

Grammatical theory

From transformational grammar to
constraint-based approaches

Open Review Part 1

Stefan Müller

Textbooks in Language Sciences 1



Textbooks in Language Sciences

Editors: Stefan Müller, Martin Haspelmath

Editorial Board: Claude Hagège, Marianne Mithun, Anatol Stefanowitsch, Foong Ha Yap

In this series:

1. Müller, Stefan. Grammatical Theory: From Transformational grammar to constraint-based approaches.
2. Schäfer, Roland. Einführung in die grammatische Beschreibung des Deutschen.

Grammatical theory

From transformational grammar to
constraint-based approaches

Stefan Müller

Stefan Müller. 2015. *Grammatical theory: From transformational grammar to constraint-based approaches* (Textbooks in Language Sciences 1). Berlin: Language Science Press.

This title can be downloaded at:

<http://langsci-press.org/catalog/book/14>

© 2015, Stefan Müller

Published under the Creative Commons Attribution-NoDerivatives 4.0 Licence (CC BY-ND 4.0): <http://creativecommons.org/licenses/by-nd/4.0/>

ISBN: 978-3-944675-21-3

Cover and concept of design: Ulrike Harbort

Translator: Andrew Murphy

Typesetting: Stefan Müller

Proofreading: Viola Auermann

Fonts: Linux Libertine, Arimo

Typesetting software: \LaTeX

Language Science Press

Habelschwerdter Allee 45

14195 Berlin, Germany

langsci-press.org

Storage and cataloguing done by FU Berlin

Freie Universität



Berlin

Language Science Press has no responsibility for the persistence or accuracy of URLs for external or third-party Internet websites referred to in this publication, and does not guarantee that any content on such websites is, or will remain, accurate or appropriate. Information regarding prices, travel timetables and other factual information given in this work are correct at the time of first publication but Language Science Press does not guarantee the accuracy of such information thereafter.

For Max

Preface

This book briefly introduces various grammatical theories that play a role in current theorizing or have made contributions in the past which are still relevant today. I explain some foundational assumptions and then apply the respective theories to what can be called the “core grammar” of German. I have decided to stick to the object language that I used in the German version of this book since many of the phenomena that will be dealt with cannot be explained with English as the object language. Furthermore, many theories have been developed by researchers with English as their native language and it is illuminative to see these theories applied to another language. I show how the theories under consideration deal with arguments and adjuncts, active/passive alternations, local reorderings (so-called scrambling), verb position, and fronting of phrases over larger distances (the verb second property of the Germanic languages without English).

A final chapter deals with foundational questions that are important for developing theories. This includes a discussion of the question of whether we have innate domain specific knowledge about language (UG), the discussion of psycholinguistic evidence concerning the processing of language by humans, a discussion of the status of empty elements and of the question whether we construct and perceive utterances holistically or rather compositionally, that is, whether we use phrasal or lexical constructions.

There is an unbelievable terminological chaos in linguistics. I therefore wrote an introductory chapter that introduces terminology in the way it is used later on in the book. The second chapter introduces phrase structure grammars, which plays a role for many of the theories that are covered in this book. I use these two chapters (excluding the Section 2.3 on interleaving phrase structure grammars and semantics) in introductory courses of our BA curriculum for German studies. Advanced readers may skip these introductory chapters. The following chapters are structured in a way that should make it possible to understand the introduction of the theories without any prior knowledge. The sections regarding new developments and classification are more ambitious: They refer to chapters still to come and also point to other publications that are relevant in the current theoretical discussion but cannot be repeated or summarized in this book. These parts of the book address advanced students and researchers. I use this book for teaching the syntactic aspects of the theories in a seminar for advanced students in our BA. The slides are available on my web page.

This book only deals with relatively recent developments. For a historical overview, see for instance Jungen & Lohnstein (2006). I am aware of the fact that chapters on Integrational Linguistics (Lieb 1983; Eisenberg 2004; Nolda 2007), Optimality Theory (Prince & Smolensky 1993, Grimshaw 1997, G. Müller 2000a), Role and Reference Grammar (Van Valin 1993) and Relational Grammar (Perlmutter 1983; 1984) are missing. I will leave these

theories for later editions.

The original book was planned to have 400 pages. This book now has 789 pages. I tried to represent the chosen theories appropriately and to cite all important work. Although the list of references is over 75 pages long, I was probably not successful. I apologize for this and any other shortcomings.

Acknowledgments

I would like to thank Jason Baldridge, Felix Bildhauer, Stefan Evert, Gisbert Fanselow, Sandiway Fong, Hans-Martin Gärtner, Kim Gerdes, Jakob Maché, Guido Mensching, Laura Michaelis, Geoffrey Pullum, Uli Sauerland, Roland Schäfer, Jan Strunk, Remi van Trijp, and Shravan Vasishth for discussion and Monika Budde, Philippa Cook, Laura Kallmeyer, Thomas Groß, Tibor Kiss, Gisela Klann-Delius, Jonas Kuhn, Timm Lichte, Anke Lüdeling, Jens Michaelis, Bjarne Ørsnes, Tim Osborne, Andreas Pankau, Christian Pietsch, Frank Richter, Ivan Sag, and Eva Wittenberg for comments on earlier versions of the German edition of this book and Martin Haspelmath and Sylvain Kahane for comments on earlier versions of this book. Thanks to Leonardo Boiko for pointing out typos.

