of sentences a formal grammar should account for. Rather, we obtain information about a speaker's competence.

Given that competence is individual, it is obvious that grammaticality judgements are individual, too. This is not a flaw. It would not make sense to take a poll among many speakers with grammaticality judgements on the same set of sentences in order to get a percentage. Given (6), such a percentage does not have any theoretical interpretation. The intermediate judgement of (10c) does not mean that some speakers find it grammatical and others not.

Let us now turn to corpora. Before the emergence of Chomskyan linguistics, most linguistic research, at least in America, was firmly corpus-based. Harris (1951, 12) expresses this in (11).

(11) "Investigation in descriptive linguistics consists of recording utterances in a single dialect and analyzing the recorded material. The stock of recorded utterances constitutes the corpus of data, and the analysis which is made of it is a compact description of the distribution of elements within it."

In Chomsky's (1965) terms, (11) restricts the scope of linguistic research to performance. Instead, Chomsky is interested in data that can be used to come up with a theory of competence. In rejecting Harris's (1951) approach, Chomsky does not reject the use of corpora, but only the restriction of the scope to performance.

Given the general approach in (8), rejecting the use of corpora would be irrational. However, as with any other type of data, the limitations have to be taken into account. These limitations are of three types. First, a corpus may contain errors. Thus, COCA (2008–2016) contains (12).

(12) Internet use for health inforamtion among college students.

Clearly, *inforamtion* is a misspelling for *information*. The error can only be recognized, however, by using one's competence. A second limitation is that a corpus may not contain a particular expression for accidental reasons. With the increasing size of corpora, it is possible to find evidence even of rare phenomena, but still, an idiom such as *a bad carpenter blames his tools* is difficult to find. Even by looking for variants by using various two-word combinations, I only found 2 occurrences in COCA, which contains over 520,000,000 words.

The third limitation of corpora is that they are typically the result of performance by many different speakers. For (12), every speaker of English agrees that *inforamtion* is an error. When we are in doubt, however, we should ask the speaker. On the basis of corpus data, Bauer et al. (2013, 76–78) claim that many irregular verbs also have regular forms competing with the irregular ones. Thus, COCA contains 199 occurrences of *knowed*, 22 occurrences of *blowed*, and 6 occurrences of *catched*. The regular forms *knew* (161478), *blew* (9869), and *caught* (49269) are much more frequent. The question is then how to interpret these data. From the perspective of Chomskyan linguistics, the first issue to settle is whether these are oversights or intended uses. Oversights are performance errors that are easily identified by applying one's competence. This requires knowing who produced the utterances and asking them. Ultimately, Chomskyan linguistics is interested in individual speakers' competence, not in the description of what Harris in (11) calls "a single dialect".

Let us finally turn to psycho- and neurolinguistic experiments. The idea of such experiments is to find external evidence for a theory that claims to explain (certain aspects of) competence. In the late 1960s and early 1970s, there were many attempts to elicit data about language processing by means of experiments that asked for a response (such as pressing a button) while listening to an utterance. Fodor et al. (1974) give an overview of these experiments and their results. The results were not very promising. The problem is that there is no direct, obvious connection between the data collected and the underlying question to be answered.

A famous experiment concerns listening to an utterance while a click is produced in the headphone. Subjects are asked to press a button when they hear the click. The hypothesis is that the constituent structure of the sentence influences the moment when the click is heard, so that it is possible to establish, for instance, which of the two structures in (13) is correct.

(13) a. [Anna expects Bernhard] to leave
b. Anna expects [Bernhard to leave]

Depending on which of the two structures in (13) corresponds to the actual one in the speaker's mind, a click should rather be attracted by the boundary before or after *Bernhard*. Before we can draw such a conclusion, however, it has to be established that the attraction takes place at all and how strong it is in general. In practice, the research did not get beyond this stage, because these so-called *calibration experiments* did not yield very encouraging results.

Like psycholinguistic experiments, also neurolinguistic experiments aim to find evidence that is more direct and objective than grammaticality judgements, because it is not consciously mediated by the subject of the experiment. The problem in such experiments is how to link neural activity to linguistic observations. The most common technique in neurolinguistic experiments is functional Magnetic Resonance Imaging (fMRI). Experiments with fMRI are generally not directed towards distinguishing structural contrasts such as (13). Friederici (2012) gives a good impression of what is aimed at, formulated by one of the most prominent researchers in this area. Based on a large number of studies, Friederici (2012, 263) proposes a map of the left hemisphere of the brain identifying the main areas involved in language processing and indicating which ones exchange information. Friederici (2012, 266) includes a *box* with "Questions for future research". These questions are directed towards a further specification of the map. Units are syntax and phonology, not linguistic representations or rules.

Zaccarella and Friederici (2015) extend this idea by focusing on the localization of the *merge* operation. In this way, they bridge the gap between units identified by linguistic theory and locations identified by fMRI. Another type of result is illustrated by Fengler et al. (2016). They investigate how the ability to process complex sentences develops in children.

In accordance with the approach formulated in (8), all such data are of interest to linguistic research. By their nature, however, psycho- and neurolinguistic data do not replace the kind of data that can be gathered from grammaticality judgements and corpora.

2.3 Evaluation criteria

In order to select and evaluate a theory, it is necessary to know the purpose of the theory. In philosophy of science, there has been a move from a theory as firm knowledge to a theory as a hypothesis. The *Wiener Kreis* emphasized that a theory is derived from basic observations. This corresponds to Harris's (1951) view of a grammar in (11) as a compact description of the distribution of linguistic elements in a corpus. In Popper's and Kuhn's systems, theories have the status of a hypothesis. They cover the observations in the sense that they describe a model of the hypothesized underlying reality from which the observations follow. In this sense, the theory can be said to explain the observations. This is also the view of theory adopted in Chomskyan linguistics.

Chomsky (1957, 28–29) introduces three levels of adequacy for grammatical descriptions, observational, descriptive and explanatory adequacy. He calls the aim Harris (1951) formulates in (11) *observational adequacy* and rejects it as a sufficient criterion for the selection of a grammatical description. The reason is that there are numerous possible descriptions that are observationally adequate and there is no non-arbitrary way of choosing one.

The levels of descriptive and explanatory adequacy are tightly interwoven. For descriptive adequacy, the grammar has to describe the organization of a speaker's competence. For explanatory adequacy, the theory of grammar has to provide criteria to select a grammar. Chomsky (1965, 36) observes that it is not necessary to achieve descriptive adequacy before raising the question of explanatory adequacy. Together, they evaluate linguistic theories.

Chomsky (1965) proposes a system of rewrite rules and transformations as a grammatical formalism, which takes a lexicon as the storage of basic elements. This is referred to as the *Standard Theory*. How exactly the rewrite rules and the transformation rules are formulated determines which sentences can be generated. However, the step from sentences to rules is underdetermined. The fact that there are numerous grammars compatible with any finite set of data makes it impossible to select a single grammar on the basis of a collection of grammaticality judgements and corpus data.

Chomsky (1959) had already introduced language acquisition as the central question to be answered by a linguistic theory. The idea is that children are born with a genetically determined predisposition to learn a language. Because of this, all human languages must share the properties that correspond to this predisposition. It is only with the Principles and Parameters (P&P) model of Chomsky (1981) that this condition can be fully activated.

In the P&P model, explanatory adequacy can be achieved by explaining that a particular grammar can be the result of language acquisition. As described by van Riemsdijk and Williams (1984), P&P emerged from a development towards more general statements than individual rewrite rules and transformations. Chomsky's (1970) X-bar Theory was one of the first of such generalizations over constraints on the formulation of rules. Another one was Chomsky's (1973) Subjacency Condition, which constrains movement. The idea is that a constituent can only cross one bounding node at a time. In English, bounding nodes are NP and S.