I thank Wolfgang Sternefeld and Frank Richter, who wrote a detailed review of the German version of this book (Sternefeld & Richter 2012). They pointed out some mistakes and omissions that were corrected in the second edition of the German book and which are of course not present in the English version.

Thanks to all the students who commented on the book and whose questions lead to improvements. Lisa Deringer, Aleksandra Gabryszak, Simon Lohmiller, Theresa Kallenbach, Steffen Neuschulz, Reka Meszaros-Segner, Lena Terhart and Elodie Winckel deserve special mention.

Since this book is built upon all my experience in the area of grammatical theory, I want to thank all those with whom I ever discussed linguistics during and after talks at conferences, workshops, summer schools or via email. Werner Abraham, John Bate-man, Dorothee Beermann, Rens Bod, Miriam Butt, Manfred Bierwisch, Ann Copestake, Holger Diessel, Kerstin Fischer, Dan Flickinger, Peter Gallmann, Adele Goldberg, Petter Haugereid, Lars Hellan, Paul Kay, Wolfgang Klein, Hans-Ulrich Krieger, Emily M. Bender, Andrew McIntyre, Detmar Meurers, Laura Michaelis, Martin Neef, Manfred Sailer, Anatol Stefanowitsch, Peter Svenonius, Michael Tomasello, Hans Uszkoreit, Gert Webelhuth, Stephen Wechsler, Daniel Wiechmann and Arne Zeschel deserve special mention.

I thank Sebastian Nordhoff for a comment regarding the completion of the subject index entry for *recursion*.

During the past years, a number of workshops on theory comparison have taken place. I was invited to three of them. I thank Helge Dyvik and Torbjørn Nordgård for inviting me to the fall school for Norwegian PhD students *Languages and Theories in Contrast*, which took place 2005 in Bergen. Guido Mensching and Elisabeth Stark invited me to the workshop *Comparing Languages and Comparing Theories: Generative Grammar and Construction Grammar*, which took place in 2007 at the Freie Universität Berlin and Andreas Pankau invited me to the workshop *Comparing Frameworks* in 2009 in Utrecht. I

really enjoyed the discussion with all participants of these events and this book benefited enormously from the interchange.

I thank Peter Gallmann for the discussion of his lecture notes on GB during my time in Jena. The Sections 3.1.3–3.4 have a structure that is similar to the one of his script and take over a lot. Thanks to David Reitter for the \LaTeX macros for Combinatorial Categorical Grammar, to Mary Dalrymple and Jonas Kuhn for the LFG macros and example structures, and to Laura Kallmeyer for the \LaTeX sources of most of the TAG analyses. Most of the trees have been adapted to the `forest` package because of compatibility issues with \LaTeX , but the original trees and texts were a great source of inspiration and without them the figures in the respective chapters would not be half as pretty as they are now.

I thank Sašo Živanović for implementing the \LaTeX package `forest`. It really simplifies typesetting of trees, dependency graphs, and type hierarchies. I also thank him for individual help via email and on stackexchange. In general, those active on stackexchange could not be thanked enough: Most of my questions regarding specific details of the typesetting of this book or the implementation of the \LaTeX classes that are used by Language Science Press now have been answered within several minutes. Thank you!

Viola Auermann and Antje Bahlke, Lea Helmers, and Chiara Jancke cannot be thanked enough for their work at the copy machines. Viola also helped a lot with proof reading prefinal stages of the translation. I also want to thank my (former) lab members Felix Bildhauer, Philippa Cook, Janna Lipenkova, Jakob Maché, Bjarne Ørsnes and Roland Schäfer, which were mentioned above already for other reasons, for their help with teaching. During the years from 2007 until the publication of the first German edition of this book two of the three tenured positions in German Linguistics were unfilled and I would have not been able to maintain the teaching requirements without their help and would have never finished the Grammatiktheorie book.

I thank Tibor Kiss for advice in questions of style. His diplomatic kind always was a shining example for me and I hope that this is also reflected in this book.

On the way this book is published

I started to work on my dissertation in 1994 and defended it in 1997. During the whole time the manuscript was available on my web page. After the defense, I had to look for a publisher. I was quite happy to be accepted to the series *Linguistische Arbeiten* by Niemeyer, but at the same time I was shocked about the price, which was 186.00 DM for a paper back book that was written and typeset by me without any help by the publisher (twenty times the price of a paperback novel).¹ This basically meant that my book was depublished: Until 1998 it was available from my web page and after this it was available in libraries only. My Habilitationsschrift was published by CSLI Publications for a much more reasonable price. When I started writing text books, I was looking for alternative

¹ As a side remark: In the meantime Niemeyer was bought by De Gruyter and closed down. The price of the book is now 139.95 €/ \$ 196.00. The price in Euro corresponds to 273.72 DM.

distribution channels and started to negotiate with no-name print on demand publishers. Ms. Narr who runs the Stauffenburg publishing house convinced me to publish my HPSG text book with her. The copyrights for the German version of the book remained with me so that I could publish it on my web page. The collaboration was successful so that I also published my second text book about grammatical theory with Stauffenburg. I think that this book has a broader relevance and should be accessible for non-German-speaking readers as well. I therefore decided to have it translated into English. Since Stauffenburg is focused on books in German, I had to look for another publisher. Fortunately the situation in the publishing sector changed quite dramatically in comparison to 1997: We now have high profile publishers with strict peer review that are entirely open access. I am very glad about the fact that Ms. Narr sold the English rights of my book back to me and that I can now publish the English version with Language Science Press.