The contrastive studies on Italian and English collected in Rizzi (1982) and on French and English collected in Kayne (1984) contributed significantly to the insight that differences between languages are often correlated. In a chapter which, according to Rizzi (1982, xii), was written in 1977, Rizzi (1982, 49–76) shows that Italian displays systematic violations of subjacency, which can be explained if we assume that English and Italian have different sets of bounding nodes. In Italian, they are NP and S'. A discussion with examples of the consequences of this difference can also be found in van Riemsdijk and Williams (1984, 58–79).

In P&P, the principles are generalizations such as X-bar Theory and Subjacency. They are universal and are claimed to characterize the predisposition for language acquisition in children. The fact that languages differ can be accounted for in the lexicon and in the parameters that may be associated with the principles. In the case of Subjacency, the child has to find out which bounding nodes the language has.

The transition from the Standard Theory to the P&P model was generally seen by proponents of Chomskyan linguistics as an implementation of language acquisition as a criterion for achieving explanatory adequacy. As Hornstein and Lightfoot (1981) emphasized, language acquisition is in this context first of all a "logical problem". The question is how principles and parameters can be formulated in such a way that children can set the parameters on the basis of the input data available to them.

Despite the priority assigned to the logical problem of language acquisition, the use of language acquisition as a criterion for empirical adequacy also stimulated the use of data from empirical language acquisition studies, cf. White (1981). Lightfoot (1981) draws attention to another aspect, the possibility of studying syntactic change along these lines. Language acquisition by children is determined by the genetic specification of the predisposition for language acquisition and the input. Language change takes place between generations, when the performance of the older generation leads to a different setting of a parameter in the competence of the children. This leads to the identification of minimal conditions for different parameter settings. Another type of study that can be used to collect this type of data involves closely related dialects.

The criterion of language acquisition had been part of the research programme of Chomskyan linguistics from its inception. In ten Hacken (2007, 103–105), I review the evidence for the claim that there was a significant change in the research programme before the Standard Theory of Chomsky (1965) and I conclude that there is hardly any basis to assume such a change. The competition between Standard Theory and P&P, as well as all intermediate variants and alternatives, was framed in a discussion on the basis of the evaluation criteria for theories that the research programme of Chomskyan linguistics provided. Linguists who rejected this transition also rejected (parts of) the research programme.

Progress in science can be measured as an extension of the set of data that are explained or as a deepening of the level of explanation. The transition to P&P achieved both. By providing a framework for a theory of the genetic predisposition for language acquisition, data from one language could be used to decide contentious issues in the description of the grammar of another language. When we formulate a grammar of Russian, everything that is different from English must be the consequence of different parameter settings. However, all these differences must be learnable on the basis of English and Russian monolingual performance data. By increasing the constraints on formulating the grammars of individual languages in this way, we can explain their properties on the basis of language acquisition.

The emergence of the Minimalist Program (MP) is a second episode that has been interpreted differently with regards to the question whether it is a rational continuation of the research programme of Chomskyan linguistics. There are historical, theoretical and epistemological issues to be discussed in this context.

Historically, the transition from P&P to MP can be seen as a reaction to a proliferation of parameters. The problem is similar to the one seen in the transition from Standard Theory to P&P. Earlier, in the absence of constraints on rules, it was impossible to determine the correct rules. Now, in the absence of constraints on parameters, any difference between two languages could be accounted for as the result of a different parameter setting. There was a need for a constraint on parameters.

Theoretically, the most striking difference between P&P and MP is the account of movement. In P&P, Move *α* applies generally, except when constraints make it impossible. In MP, movement is governed by the Last Resort principle. Movement is only possible if it serves to avert immediate crashing of the derivation, i. e. ungrammaticality, because a crucial condition is not met.

Epistemologically, the emergence of MP marks the operationalization of the evaluation criterion of accounting for the evolution of language. Whereas P&P intended to provide a possible answer to the question of how language comes into being in the individual, an answer that relies on the genetically specified predisposition for language acquisition, MP intends also to answer the question of how this predisposition emerged in the evolution of the human species.

The historical and theoretical aspects of the emergence of MP are discussed in some articles in the journal *Natural Language and Linguistic Theory*. Lappin et al. (2000a) defend the thesis that the transition was based more on the authority of Noam Chomsky than on rational considerations. In response, a number of proponents of MP took issue with their analysis. Holmberg (2000), Reuland (2000), Roberts (2000), Piattelli-Palmarini (2000), and Uriagereka (2000) raise various objections to the analysis by Lappin et al. (2000a). In their response, Lappin et al. (2000b) defend and maintain their analysis.

The issue can be framed in terms of the question whether the transition took place within the research programme or not. In order to answer this question, we have to consider the evaluation criteria for the selection of theories. If P&P and MP can be meaningfully compared by applying the same set of evaluation criteria, they belong to the same research programme. Lappin et al. (2000a,b) maintain that it is rather Chomsky's authority than the application of evaluation criteria which determined the outcome, a situation that is typical of scientific revolutions.

Crucial in this analysis is the status of the criterion of evolution. The criterion requires that the theory of the individual speaker's competence and of the genetic predisposition for language acquisition is compatible with an account of the latter's emergence in the evolutionary history of the human species. As I show in ten Hacken (2007, 111), the criterion has a similar status to the criterion of language acquisition in the emergence of P&P. Chomsky (1965, 59) refers to the issue of evolution in very general terms and Chomsky (1980, 239) indicates that little is known about the question, but the theoretical development in the early 1990s created a new situation in which the criterion could be operationalized.

A detailed analysis of the criterion is provided by Hauser et al. (2002). They distinguish FLN (Faculty of Language in the Narrow sense) from FLB (Faculty of Language in the Broad sense). FLB includes FLN with the sensory-motor system and the conceptual-intentional system. FLN is only the abstract computational system. Hauser et al. (2002, 1573) claim that only FLN is uniquely human. Therefore, in the evolution of the human species, it must be possible for FLN to arise in one step. This imposes the condition on the linguistic theory that everything except what emerged in this one step has an independent use for the species. It is this criterion which for them leads to a much superior evaluation of MP over P&P, because MP uses fewer and simpler operations instead of the set of rather complex principles in P&P.

In sum, the evaluation criteria used in Chomskyan linguistics to compare the relative merits of grammatical theories are based on the three issues in (14).

(14) a. Accounting for the speaker's competence.
b. Accounting for language acquisition.
c. Accounting for the evolution of language in the human species.

The criteria in (14) were deployed one by one as progress in theoretical insight required additional criteria to choose between theories.

3 Some alternative answers

In the course of the history of generative grammar, there have been many linguists who accepted some but not all parts of Chomsky's theories. As long as this is only a difference in theory, a completely rational discussion is possible, at least in principle. The problem consists in identifying data that distinguish between the two theories, i. e. data that show that one theory is better than another.

A different kind of problem arises when it is the evaluation criteria rather than the theory that are challenged. There are no meta-evaluation criteria that can be applied to evaluate the merits of different evaluation criteria in the same purely rational way as in the comparison of two theories. Typically, if the evaluation criteria are challenged, also the theory that is formulated is rather different.

In this section, we will look at three challenges to the evaluation criteria in (14). Formalist approaches, including Generalized Phrase Structure Grammar, reject (14a). Lexical-Functional Grammar replaces (14b) by the criterion of language processing. Parallel Architecture takes a different stance on the elaboration of (14c). Each of these call themselves *Generative Grammar*.