Berlin, May 20, 2015

Stefan Müller

Contents

I	Background and specific theories	1
1	Introduction and basic terms	3
1.1	Why do syntax?	3
1.2	Why do it formally?	6
1.3	Constituents	6
1.3.1	Constituency tests	7
1.3.2	Some comments on the status of constituent tests	11
1.4	Parts of speech	17
1.5	Heads	29
1.6	Arguments and adjuncts	31
1.7	Grammatical functions	36
1.7.1	Subjects	36
1.7.2	Adverbials	42
1.7.3	Predicatives	43
1.7.4	Valency classes	44
1.8	A topological model of the German clause	45
1.8.1	The position of the verb	45
1.8.2	The sentence bracket, prefield, middle field and postfield	46
1.8.3	Assigning elements to fields	46
1.8.4	Recursion	51
2	Phrase Structure Grammar	55
2.1	Symbols and rewrite rules	55
2.2	Expanding PSG with features	61
2.3	Semantics	62
2.4	Phase structure rules for some aspects of German syntax	65
2.4.1	Noun phrases	65
2.4.2	Prepositional phrases	74
2.5	\bar{X} Theory	76
3	Transformational grammar – Government & Binding	85
3.1	General remarks on the representational format	86
3.1.1	Transformations	86
3.1.2	The hypothesis of language acquisition: Principles & Parameters	88

Contents

3.1.3	The T-Model	89
3.1.4	Phonetic Form	91
3.1.5	Logical Form	91
3.1.6	The Lexicon	93
3.1.7	\bar{X} Theory	95
3.1.8	CP and IP in English	98
3.1.9	The structure of the German clause	103
3.2	Verb position	104
3.3	Long-distance dependencies	107
3.4	Passive	111
3.4.1	Structural and lexical case	111
3.4.2	Case assignment and the Case Filter	112
3.5	Local reordering	116
3.6	Summary and classification	118
3.6.1	Explaining language acquisition	119
3.6.2	Formalization	119
4	Transformational Grammar – Minimalism	127
4.1	General remarks on the representational format	127
4.1.1	Basic architecture	128
4.1.2	Valence, feature checking, and agreement	129
4.1.3	Phrase structure and \bar{X} Theory	131
4.1.4	Little v	132
4.1.5	CP, TP, v P, VP	134
4.1.6	Adjuncts	138
4.2	Verb position	138
4.3	Long-distance dependencies	138
4.4	Passive	139
4.5	Local reordering	142
4.6	New developments and theoretical variants	142
4.6.1	Move, Merge, feature-driven movement and functional projections	144
4.6.2	Labeling	152
4.6.3	Specifiers, complements, and the remains of \bar{X} Theory	158
4.6.4	Minimalism, Categorical Grammar, and HPSG	162
4.6.5	Selection of atomic features vs. selection of complex categories	171
4.6.6	Summary	172
4.7	Summary and classification	173
4.7.1	Explaining language acquisition	173
4.7.2	Formalization	173
5	Generalized Phrase Structure Grammar	179
5.1	General remarks on the representational format	179
5.1.1	Complex categories, the Head Feature Convention, and \bar{X} rules	180

5.1.2	Local reordering	182
5.1.3	metarules	184
5.1.4	Semantics	186
5.1.5	Adjuncts	187
5.2	Passive as a metarule	189
5.3	Verb position	192
5.4	Long-distance dependencies as the result of local dependencies	193
5.5	Summary and classification	196
6	Feature Descriptions	201
6.1	Feature Descriptions	201
6.2	Types	204
6.3	Disjunction	207
6.4	Structure sharing	207
6.5	Cyclic structures	209
6.6	Unification	210
6.7	Phenomena, models and formal theories	212
7	Lexical Functional Grammar	217
7.1	General remarks on the representational format	219
7.1.1	Functional structure	219
7.1.2	Completeness	222
7.1.3	Coherence	222
7.1.4	Restrictions on the c-structure/f-structure relation	223
7.1.5	Semantics	224
7.1.6	Adjuncts	226
7.2	Passive	227
7.3	Verb position	230
7.4	Local reordering	231
7.5	Long-distance dependencies and functional uncertainty	233
7.6	Summary and classification	237
8	Categorial Grammar	241
8.1	General remarks on the representational format	242
8.1.1	Representation of valency information	242
8.1.2	Semantics	243
8.1.3	Adjuncts	245
8.2	Passive	246
8.3	Verb position	247
8.4	Local reordering	249
8.5	Long-distance dependencies	250
8.5.1	Type Raising	250
8.5.2	Forward and backward composition	250
8.5.3	Analysis of long-distance dependencies	251