3.1 Generalized Phrase Structure Grammar (GPSG) and the formalist approach

Generalized Phrase Structure Grammar (GPSG) is a grammatical framework that is presented by Gazdar et al. (1985). It exemplifies a formalist approach to linguistics in which a grammar does not primarily describe a speaker's competence. GPSG is by no means the only representative of such an approach. It is particularly suitable as an example, because Gazdar et al. (1985) describe the central aspects of the approach explicitly and in contrast to Chomskyan linguistics. Moreover, there have been several published discussions between representatives of GPSG and proponents of Chomskyan linguistics. In ten Hacken (2007, 209–230), I give a more detailed analysis of these discussions and of the research programme adopted in GPSG.

One of the claims in GPSG is that GPSG belongs to Generative Grammar, but Chomsky's work, at least from the 1970s onwards, does not (Gazdar et al. 1985, 6). They see themselves as a better continuation of the objectives of Chomsky's early work than the work in the P&P model. In support of this idea, they refer to statements such as (15) from Chomsky (1965, 8).

(15) "by a generative grammar I mean simply a system of rules that in some explicit and well-defined way assigns structural descriptions to sentences."

Gazdar et al. (1985, 1) formulate what they consider a generative grammar in (16).

(16) "The basic assumption made in generative grammar is that languages can be regarded as collections whose membership is definitely and precisely specifiable. The elements of such a collection are the expressions in the language."

Whereas (15) is compatible with Chomskyan linguistics in the sense that at every stage of its history, theories are expected to make their formalism explicit enough to evaluate it, (16) makes a point that is not compatible with the Chomskyan emphasis on competence. In Chomskyan linguistics, a language is primarily a speaker's competence. This is what Chomsky (1986) calls *I-language*. Judgements about expressions are rooted in this competence and can be used as data. In (16), however, the expressions are themselves part of the language to be described and judgements are used as evidence that an expression belongs to the language. Gazdar et al. (1985, 2) elaborate this in (17).

(17) "A necessary precondition to 'explaining' some aspect of the organization of natural languages is a description of the relevant phenomena which is thorough enough and precise enough to make it plausible to suppose that the language under analysis really is organized in the postulated way."

From the perspective of Chomskyan linguistics, (17) is somewhat hard to interpret. As we saw in Section 2.3, Chomsky (1965) does not see a need to achieve descriptive adequacy before turning to explanatory adequacy. Indeed, the two can only be achieved together, because only the criterion used for explanatory adequacy makes it possible to choose among the several grammars compatible with the judgements and corpus data. In (18), Gazdar et al. (1985, 5) express their attitude towards competence.

(18) "we feel it is possible, and arguably proper, for a linguist (*qua* linguist) to ignore matters of psychology."

There is no explicit claim in (18) that language is not related to competence. However, (18) is not compatible with the view that a linguistic theory must be a description of a speaker's competence. Chomsky (1986) calls notions of language that are not related to competence *E-language*. What (18) implies, then, is that GPSG studies E-language.

The study of E-language can of course start with establishing the expressions of the language, as (16) requires, and concentrate on the description of this set. In terms of Chomsky's (1957) levels of adequacy, this means that the aim is observational adequacy, not descriptive adequacy. For descriptive adequacy, Chomsky (1957, 28) requires (19).

(19) "the grammar gives a correct account of the linguistic intuition of the native speaker, and specifies the observed data (in particular) in terms of significant generalizations that express underlying regularities in the language."

It is striking that in (19) Chomsky relates descriptive adequacy explicitly to the native speaker's intuition. As such, it is hardly compatible with (18). In addition, it should also be noted that *the language* at the end of (19) is meant as competence or I-language. It is not an E-language as (16) assumes as the object of study.

Therefore, GPSG does not belong to the same research programme as Chomskyan linguistics. The conflicting interpretations of basic terms such as *language* gives rise to incommensurability effects in discussions between proponents of the two sides. When in GPSG a claim is made that a particular expression exists in English, it is a claim of the type required by (17). English is a set of expressions and it is somehow possible to establish whether an expression belongs to English or not. In Chomskyan linguistics, such a claim is not considered relevant. English does not exist as an empirical entity, only speakers and their competence do. Expressions do not exist as elements of English, but judgements of speakers can be used as information about the speakers' competence.

3.2 Lexical-Functional Grammar (LFG) and language processing

Lexical-Functional Grammar (LFG) is a framework that is presented in Bresnan (ed.) (1982), with an update in Bresnan (2001). A fundamental assumption is the Competence Hypothesis in (20).

(20) *Competence Hypothesis*
A reasonable model of language use will incorporate, as a basic component, the generative grammar that expresses the speaker-hearer's knowledge of the language.

It is clear from (20) that LFG adheres to (14a). At the same time, (20) emphasizes the role of language use in a way that Chomskyan linguistics does not. Bresnan (1978, 1) attributes (20) to Chomsky (1965, 9). In fact, the text of (20) can be found there except for the name and the initial capital. As shown in ten Hacken (2007, 196–197), however, the context in Chomsky (1965, 9) excludes that he intended to make this statement into a hypothesis of the kind represented in (20). This does not disqualify the Competence Hypothesis, it only shows that it is not Chomsky's hypothesis. On the basis of (20), Bresnan and Kaplan (1982, xxxi) formulate the Strong Competence Hypothesis in (21).

(21) "We can now say that a model satisfies the *strong competence hypothesis* if and only if its representational basis is isomorphic to the competence grammar."

The "competence grammar" in (21) is the "generative grammar" in (20), which is then used in language processing. This leads to an interpretation of Universal Grammar which Bresnan and Kaplan (1982, xviii) allude to in (22).

(22) "For example, principles of Universal Grammar might characterize aspects of the structure of the language-using device."

The formulation in (22) results in a direct counterpart to (14b) as a criterion for the evaluation of grammars. In Chomskyan linguistics, Universal Grammar is taken as the description of what Chomsky (1965, 47) calls the "language-acquisition device".

Kaplan and Bresnan (1982) present what they call "a formalism for representing the native speaker's syntactic knowledge" (1982, 173), which is in line with the Competence Hypothesis in (20). They call this formalism *Lexical-Functional Grammar*. Its core is a unification-based mapping mechanism for determining a functional structure from a constituent structure. This mechanism, which avoids any transformation or movement, is not substantially changed in later work and appears also in Bresnan (2001).

Language processing and language acquisition are both phenomena that a theory of language will have to account for. The difference between Chomskyan linguistics and LFG is that the former uses language acquisition, the latter language processing as an evaluation criterion for theories. Either choice is coherent, but they have different consequences.

The consequence of using language acquisition as an evaluation criterion is, as we saw in Section 2.3, that language acquisition is not only studied as an empirical problem, but also as a logical problem. Every aspect of a grammar of a particular language must be either innate or learnable. This creates a tension between reasons to make more aspects innate and reasons to make more aspects language-specific. If they are innate, they must be valid for all languages and if they are language-specific they must be learnable on the basis of data available to the child.

When we take language processing as the factor used to select a grammar, it is less obvious how a similar tension might emerge. Kaplan and Bresnan's (1982) formalism provides a basis for writing grammars of individual languages. However, given a set of empirical data about the language, there are many possible grammars that are compatible with their formalism. Bresnan and Kaplan (1982) propose to use psycholinguistic experiments to select the right grammar, but, as we saw in Section 2.2, it is in general hard to link any actual experiments directly to the form of the grammar. There is no obvious tension between factors pulling in opposite directions.

In ten Hacken (2007, 184–209), I give a more detailed analysis of the research programme of LFG, including the history of its emergence and some examples of incommensurability effects resulting from the differences between the research programmes of Chomskyan linguistics and LFG. One area where such effects can be observed is in discussions of language acquisition. Working within LFG, Pinker (1982) proposes a theory of language acquisition which focuses on the empirical aspects while neglecting the logical problem. In his discussion, he evaluates aspects of theories originally proposed in Chomskyan linguistics by the criteria of LFG. From the perspective of Chomskyan linguistics, this evaluation is simply irrelevant, because it ignores the role of (14b) as an evaluation criterion.

3.3 Parallel Architecture (PA) and the concept of evolution

Whereas GPSG and LFG were from the outset alternative syntactic theories to Chomsky's, Jackendoff's Parallel Architecture (PA) has the alternative position of syntax in the model of language as one of its most prominent differences to Chomsky's theories.

In fact, Jackendoff (1983, 1990) first developed a theory of semantics. PA made a first appearance in Jackendoff (1997) and was fully elaborated in Jackendoff (2002).

There is a large degree of compatibility between PA and Chomskyan linguistics. This should not be a surprise, because Jackendoff (1993) argues for a view of linguistics in which (14a) and (14b) are crucial for the study of language. The agreement between Chomsky and Jackendoff also extends to the nature of language and the approach to language data, so that it encompasses (1a) and (1b).

Superficially, the main difference between Chomskyan linguistics and PA seems to be one of the place of syntax in the architecture of language. Although Chomsky's views developed significantly over the decades, he always maintained that syntax was the only truly generative component. Phonology and semantics are components that derive representations from a syntactic structure. Jackendoff (2002) proposes a model in which Phonological, Syntactic and Conceptual Structures each have their own set of formation rules. The mapping between them is not a matter of interpreting, as it is for Chomsky, but of linking. In fact, formation rules and linking rules in PA are formally of the same type.

Obviously, the formulation of a new theory does not imply that we are dealing with a new research programme. Standard Theory, P&P, and MP are clearly different theories, but they were all formulated from within the same research programme. As we saw in Section 2.3, the discussion about the question of which of these theories is better can be conducted within the framework offered by the research programme of Chomskyan linguistics.

As argued in detail in ten Hacken (2007, 94–124), proponents of the older theory, i. e. Standard Theory in the transition to P&P, but P&P in the transition to MP, typically insisted on two types of problem with the newer theory. One is that the newer theory was less explicit than the older one, the other that the newer theory required reformulations of explanations available in the old theory and it was not obvious how these reformulations were possible. Proponents of the newer theory, however, invoked the evaluation criteria of the research programme. P&P offers a better account than Standard Theory, because P&P is better at explaining language acquisition. MP is better than P&P, because MP has a better account for the evolution of the language faculty.

The discussion between Jackendoff and Chomsky about the relative merits of MP and PA is of a different nature. Jackendoff (2003, 661–665) argues that PA contributes to the reintegration of linguistics with cognitive neuroscience. His argument invokes four properties that can be constructed as criteria for the evaluation of linguistic theories, summarized in (23).

(23) a. Compatibility with a modular structure of the mind/brain
b. Compatibility with a theory of meaning as a separate module
c. Compatibility with a theory of language processing
d. Compatibility with an incremental adaptation in evolution

The requirement in (23a) refers to the way language is implemented in the brain. The idea that the mind/brain has a modular structure has been put forward by many researchers. In the same way as the origin of the language faculty, the implementation of language in the brain is a question that does not come up in Chomsky's earlier writings. They are not in the classical triad of the nature, origin and use of language as in the title of Chomsky (1986). The first appearance of the implementation question seems to be in Chomsky (1988, 6). Unlike the question of the evolution, however, the implementation is not selected as an evaluation criterion. This decision is similar to the choice of language acquisition over language processing in (14).

The criterion in (23b) is linked to (23a), but not entirely subsumed by it. At an early stage in his research, Jackendoff (1983, 17) had postulated the Conceptual Structure Hypothesis in (24).

(24) *Conceptual Structure Hypothesis*
There is a *single* level of mental representation, *conceptual structure*, at which linguistic, sensory, and motor information are compatible.

In (23b), the status of (24) is changed. Instead of a hypothesis that is used as a basis for the development of a theory, it becomes a criterion for evaluating theories.

The use of language processing in (23c) should be distinguished from the use in LFG, as discussed in Section 3.2. In LFG, Universal Grammar is a language-independent processing mechanism. This makes language processing an alternative to language acquisition in (14b). Such a position is rejected by Chomskyan linguistics and also by Jackendoff. What Jackendoff (2003, 662–664) means, as summarized in (23c) is that the grammar as it is described by linguists plays a role in language processing. He rejects the idea that processing considerations are entirely irrelevant for the evaluation of a theory. This reduces the weight of (14b) to the extent that language acquisition is no longer the only criterion at this level of generalization, but it does not replace (14b), as LFG does.

Most of the subsequent discussion between proponents of Chomskyan linguistics (MP) and proponents of PA centred on (14c) and (23d). Important contributions to this discussion were Pinker and Jackendoff (2005), Fitch et al. (2005), and Jackendoff and Pinker (2005). Interestingly, while both frameworks include (14c) as an evaluation criterion, they have diverging views on how this criterion should be interpreted. As we saw in Section 2.3, Hauser et al. (2002) interpret (14c) on the basis of the hypothesis that language emerged as an exaptation. The term *exaptation* was introduced by Gould and Vrba (1982) for properties that evolved for one function and happened to be useful for another. By contrast, Jackendoff takes (23d) as the interpretation of (14c). He assumes that language as we know it today is the result of a gradual evolutionary process.

As a consequence of these different interpretations of (14c), MP and PA end up with a quite different heuristic for finding a theory of language. Hauser et al. (2002) reduce FLN as much as possible. The aim is to find a mechanism that is at the same time powerful enough to unlock the potential of language and simple enough to emerge as a single-step exaptation. In their view, the key to language must be a formally simple operation that connects independently existing phonological and semantic components. These phonological and semantic components are then shared with other species.

In PA, (23d) requires that intermediate stages in the evolution should be identified. Jackendoff (2002, 231–264) discusses potential evolutionary paths. A crucial point is that each incremental step should lead to an evolutionary advantage. Jackendoff (2002, 238) proposes a number of such incremental developments. It is not sufficient to have a single *protolanguage* stage in between states labelled as *no language* and *full language*.

The question whether language resulted from exaptation or from incremental evolution is an empirical one. Currently, we do not know the answer, but at least in principle it is possible that at some future point in time, we will find out. The question is rather a question of evolutionary biology than one of grammatical theory. However, it has wide-ranging consequences for the evaluation of theories of grammar. Clearly, Chomskyan linguistics will be in a better position if language turns out to be indeed an exaptation, because it has included this hypothesis as an evaluation criterion into its research programme. Conversely, PA will be better placed if language turns out to be actually the result of an incremental evolutionary process, because Jackendoff developed PA with (23d) as one of the criteria in mind.

4 Conclusion

In this chapter, I presented the concept of *research programme* and applied it to Chomskyan linguistics and a selection of competing frameworks. The notion of *research programme* as it is used here derives from Kuhn's (1962) theory of science. Whereas Kuhn intends to account for the way science is conducted and develops, in a research programme only the intellectual aspects are included. This separates the discussion from approaches that prioritize social interaction and authority as dominating factors. In linguistics, an example of such an approach is the one adopted by Huck and Goldsmith (1995). Instead, the central questions in a research programme are listed in (1).