8.6	Summary and classification	253
9	Head-Driven Phrase Structure Grammar	259
9.1	General remarks on the representational format	261
9.1.1	Representation of valency information	262
9.1.2	Representation of constituent structure	264
9.1.3	Linearization rules	268
9.1.4	Projection of head properties	269
9.1.5	Inheritance hierarchies and generalizations	270
9.1.6	Semantics	273
9.1.7	Adjuncts	276
9.2	Passive	279
9.2.1	Valency information and the Case Principle	282
9.3	Verb position	284
9.4	Local reordering	289
9.5	Long-distance dependencies	292
9.6	New developments and theoretical variants	296
9.7	Summary and classification	298
10	Construction Grammar	305
10.1	General remarks on the representational format	307
10.1.1	The head-complement construction	307
10.1.2	Representation of valency information	309
10.1.3	Semantics	309
10.1.4	Adjuncts	309
10.2	Passive	309
10.3	Verb position	314
10.4	Local reordering	316
10.5	Long-distance dependencies	317
10.6	New developments and theoretical variants	317
10.6.1	Berkeley Construction Grammar	317
10.6.2	Sign-Based Construction Grammar	320
10.6.3	Embodied Construction Grammar	325
10.6.4	Fluid Construction Grammar	329
10.7	Summary and classification	348
11	Dependency Grammar	351
11.1	General remarks on the representational format	353
11.1.1	Valency information, nucleus and satellites	353
11.1.2	Adjuncts	355
11.1.3	Linearization	356
11.1.4	Semantics	357
11.2	Passive	358
11.3	Verb position	360

11.4	Local reordering	361
11.5	Long-distance dependencies	362
11.6	New developments and theoretical variants	365
11.6.1	Tesnière's part of speech classification	366
11.6.2	Connexion, junction, and transfer	367
11.6.3	Scope	373
11.7	Summary and classification	373
11.7.1	Linearization	374
11.7.2	Dependency Grammar vs. Phrase Structure Grammar	381
12	Tree Adjoining Grammar	395
12.1	General remarks on representational format	396
12.1.1	Representation of valency information	396
12.1.2	Substitution	396
12.1.3	Adjunction	396
12.1.4	Semantics	398
12.2	Local reordering	400
12.3	Verb position	403
12.4	Passive	405
12.5	Long-distance dependencies	406
12.6	New developments and theoretical variants	406
12.6.1	FTAG	406
12.6.2	V-TAG	410
12.6.3	The competence-performance distinction and the generative capacity of tree-local MC-LTAG	412
12.7	Summary and classification	417
II	General discussion	421
13	The innateness of linguistic knowledge	423
13.1	Syntactic universals	425
13.1.1	Head Directionality Parameter	425
13.1.2	\bar{X} structures	427
13.1.3	Grammatical functions such as subject and object	429
13.1.4	Binding principles	432
13.1.5	Properties of long-distance dependencies	433
13.1.6	Grammatical morphemes for tense, mood and aspect	441
13.1.7	Parts of speech	441
13.1.8	Recursion and infinitude	443
13.1.9	Summary	448
13.2	Speed of language acquisition	448
13.3	Critical period for acquisition	449

Contents

13.4	Lack of acquisition among non-human primates	450
13.5	Creole and sign languages	451
13.6	Localization in special parts of the brain	452
13.7	Differences between language and general cognition	453
13.7.1	Williams Syndrome	454
13.7.2	KE family with FoxP2 mutation	454
13.8	Poverty of the Stimulus	455
13.8.1	Gold's Theorem	456
13.8.2	Four case studies	459
13.8.3	Unsupervised Data-Oriented Parsing (U-DOP)	470
13.8.4	Negative evidence	476
13.9	Summary	478
14	Generative-enumerative vs. model-theoretic approaches	481
14.1	Graded acceptability	483
14.2	Utterance fragments	484
14.3	A problem for model-theoretic approaches?	486
15	The competence/performance distinction	489
15.1	The derivational theory of complexity	491
15.2	Incremental processing	494
16	Language acquisition	501
16.1	Principles and Parameters	501
16.2	Principles and the lexicon	509
16.3	Pattern-based approaches	510
16.4	Selection-based approaches	516
16.5	Summary	517
17	Binary branching	519
18	Generative capacity and grammatical formalisms	523
19	Locality	527
20	Recursion	537
21	Empty elements	539
21.1	Views on empty elements	539
21.2	Eliminating empty elements from grammars	541
21.3	Empty elements and semantic interpretation	547
21.4	Evidence for empty elements	553
21.5	Transformations, lexical rules, and empty elements	553
22	Extraction, scrambling, and passive: one or several descriptive devices?	559