The discussion of the research programme of Chomskyan linguistics in Section 2 shows the answers to the questions in (1). At a metatheoretical level, it demonstrates the historical unity of Chomskyan linguistics and its intellectual coherence. One can disagree with the assumptions, but they are not incoherent. Chomskyan linguistics focuses on language as competence, accepts all types of data to the extent that they can be related to competence and uses the evaluation criteria in (14). In the history of Chomskyan linguistics, the set of evaluation criteria was extended in the sense that language acquisition in (14b) and evolution in (14c), which had been present from the outset as general ideas, became actual criteria for the selection of theories.

In Section 3, another application of the concept of *research programme* is illustrated. It is shown how other types of generative grammar differ from Chomskyan linguistics. For each of the evaluation criteria in (14), one framework was selected that takes a different course in this respect. GPSG rejects the priority of competence and aims to account for a language as a set of form-meaning pairs. LFG rejects the priority of language acquisition over language processing in the choice of theories. PA rejects the idea of language emerging as an exaptation rather than as a series of gradual adaptations in the evolution of the human species. It is not my intention here to indicate which approach is 'right'. What is clear, however, is that GPSG, LFG and PA do not belong to the same research programme as Chomskyan linguistics, but each to a different one.

By separating the intellectual aspects of Kuhn's theory of science from the social aspects, it is possible to indicate the intellectual coherence of a research programme and identify the differences between theories developed from within different research programmes. In this way, the concept of *research programme* contributes to a better understanding of linguistic discussions. In particular, it opens a perspective for the analysis of incommensurability effects in such discussions, which in turn indicates the potential for the successful reuse of theoretical ideas across frameworks.

References

Bauer, Laurie, Rochelle Lieber & Ingo Plag. 2013. *The Oxford reference guide to English morphology*. Oxford: Oxford University Press.

Bresnan, Joan. 1978. A realistic transformational grammar. In Morris Halle, Joan Bresnan & George A. Miller (eds.), *Linguistic theory and psychological reality*, 1–59. Cambridge (Mass.): MIT Press.

Bresnan, Joan & Ronald M. Kaplan. 1982. Introduction: Grammars as mental representations of language. In Joan Bresnan (ed.), *The mental representation of grammatical relations*, xvi–lii. Cambridge (Mass.): MIT Press.

Bresnan, Joan. 2001. *Lexical-functional syntax*. Oxford: Blackwell.

Bresnan, Joan (ed.). 1982. *The mental representation of grammatical relations*. Cambridge (Mass.): MIT Press.

Chomsky, Noam. 1957. *Syntactic structures*. Den Haag: Mouton.

Chomsky, Noam. 1959. Review of: Skinner, Burrhus F. 1957. *Verbal behavior*. Englewood Cliffs NJ: Prentice-Hall. *Language* 35. 26–58.

Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge (Mass.): MIT Press.

Chomsky, Noam. 1970. Remarks on nominalization. In Roderick A. Jacobs & Peter S. Rosenbaum (eds.), *Readings in English transformational grammar*, 11–61. Waltham (Mass.): Ginn.

Chomsky, Noam. 1973. Conditions on transformations. In Stephen R. Anderson & Paul Kiparsky (eds.), *A festschrift for Morris Halle*, 232–286. New York: Holt, Rinehart & Winston.

Chomsky, Noam. 1976. On the biological basis of language capacities. In Robert W. Rieber (ed.), *The neuropsychology of language: Essays in honor of Eric Lenneberg*, 1–24. New York: Plenum.

Chomsky, Noam. 1980. *Rules and representations*. New York: Columbia University Press.

Chomsky, Noam. 1981. *Lectures on government and binding*. Dordrecht: Foris.

Chomsky, Noam. 1986. *Knowledge of language: Its nature, origin, and use*. Westport (Conn.): Praeger.

Chomsky, Noam. 1988. *Language and problems of knowledge*. Cambridge (Mass.): MIT Press.
Chomsky, Noam. 1993. *Language and thought*. Wakefield RI: Moyer Bell.
Chomsky, Noam. 1997. *Perspectives on power: Reflections on human nature and the social order*. Montréal: Black Rose.
Chomsky, Noam. 2000. Minimalist Inquiries: The Framework. In Roger Martin, David Michaels & Juan Uriagereka (eds.), *Step by step: Essays on minimalist syntax in honor of Howard Lasnik*, 89–155. Cambridge MA: MIT Press.
Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.
COCA (2008–2016). *The Corpus of Contemporary American English*. Mark Davies (ed.). http://corpus.byu.edu/coca/ (accessed 23 April 2017).
Fengler, Anja, Lars Meyer & Angela D. Friederici. 2016. How the brain attunes to sentence processing: Relating behavior, structure, and function. *NeuroImage* 129. 268–278.
Fitch, W. Tecumseh, Marc D. Hauser & Noam Chomsky. 2005. The evolution of the language faculty: Clarifications and implications. *Cognition* 97. 179–210.
Fodor, Jerry A., Thomas G. Bever & Merrill F. Garrett. 1974. *The psychology of language: An introduction to psycholinguistics and generative grammar*. New York: McGraw-Hill.
Friederici, Angela D. 2012. The cortical language circuit: From auditory perception to sentence comprehension. *Trends in Cognitive Sciences* 16. 262–268.
Fuller, Steve. 2003. *Kuhn vs Popper: The struggle for the soul of science*. Cambridge: Icon Books.
Gazdar, Gerald, Ewan Klein, Geoffrey Pullum & Ivan A. Sag. 1985. *Generalized Phrase Structure Grammar*. Cambridge (Mass.): Harvard University Press.
Gould, Stephen Jay & Elisabeth S. Vrba. 1982. Exaptation – a missing term in the science of form. *Paleobiology* 8. 4–15.
ten Hacken, Pius. 2007. *Chomskyan linguistics and its competitors*. London: Equinox.
Harris, Zellig S. 1951. *Methods in structural linguistics*. Chicago: University of Chicago Press.
Hauser, Marc D., Noam Chomsky & W. Tecumseh Fitch. 2002. The faculty of language: What is it, who has it, and how did it evolve? *Science* 298. 1569–1579.
Holmberg, Anders. 2000. Am I unscientific? A replay to Lappin, Levine, and Johnson. *Natural Language and Linguistic Theory* 18. 837–842.
Hoyningen-Huene, Paul. 1989. *Die Wissenschaftsphilosophie Thomas S. Kuhns*. Vieweg: Braunschweig.
Hornstein, Norbert. 2005. Empiricism and rationalism as research strategies. In James McGilvray (ed.), *The Cambridge companion to Chomsky*, 145–163. Cambridge: Cambridge University Press.
Hornstein, Norbert & David Lightfoot. 1981. Introduction. In Norbert Hornstein & David Lightfoot (eds.), *Explanation in linguistics: The logical problem of language acquisition*, 9–31. London & New York: Longman.
Huck, Geoffrey J. & John A. Goldsmith. 1995. *Ideology and linguistic theory: Noam Chomsky and the deep structure debates*. London: Routledge.
Jackendoff, Ray 1983. *Semantics and cognition*. Cambridge (Mass.): MIT Press.
Jackendoff, Ray 1990. *Semantic structures*. Cambridge (Mass.): MIT Press.
Jackendoff, Ray 1993. *Patterns in the mind: Language and human nature*. New York: Harvester/Wheatsheaf.
Jackendoff, Ray 1997. *The architecture of the language faculty*. Cambridge (Mass.): MIT Press.
Jackendoff, Ray 2002. *Foundations of language: Brain, meaning, grammar, evolution*. Oxford: Oxford University Press.
Jackendoff, Ray 2003. Précis of Foundations of language: Brain, meaning, grammar, evolution. *Behavioral and Brain Sciences* 26. 651–707.

Jackendoff, Ray & Steven Pinker. 2005. The nature of the language faculty and its implications for evolution of language (Reply to Fitch, Hauser, and Chomsky). *Cognition* 97. 211–225.
Kayne, Richard S. 1984. *Connectedness and binary branching*. Dordrecht: Foris.
Kuhn, Thomas S. 1962. *The structure of scientific revolutions*. Chicago: University of Chicago Press.
Lappin, Shalom, Robert D. Levine & David E. Johnson. 2000a. The structure of unscientific revolutions. *Natural Language and Linguistic Theory* 18. 665–671.
Lappin, Shalom, Robert D. Levine & David E. Johnson. 2000b. The revolution confused: A response to our critics. *Natural Language and Linguistic Theory* 18. 873–890.
Lasnik, Howard & Robert W. Fiengo. 1974. Complement object deletion. *Linguistic Inquiry* 5. 535–571.
Lewis, David. 1975. Languages and language. In Keith Gunderson (ed.), *Language, mind, and knowledge*, 3–35. Minneapolis: University of Minnesota Press.
Lightfoot, David. 1981. Explaining syntactic change. In Norbert Hornstein & David Lightfoot (eds.), *Explanation in linguistics: The logical problem of language acquisition*, 209–240. London & New York: Longman.
Matthews, Peter H. 1993. *Grammatical theory in the United States from Bloomfield to Chomsky*. Cambridge: Cambridge University Press.
Nagel, Ernest. 1961. *The structure of science: Problems in the logic of scientific explanation*. Indianapolis: Hackett.
Piattelli-Palmarini, Massimo. 2000. The metric of open-mindedness. *Natural Language and Linguistic Theory* 18. 859–862.
Pinker, Steven. 1982. A theory of the acquisition of lexical interpretive grammars. In Joan Bresnan (ed.), *The mental representation of grammatical relations*, 655–726. Cambridge (Mass.): MIT Press.
Pinker, Steven & Ray Jackendoff. 2005. The faculty of language: What's special about it? *Cognition* 95. 201–236.
Popper, Karl R. 1959. *The logic of scientific discovery*. London: Routledge.
Reuland, Eric. 2000. Revolution, discovery, and an elementary principle of logic. *Natural Language and Linguistic Theory* 18. 843–848.
Roberts, Ian. 2000. Caricaturing dissent. *Natural Language and Linguistic Theory* 18. 849–857.
van Riemsdijk, Henk & Edwin Williams. 1984. *Introduction to the theory of grammar*. Cambridge (Mass.): MIT Press.
Rizzi, Luigi. 1982. *Issues in Italian syntax*. Dordrecht: Foris.
Smith, Neil & Nick Allott. 2016. *Chomsky: Ideas and Ideals*, Third edn. Cambridge: Cambridge University Press.
Uriagereka, Juan. 1998. *Rhyme and reason: An introduction to minimalist syntax*. Cambridge (Mass.): MIT Press.
Uriagereka, Juan. 2000. On the emptiness of 'design' polemics. *Natural Language and Linguistic Theory* 18. 863–871.
White, Lydia. 1981. The responsibility of grammatical theory to acquisition data. In Norbert Hornstein & David Lightfoot (eds.), *Explanation in linguistics: The logical problem of language acquisition*, 241–271. London & New York: Longman.
Wiener Kreis. 1929. *Wissenschaftliche Weltauffassung*. Wien: Wolf.
Woolgar, Steve. 1988. *Science: The very idea*. Chichester: Horwood & London: Tavistock.
Zaccarella, Emiliano & Angela D. Friederici. 2015. Merge in the human brain: A sub-region based functional investigation in the left pars opercularis. *Frontiers in Psychology* 6. https://doi.org/10.3389/fpsyg.2015.01818 (accessed 23 April 2017).

András Kertész and Csilla Rákosi

21 Conclusions: On the use of the comparison of syntactic theories

1 Introductory remark

The present handbook is unique in at least two respects. In Part I – besides demonstrating how variegated object-scientific syntactic theories focusing on the syntax of human language are, which other handbooks of syntactic theories on the market may also achieve – it invites the reader to make a comparison of the syntactic theories overviewed. At the outset, the comparison of theories requires that they be reflected on at a meta-level. However, in addition, metatheoretical reflection is present even more explicitly in Part II, since it consists of a collection of metatheoretical approaches to syntactic theorizing. Accordingly, both of these properties suggest that the present handbook is basically a *metatheoretical enterprise*. Why should syntactic research be accompanied by metatheoretical reflection? In order to answer this question, we will now focus on revealing two factors: we will clarify some misconceptions affecting the use of metatheoretical reflection on syntactic inquiry; and we will touch on a kind of metatheoretical perspective that we believe to be both well-motivated and useful.

Let us begin with a very short overview of the state of the art in the philosophy of science and its relation to linguistic metatheory.

2 On the history of the philosophy of science

2.1 17–19th centuries

Traditionally, the discipline responsible for raising and solving metatheoretical problems pertaining to scientific inquiry has been the philosophy of science. Metascientific problems were regarded as the territory of philosophers rather than of practicing scientists. At the birth of modern science, that is to say, in the 17–18th century, philosophers of science aimed at establishing the methods of scientific discovery, i. e., "an orderly, systematic procedure for conducting inquiry that virtually guaranteed the *discovery of abundant new knowledge*" (Nickles, 2000, 86; emphasis added). They did not study and generalize about the successful practice of working scientists: the methodologies they put forward were based on abstract logical-epistemological considerations. Following this, from the 19th century onwards, discovery was not seen as a systematic, rule-governed procedure but as the result of an unpredictable, sudden creative moment. As a consequence, reflection on the process of theory formation was deemed uninteresting and irrelevant by philosophers.

https://doi.org/10.1515/9783110540253-021

2.2 The Standard View of the analytical philosophy of science

In the first half of the 20th century, logical positivism and Karl Popper's falsificationism further narrowed down the meta-scientific analysis of science to the *justification* of the end product of scientific activities, in particular, theories, by the use of logical tools. Given that, besides their well-known differences, these two branches (i. e. logical positivism and falsificationism) also share a series of common background assumptions – above all, their normativity –, they have been subsumed under *the Standard* (or, alternatively: *Received*) *View* of the analytical philosophy of science, henceforth: SV. Thereby, it was assumed that there were generally valid norms of scientific rationality. Theories in the natural sciences were expected to meet these norms and the philosopher's task was to justify theories by evaluating them against these norms. In this way, there was a split between the philosophy of science and actual research practice, which had serious consequences for both sides. Reflection on several methodological aspects of scientific inquiry was understated and expelled from philosophical discourse to textbooks for students of science; and vice versa, scientific inquiry was mostly performed without higher-level meta-theoretical reflection. Philosophy of science (which dealt with the highly abstract epistemological features of science) and scientific research practice (with its unreflected, everyday and specific methodological problems) were disconnected from each other. As a criticism of this relationship between the philosophy of science and the specific sciences, it is often ironically mentioned that "the philosophy of science is about as useful to scientists as ornithology is to birds" (Weinberg, 1987, 433).

2.3 Current trends in the philosophy of science

This situation, however, has dramatically changed since the 1960s. Due to Kuhn's (1962) influential initiative, followed by seminal works by Lakatos (1970a, 1970b), Laudan (1977), Feyerabend (1975) and others, the tenets of SV were rejected as untenable and unrealistic. Instead, the history of science was recognized as the main source of our understanding of the nature of science. This historical turn legitimized the social and psychological aspects of science as interesting and worthy research objects. As a consequence, philosophers of science also engaged with the living, searching, creative, but also fallible and continuously changing *practice* of scientific research, as well as with the process of scientific *discovery*.