23 Phrasal vs. lexical analyses	569
23.1 Some putative advantages of phrasal models	570
23.1.1 Usage-based theories	570
23.1.2 Coercion	572
23.1.3 Aspect as a clause level phenomenon	574
23.1.4 Simplicity and polysemy	574
23.2 Evidence for lexical approaches	577
23.2.1 Valence and coordination	577
23.2.2 Valence and derivational morphology	581
23.3 Radical underspecification: The end of argument structure?	587
23.3.1 Neo-Davidsonianism	587
23.3.2 Little <i>v</i> and idiom asymmetries	589
23.3.3 Deverbal nominals	591
23.3.4 Idiosyncratic syntactic selections	594
23.3.5 Expletives	596
23.3.6 An exoskeletal approach	597
23.3.7 Is there an alternative to lexical valence structure?	601
23.3.8 Summary	602
23.4 Relations between constructions	603
23.4.1 Inheritance hierarchies for constructions	603
23.4.2 Mappings between different levels of representations	607
23.4.3 Is there an alternative to lexical rules?	610
23.5 Further problems for phrasal approaches	610
23.5.1 Particle verbs and commitment to phrase structure configurations	610
23.6 Arguments from language acquisition	611
23.6.1 Recognizability of constructions	611
23.6.2 Coordination and discontinuousness	613
23.7 Arguments from psycho- and neurolinguistics	618
23.7.1 Lexical rules vs. phrasal constructions	618
23.7.2 Light verbs	620
23.7.3 Arguments from Neurolinguistics	621
23.8 Arguments from statistical distribution	625
23.8.1 Unsupervised Data-Oriented Parsing	625
23.8.2 Collostructions	628
23.9 Conclusion	628
23.10 Why (phrasal) constructions?	629
23.10.1 Verbless directives	629
23.10.2 Serial verbs	637
23.10.3 Relative and interrogative clauses	638
23.10.4 The NPN construction	640
24 Universal Grammar and comparative linguistics without UG	643
24.1 Formal tools for capturing generalizations	643

Contents

24.2	How to arrive linguistic theories that capture cross-linguistic generalizations	647
25	Conclusion	655
A	Solutions to the exercises	659
A.1	Introduction and basic terms	659
A.2	Phrase structure grammars	659
A.3	Transformational Grammar – Government & Binding	663
A.4	Generalized Phrase Structure Grammar	665
A.5	Feature descriptions	667
A.6	Lexical Functional Grammar	668
A.7	Categorial Grammar	669
A.8	Head-Driven Phrase Structure Grammar	670
A.9	Construction Grammar	672
A.10	Dependency Grammar	673
A.11	Tree Adjoining Grammar	674
	List of references	677
	Bibliography	677
	Index	765
	Name index	765
	Language index	778
	Subject index	780

Part I

Background and specific theories

1 Introduction and basic terms

The aim of this chapter is to explain why we actually study syntax (Section 1.1) and why it is important to formalize our findings (Section 1.2). Some basic terminology will be introduced in Sections 1.3–1.8: Section 1.3 deals with criteria for dividing up utterances into smaller units. Section 1.4 shows how words can be grouped into classes, that is I will introduce criteria for assigning words to categories such as verb or adjective. Section 1.5 introduces the notion of heads, in Section 1.6 the distinction between arguments and adjuncts is explained, Section 1.7 defines grammatical functions and Section 1.8 introduces the notion of topological fields which can be used to characterize certain areas of the clause in languages such as German.

Unfortunately, linguistics is a scientific field with a considerable amount of terminological chaos. This is partly due to the fact that terminology originally defined for certain languages (e. g. Latin, English) was later simply adopted for the description of other languages as well. However, this is not always appropriate since languages differ from one another greatly and are constantly changing. Due to the problems this caused, the terminology started to be used differently or new terms were invented.

When new terms are introduced in this book, I will always mention related terminology or differing uses of each term so that readers can relate this to other literature.

1.1 Why do syntax?

Every linguistic expression we utter has a meaning. We are therefore dealing with what have been referred to as pairs of form and meaning (de Saussure 1916b). A word such as *tree* in its specific orthographical form or in its corresponding phonetic form is assigned the meaning *tree'*. Larger linguistic units can be built up out of smaller ones: words can be joined together to form phrases and these in turn can form sentences.

The question which now arises is the following: Do we need a formal system which can assign a structure to these sentences? Would it not be sufficient to formulate a pairing of form and meaning for complete sentences just as we did for the word *tree* above?