Subsequently, current thinking about the nature of science was also shaped by *another turn*. This turn came from a growing need to distinguish between the general philosophy of science and the philosophies of the special sciences (for an instructive and relatively detailed overview of this turn, see Kuhlmann, 2017). As Machamer puts it:

> Philosophers of science could no longer get along without knowing science and/or its history in considerable depth. They, hereafter, would have to work within science as actually practiced, and be able to discourse with *practicing* scientists about what was going on. [. . .] This emphasis on the details of science led various practitioners into doing *the philosophy of the special sciences*. [. . .] One interesting implication of this work in the specialized sciences is that many philosophers have clearly *rejected any form of a science/philosophy dichotomy*, and find it quite congenial to conceive of themselves as, at least in part of their work, "theoretical" scientists. Their goal is to actually make clarifying and, sometimes, substantive changes in the theories and practices of the sciences they study. (Machamer, 2002, 9–11; emphasis added)

As a result, metatheoretical reflection is continuous with object-scientific inquiry in the special sciences. Basically, metatheories are concerned with the foundational problems of object-scientific inquiry which are also tackled by the special sciences themselves. They have the same goals, although their perspective and means may differ. Accordingly, the topics they investigate overlap.

3 Linguistics and the methodology of the natural sciences

The scientific status of linguistics has often been felt to be unsatisfactory in comparison to the standards of the natural sciences. Traditionally, there have been three typical views among linguists.

The first one is, as Simone calls it, "*Saussure's dream*", according to which linguistics does not belong to the natural sciences, and thus it should not apply or imitate their methodologies and meta-theories. Instead, one should

> provide linguistics with an appropriate *method*, one not borrowed more or less mechanically from other sciences, but designed to be peculiarly and strictly of its own. (Simone, 2004, 238; emphasis as in the original)

The second takes the opposite view, since it emphasizes that

> language should be analysed by the methodology of the natural sciences, and there is no room for constraints on linguistic inquiry beyond those typical of all scientific work. (Smith, 2000, vii)

> The methods and results of linguistics, in spite of their modest scope, resemble those of natural sciences, the domain in which science has been most successful. It is only a prospect, but not hopelessly remote, that the study of language may help us toward the understanding and control of human events. (Bloomfield, 1935, 509)

> The statements of grammar [. . .] are not different in principle from the statements of natural science theories; they are factual, in whatever sense statements about valence or visual processing mechanisms are factual and involve truth claims. (Chomsky, 1986, 224)

In syntax, but also in other sub-disciplines of linguistics, the following general requirements have been imposed and have found wide acceptance since the 1950s:

(a) Theory formation (that is, the generation of hypotheses) and the testing of the theory have to be strictly separated.
(b) The hypotheses of empirical linguistic theories have to be connected by valid deductive inferences.
(c) The hypotheses of empirical linguistic theories have to be tested with the help of reliable data that can be regarded as facts constituting a firm and secure basis of research. Such data are called 'evidence'.
(d) Data are immediately given and primary to the theory.

These requirements, however, are not methodological principles guiding research in natural sciences but the tenets of the Standard View of the analytical philosophy of science (SV). This means that linguistics, similarly to many other social sciences, incorrectly identified meta-scientific standards with the norms which SV projected onto the natural sciences:

> The logical positivists, though some of them had studied physics, had little influence on the practice of physics, though their criteria for an ideal science and their models for explanations did have substantial influence on the social sciences as they tried to model themselves on physics, i. e. on 'hard' science. (Machamer, 2002, 12)

Since many linguists accepted SV as a set of unquestionably valid principles of research and believed that it correctly describes the rules they follow every day, in many cases there was a *discord* between the practice of linguistic research and the self-image of linguists.

The third camp agrees with the first standpoint in rejecting SV because it forces abstract, unachievable and unproductive norms on the linguist, but argues that linguistics does not need a comprehensive and uniform meta-theory and methodology because each linguistic field has its very specific routines and conventions:

> There are several reasons to avoid becoming obsessed with any particular set of methodological strictures when evaluating alternative theories. [...] There is a wide gap between general methodological principles on the one hand, and on the other hand the specific procedural, analytical, descriptive, and theoretical decisions one faces in conducting the daily affairs of a particular field. (Langacker, 1987, 33)

Against the background of these traditional stances, it is no wonder that metatheoretical reflection has been misunderstood in linguistics for decades.

However, beginning in the early 21st century, a fourth viewpoint emerged when a series of debates on data and evidence in linguistics brought a turning point (cf. Kepser and Reis (eds.), 2005, Borsley (ed.), 2005, Lehmann, 2004, Sternefeld (ed.), 2007, Penke and Rosenbach (eds.), 2007, Featherston and Winkler (eds.), 2009, Winkler and

Featherston (eds.), 2009, Kertész and Rákosi, 2012 etc.). It was recognized that SV is not only unrealistic but also thematically considerably impoverishes meta-scientific reflection, because it narrows down its scope to the subsequent reconstruction and analysis of the structure of theories, and does not touch upon topics which were diagnosed as urgent problems of linguistic theorizing. Therefore, linguists started to elaborate on the revision of different elements of SV, as well as on establishing guidelines on the issues they faced during their everyday research practice. The majority of these attempts were made without an elaborated meta-theoretical background, from a perspective very close to the object-scientific problems themselves.

Their most important insights pertain to the acknowledgement of the pluralism and uncertainty of data:

- The combination of several data types within a given piece of research is considered to be preferable to the application of a single data type.
- All data types are assumed to be uncertain and problematic; therefore, it seems indispensable to clarify their structure, complexity, directness, abstractness, and function.
- The relationship between data and theory is assumed to be cyclic rather than linear.
- Linguistic data are considered to be unavoidably theory- and problem-dependent.
- Linguistic data may also generate contradictions. This is no longer regarded as a failure in every case, but a constitutive property of theorizing which inspires researchers to a continuous revision of their findings, the frameworks they apply, and the empirical basis of their research.

4 Metatheoretical reflection and object-theoretical research in syntax

At the same time, the need for a more radical strategy has also been recognized:

> There is obscurity and controversy not only over the problems for which we need theories of language but also over the status of the theories themselves. This issue of status is highly abstract: it requires a theory of theories of language, a 'meta-theory'. It would be nice to ignore the meta-theory and get on with the theory, but that is a luxury we cannot afford. We think that *many mistakes in the theory of language arise from a mistaken meta-theory*. Further, we think that *these mistakes are often facilitated by a failure to be explicit about the meta-theory*: once the implicit meta-theory is exposed, it can be seen to be implausible and unsupportable. (Devitt and Sterelny, 1999, 9; emphasis added)

Consequently, unlike SV, there is no single, generally valid metatheory that all kinds of scientific theories should take for granted. Rather, as Part II of the present book illustrates, there is a great number of different meta-theoretical approaches to linguistics – *the meta-theoretical level is just as varied as the object-scientific level*. Most of them fit into the current tendencies in the philosophy of science mentioned in Section 2.3.

Within this metatheoretical pluralism, the present volume advocates a perspective according to which the scope of syntactic theorizing can be broadened: by blurring the rigid distinction between pure object theoretical research and metatheories, one obtains *metatheoretically reflected object theoretical research*. Extending the scope of syntax in this way serves the solution of *the same* object-scientific problems which working linguists tackle, although from a different point of view and perhaps more deeply, less technically and in a more abstract way.

5 Some possible uses of the four parameters

As explained in the Introduction to this handbook, the three editors have suggested four parameters that may assist us in any comparison of syntactic theories. Their application in Part I, as well as the presentation of meta-linguistic approaches in Part II, invite the reader to perform *meta*-theoretical reflection in order to gain new insights into the syntactic structure *of human language in general and the individual languages in particular*. Such metatheoretical reflection may question features of research practice which seem to be taken for granted, unproblematic and natural, and prompt us to look actively for revisions and new possibilities. For example, the comparison of syntactic theories along the four parameters discussed in the Introduction may be useful in the following respects:

– **Data.** There are several syntactic theories which apply predominantly introspective data even today. They may rule out the use of other data types for principled reasons or enable them in theory, but leave them unutilized in practice. Other approaches make use of a wide range of data types. This finding raises a series of questions which cannot be answered easily but require further investigations:
  Which group of theories fares better: those which make use of only one data type or those which rely on multiple data sources? What are the advantages and dangers of the use of multiple data types? How can different data types be combined with each other? Are there linguistic phenomena which can be investigated only with the help of certain data types, and, conversely: Are certain data types suitable only for the investigation of specific phenomena? Can research based on experimental or corpus data revise results obtained with the help of introspective data, or vice versa? Do certain data types provide stronger support for the hypotheses of the theory than others? Do certain data types involve problems which cannot be avoided and undermine the reliability of the hypotheses they provide support for? Or, rather, can the use of multiple data types counter-balance the shortcomings of single data types?

- **Goals**. Rival theories can set similar goals but try to attain them by using different tools. The use of different means may lead to conflicting results and deepen the rivalry between the theories; but it might also produce similar outcomes, providing converging evidence for the hypotheses at issue. It may also happen that two theories differ considerably in their goals. In such cases, their results may be complementary, which might raise the idea of uniting these two approaches to a more complex theory. One may also ask whether an approach can be extended by new goals, for example, by applying it to new areas, or striving for deeper explanations, etc. More generally, the question arises as to which goals might be achievable when we are in possession of a set of basic assumptions, tools, data, etc. Are there theories which seem to be more comprehensive in the sense that they set a wider range of goals than others? Is there a ranking of goals or are all goals equal and should be pursued simultaneously?
- **Tools**. A comparison of the tools applied may be beneficial, too. The conceptual system or the categories applied by different theories may seem to be different but on closer inspection, they might be convertible into each other. It might be fruitful to study the circumstances under which tools from rival theories can be borrowed and built into one's own approach. Of course, the opposite may happen, too: tools which seemed to be similar may be found to be rather different, leading to contradictory results. Are there theories which can be deemed more sophisticated in the sense that they possess a richer inventory of tools? Or conversely, are there syntactic approaches which pursue the same goals effectively with the help of a smaller set of tools?
- **Evaluation**. Different syntactic theories may rely on different criteria of evaluation. The systematic comparison and the clarification of the differences between the criteria used could be a very important step towards a fruitful dialogue among rival approaches. Not only can the applied criteria differ between two theories but their ranking can be different, too. Related to this, one may also ask whether and how an approach is willing to tolerate the infringement of a criterion in order to fulfill another, more important one. For example, can the requirement for simplicity be sacrificed to retain consistency? Are there cases in which inconsistency has to be tolerated in order to achieve a greater degree of explanatory adequacy?

6 Final remark

As we have outlined above, and speaking in general terms, the relationship between metatheoretical reflection and syntactic theorizing seems to have been shaped by four factors:

- The historical turn in the philosophy of science, which also meant that the practice of scientific inquiry could be subject to metatheoretical investigation.

- The focus on the philosophies of the special sciences.
- The data/evidence debate in linguistics, which led to the acknowledgement of the pluralism of data types and data sources.
- The pluralism of linguistic metatheories.

These factors seem to legitimize a *meta*theoretical perspective whose aim is to improve syntactic theorizing by tackling its foundational problems and which, in this way, serves the enrichment of our knowledge of the syntactic structure of language and languages in cooperation with *object-theoretical* syntactic research itself. The comparison of syntactic theories may contribute to the realization of such a perspective.

References

Bloomfield, Leonard. 1935. *Language*. London: George Allen and Unwin.
Borsley, Robert D. (ed.). 2005. Data in theoretical linguistics. [Special issue]. *Lingua* 115(11).
Chomsky, Noam. 1986. *Knowledge of language: Its nature, origin, and use*. New York: Praeger.
Devitt, Michael & Kim Sterelny. 1999. *Language and reality. An introduction to the philosophy of language*. Oxford: Blackwell.
Featherston, Sam & Susanne Winkler (eds.). 2009. *The Fruits of Empirical Linguistics. Vol. 1: Process*. Berlin & New York: de Gruyter.
Feyerabend, Paul K. 1975. *Against Method*. London: Verso.
Kertész, András & Csilla Rákosi. 2012. *Data and evidence in linguistics. A plausible argumentation model*. Cambridge: Cambridge University Press.
Kepser, Stephan & Marga Reis (eds.). 2005. *Linguistic evidence. Empirical, theoretical and computational perspectives*. Berlin & New York: de Gruyter.
Kuhlmann, Meinhard. 2017. Allgemeine Wissenschaftsphilosophie und die Philosophien der Einzelwissenschaften. In Simon Lohse & Thomas Reydon (eds.), *Grundriss Wissenschaftsphilosophie. Philosophien der Einzelwissenschaften*, 17–50. Hamburg: Meiner.
Kuhn, Thomas S. 1962. *The Structure of Scientific Revolutions*. Chicago: The University of Chicago Press.
Lakatos, Imre. 1970a. History of science and its rational reconstructions. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1970. 91–136.
Lakatos, Imre. 1970b. Falsification and the methodology of scientific research programmes. In Imre Lakatos & Alan Musgrave (eds.), *Criticism and the Growth of Knowledge*, 91–195. Cambridge: Cambridge University Press.
Langacker, Ronald W. 1987. *Foundations of cognitive grammar. Vol. 1*, Stanford: Stanford University Press.
Laudan, Larry. 1977. *Progress and Its Problems*. Berkeley & Los Angeles: University of California Press.
Lehmann, Christian. 2004. Data in linguistics. *The Linguistic Review* 21. 175–210.
Machamer, Peter. 2002. A brief historical introduction to the philosophy of science. In Peter Machamer & M. Silberstein (eds.), *The Blackwell guide to the philosophy of science*, 1–17. Malden & Oxford: Blackwell.
Nickles, Thomas. 2000. Discovery. In William H. Newton-Smith (ed.), *A companion to the philosophy of science*, 85–96. Malden & Oxford: Blackwell.

Penke, Martina & Anette Rosenbach (eds.). 2007. *What counts as evidence in linguistics?* Amsterdam & Philadelphia: Benjamins.
Simone, Raffaele. 2004. The object, the method, and the ghosts. Remarks on a terra incognita. *The Linguistic Review* 21. 235–256.
Smith, Neil. 2000. Foreword. In Noam Chomsky (ed.), *New horizons in the study of language and mind*, vi–xvi. Cambridge: Cambridge University Press.
Sternefeld, Wolfgang (ed.). 2007. Data in generative grammar [Special issue]. *Theoretical Linguistics* 33. 269–413.
Weinberg, Steven. 1987. Newtonianism, reductionism and the art of congressional testimony. *Nature* 330. 433–437.
Winkler, Susanne & Sam Featherston (eds.). 2009. *The Fruits of empirical linguistics. Vol. 2: Product*. Berlin & New York: de Gruyter.

Author Index

Language Index

Subject Index