That would, in principle, be possible if a language were just a finite list of word sequences. If we were to assume that there is a maximum length for sentences and a maximum length for words and thus that there can only be a finite number of words, then the number of possible sentences would indeed be finite. However, even if we were to restrict the possible length of a sentence, the number of possible sentences would still be enormous. The question we would then really need to answer is: What is the maximum length of a sentence? For instance, it is possible to extend all the sentences in (1):

1 Introduction and basic terms

- (1) a. This sentence goes on and on and on and on ...
- b. [A sentence is a sentence] is a sentence.
- c. that Max thinks that Julia knows that Otto claims that Karl suspects that Richard confirms that Friederike is laughing

In (1b), something is being said about the group of words *a sentence is a sentence*, namely that it is a sentence. One can, of course, claim the same for the whole sentence in (1b) and extend the sentence once again with *is a sentence*. The sentence in (1c) has been formed by combining *that Friederike is laughing* with *that, Richard* and *confirms*. The result of this combination is a new sentence *that Richard confirms, that Friederike is laughing*. In the same way, this has then been extended with *that, Karl* and *suspects*. One thus obtains a very complex sentence which embeds a less complex sentence. This partial sentence in turn contains a further partial sentence and so on. (1c) is similar to those sets of Russian nesting dolls, also called *matryoshka*: Each doll contains a smaller doll which can be painted differently from the one that contains it. In just the same way the sentence in (1c) contains parts which are similar to it but which are shorter and which involve different nouns and verbs. This can be made clearer by using brackets in the following way:

- (2) that Max thinks [that Julia knows [that Otto claims [that Karl suspects [that Richard confirms [that Friederike is laughing]]]]]

We can build incredibly long and complicated sentences by applying extensions like the one in (1).¹

It would be arbitrary to establish some cut-off point up to which such combinations can be considered to belong to our language (Harris 1957: 208; Chomsky 1957: 23). It is also implausible to claim that such complex sentences are stored in our brains as a single complex unit. While evidence from psycholinguistic experiments shows that highly frequent or idiomatic combinations are stored as complex units, this could not be the case for sentences such as those in (1). Furthermore, we are capable of producing utterances that we have never heard before and which have also never been uttered or written down previously. Therefore, these utterances must have some kind of structure, there must be patterns which occur again and again. As humans, we are able to build such complex structures out of simpler ones and, vice-versa, to break down complex utterances into their component parts. Evidence for humans' ability to make use of rules for combining words into larger units has now also been provided by research in neuroscience (Pulvermüller 2010: 170).

¹ It is sometimes claimed that we are capable of constructing infinitely long sentences (Nowak, Komarova & Niyogi 2001: 117; Kim & Sells 2008: 3; Dan Everett in O'Neill & Wood (2012) at 25:19) or that Chomsky made such claims (Leiss 2003: 341). This is, however, not correct since every sentence has to come to an end at some point. Even in the theory of formal languages developed in the Chomskyan tradition there are also no infinitely long sentences. Rather, certain formal grammars can describe a set containing infinitely many finite sentences (Chomsky 1957: 13). See also Pullum & Scholz (2010) and Section 13.1.8 on the issue of recursion in grammar and for claims about the infinite nature of language.

It becomes particularly evident that we combine linguistic material in a rule-governed way when these rules are violated. Children acquire linguistic rules by generalizing from the input available to them. In doing so, they produce some utterances which they could not have ever heard previously:

- (3) Ich festhalte die. (Friederike, 2;6)
 I PRT.hold them
 Intended: 'I hold them tight.'

Friederike, learning German, is at the stage of acquiring the rule for the position of the finite verb (namely, second position). What she does here, however, is to place the whole verb, including a separable particle *fest* 'tight', in the second position although the particle should be realized at the end of the clause (*Ich halte die fest.*).

If we do not wish to assume that language is merely a list of pairings of form and meaning, then there must be some process whereby the meaning of complex utterances can be obtained from the meanings of the smaller components of those utterances. Syntax reveals something about the way in which the words involved can be combined, something about the structure of an utterance. For instance, knowledge about subject-verb agreement helps with the interpretation of the following sentences in German:

- (4) a. Die Frau schläft.
 the woman sleep.3SG
 'The woman sleeps.'
 b. Die Mädchen schlafen.
 the girls sleep.3PL
 'The girls sleep.'
 c. Die Frau kennt die Mädchen.
 the woman know.3SG the girls
 'The woman knows the girls.'
 d. Die Frau kennen die Mädchen.
 the woman know.3PL the girls
 'The girls know the woman.'

The sentences in (4a,b) show that a singular or a plural subject requires a verb with the corresponding inflection. In (4a,b), the verb only requires one argument so the function of *die Frau* and *die Mädchen* is clear. In (4c,d) the verb requires two arguments and *die Frau* and *die Mädchen* could appear in either argument position in German. The sentences could mean that the woman knows somebody or that somebody knows the woman. However, due to the inflection on the verb and knowledge of the syntactic rules of German, the hearer knows that there is only one available reading for (4c) and (4d), respectively.

It is the role of syntax to discover, describe and explain such rules, patterns and structures.

1.2 Why do it formally?

The two following quotations give a motivation for the necessity of describing language formally:

Precisely constructed models for linguistic structure can play an important role, both negative and positive, in the process of discovery itself. By pushing a precise but inadequate formulation to an unacceptable conclusion, we can often expose the exact source of this inadequacy and, consequently, gain a deeper understanding of the linguistic data. More positively, a formalized theory may automatically provide solutions for many problems other than those for which it was explicitly designed. Obscure and intuition-bound notions can neither lead to absurd conclusions nor provide new and correct ones, and hence they fail to be useful in two important respects. I think that some of those linguists who have questioned the value of precise and technical development of linguistic theory have failed to recognize the productive potential in the method of rigorously stating a proposed theory and applying it strictly to linguistic material with no attempt to avoid unacceptable conclusions by ad hoc adjustments or loose formulation. (Chomsky 1957: 5)

As is frequently pointed out but cannot be overemphasized, an important goal of formalization in linguistics is to enable subsequent researchers to see the defects of an analysis as clearly as its merits; only then can progress be made efficiently. (Dowty 1979: 322)

If we formalize linguistic descriptions then it is easier to recognise what exactly a particular analysis means. We can establish what predictions it makes and we can rule out alternative analyses. A further advantage of precisely formulated theories is that they can be written down in such a way that computer programs can process them. When a theoretical analysis is implemented as a computationally processable grammar fragment, any inconsistencies will become immediately evident. Such implemented grammars can then be used to process large collections of text, so-called corpora, and they can thus establish which sentences a particular grammar cannot yet analyze or which sentences are assigned the wrong structure. For more on using computer implementation in linguistics see Bierwisch (1963: 163), Müller (1999a: Chapter 22) and Bender (2008b) as well as Section 3.6.2.

1.3 Constituents

If we consider the sentence in (5), we have the intuition that certain words form a unit.

- (5) Alle Studenten lesen während dieser Zeit Bücher.
all students read during this time books
'All the students are reading books at this time.'

The words *alle* ‘all’ and *Studenten* ‘students’ thus form a unit which says something about who is reading. *während* ‘during’, *dieser* ‘this’ and *Zeit* ‘time’ also form a unit which refers to a period of time during which the reading takes place, and *Bücher* ‘books’ says something about what is being read. The first unit is itself made up of two parts, namely *alle* ‘all’ and *Studenten* ‘students’. The unit *während dieser Zeit* ‘during this time’ can also be divided into two subcomponents: *während* ‘during’ and *dieser Zeit* ‘this time’. *dieser Zeit* ‘this time’ is also composed of two parts, just like *alle Studenten* ‘all students’ is.

Recall that in connection with (1c) above we talked about the sets of Russian nesting dolls (*matryoshkas*). Here, too, when we break down (5) we have smaller units which are components of bigger units. However, in contrast to the Russian dolls, we do not just have one smaller unit contained in a bigger one but rather, we can have several units which are grouped together in a bigger one. The best way to envisage this is to imagine a system of small boxes: one big box contains the whole sentence. Inside this box, there are four other small boxes, which each contain *alle Studenten* ‘all students’, *lesen* ‘reads’, *während dieser Zeit* ‘during this time’ and *Bücher* ‘books’ respectively. Figure 1.1 illustrates this.

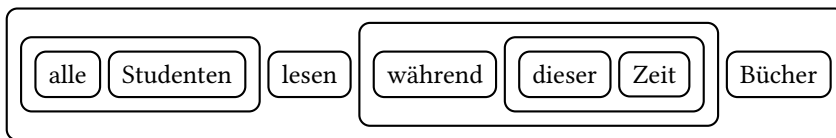


Figure 1.1: Words and phrases in boxes

In the following section, I will introduce various tests which can be used to show how certain words seem to “belong together” more than others. When I speak of a *word sequence*, I generally mean an arbitrary linear sequence of words which do not necessarily need to have any syntactic or semantic relationship, e. g. *Studenten lesen während* ‘students read during’ in (5). A group of words which form a structural entity, on the other hand, are referred to as a *phrase* or *constituent*. Individual words are a trivial case as these clearly always constitute a structural entity of exactly one element.

Following these preliminary remarks, I will now introduce some tests, which will help us to identify whether a particular string of words is a constituent or not.

1.3.1 Constituency tests

There are a number of ways to test the constituent status of a sequence of words. In the following sections, I will present some of these. In Section 1.3.2, we will see that that there are cases when simply applying a test “blindly” leads to unwanted results.

1.3.1.1 Substitution

If it is possible to replace a sequence of words in a sentence with a different sequence of words and the acceptability of the sentence remains unaffected, then this constitutes

1 Introduction and basic terms

evidence for the fact that each sequence of words forms a constituent.

In (6), *den Mann* ‘the man’ can be replaced by the string *eine Frau* ‘a woman’. This is an indication that both of these word sequences are constituents.

- (6) a. Er kennt [den Mann].
he knows the man
‘He knows the man.’
- b. Er kennt [eine Frau].
He knows a woman
‘He knows a woman.’

Similarly to (7a), the string *das Buch zu lesen* can be replaced by *der Frau das Buch zu geben*.

- (7) a. Er versucht, [das Buch zu lesen].
he tries the book to read
‘He is trying to read the book’
- b. Er versucht, [der Frau das Buch zu geben].
he tries the woman the book to give
‘He is trying to give the woman the book.’

This test is referred to as the *substitution test*.

1.3.1.2 Pronominalization

Everything that can be replaced by a pronoun forms a constituent. In (8), one can for example refer to *der Mann* ‘the man’ with the pronoun *er* ‘he’:

- (8) a. [Der Mann] schläft.
the man sleeps
‘The man is sleeping’
- b. Er schläft.
he sleeps
‘He is sleeping’

It is also possible to use a pronoun to refer to constituents such as *das Buch zu lesen* in (7a), as is shown in (9):

- (9) a. Peter versucht, [das Buch zu lesen].
Peter tries the book to read
‘Peter is trying to read the book.’
- b. Klaus versucht das auch.
Klaus tries that also
‘Klaus is trying to do that as well.’

The pronominalization test is another form of the substitution test.

1.3.1.3 Question formation

A sequence of words that can be elicited by a question forms a constituent:

- (10) a. [Der Mann] arbeitet.
 the man works
 ‘The man is working’
 b. Wer arbeitet?
 who works
 ‘Who is working?’

Question formation is a specific case of pronominalization. One uses a particular type of pronoun (an interrogative pronoun) to refer to the word sequence.

Constituents such as *das Buch zu lesen* in (7a) can also be elicited by questions, as (11) shows:

- (11) Was versucht er?
 what tries he
 ‘What does he try?’

1.3.1.4 Permutation test

If a sequence of words can be moved without adversely affecting the acceptability of the sentence in which it occurs, then this is an indication that this word group forms a constituent.

In (12), *keiner* ‘nobody’ and *diese Frau* ‘this woman’ exhibit different orderings, which suggests that *diese* and *Frau* belong together.

- (12) a. dass keiner [diese Frau] kennt
 that nobody this woman knows
 b. dass [diese Frau] keiner kennt
 that this woman nobody knows
 ‘that nobody knows this woman’

On the other hand, it is not plausible to assume that *keiner diese* forms a constituent in (12a). If we try to form other possible orderings by trying to move *keiner diese* as a whole, we see that this leads to unacceptable results:²

- (13) a. * dass Frau keiner diese kennt
 b. * dass Frau kennt keiner diese

² I use the following notational conventions for all examples: ‘*’ indicates that a sentence is ungrammatical, ‘#’ denotes that the sentence has a reading which differs from the intended one and finally ‘§’ should be understood as a sentence which is deviant for semantic or information-structural reasons, for example, because the subject must be animate, but is in fact inanimate in the example in question, or because there is a conflict between constituent order and the marking of given information through the use of pronouns.

1 Introduction and basic terms

Furthermore, constituents such as the *das Buch zu lesen* ‘to read the book’ in (7a) can be moved:

- (14) a. Er hat noch nicht [das Buch zu lesen] versucht.
 he has PART not the book to read tried
 ‘He has not yet tried to read the book.’
 b. Er hat [das Buch zu lesen] noch nicht versucht.
 he has the book to read PART not tried
 c. Er hat noch nicht versucht, [das Buch zu lesen].
 he has PART not tried the book to read

1.3.1.5 Fronting

Fronting is a further variant of the movement test. In German declarative sentences, only a single constituent may normally precede the finite verb:

- (15) a. [Alle Studenten] lesen während der vorlesungsfreien Zeit Bücher.
 all students read.3PL during the lecture.free time books
 ‘All students read books during the semester break.’
 b. [Bücher] lesen alle Studenten während der vorlesungsfreien Zeit.
 books read all students during the lecture.free time
 c. * [Alle Studenten] [Bücher] lesen während der vorlesungsfreien Zeit.
 all students books read during the lecture.free time
 d. * [Bücher] [alle Studenten] lesen während der vorlesungsfreien Zeit.
 books all students read during the lecture.free time

The possibility for a sequence of words to be fronted (that is occur in front of the finite verb) is a strong indicator of constituent status.

1.3.1.6 Coordination

If two sequences of words can be conjoined then this suggests that each word group forms a constituent.

In (16), *der Mann* ‘the man’ and *die Frau* ‘the woman’ are conjoined and the entire coordination is the subject of the verb *arbeiten* ‘to work’. This is a good indication of the fact that *der Mann* and *die Frau* each form a constituent.

- (16) [Der Mann] und [die Frau] arbeiten.
 the man and the woman work.3PL
 ‘The man and the woman work.’

The example in (17) shows that word groups with *to*-infinitives can be conjoined:

- (17) Er hat versucht, [das Buch zu lesen] und [es dann unauffällig verschwinden zu lassen].
 he had tried the book to read and it then secretly disappear to
 let
 ‘He had tried to read the book and then make it quietly disappear.’

1.3.2 Some comments on the status of constituent tests

It would be ideal if the tests presented here delivered clear-cut results in every case, as the empirical basis on which syntactic theories are built would thereby become much clearer. Unfortunately, this is not the case. There are in fact a number of problems with constituent tests, which I will discuss in what follows.

1.3.2.1 Expletives

There is a particular class of pronouns – so-called *expletives* – which do not denote people or things and are therefore non-referential. An example of this is *es* ‘it’ in (18).

- (18) a. Es regnet.
 it rains
 ‘It is raining.’
 b. Regnet es?
 rains it
 ‘Is it raining?’
 c. dass es jetzt regnet
 that it now rains
 ‘that it is raining now’

As the examples in (18) show, *es* can either precede the verb, or follow it. It can also be separated from the verb by an adverb, which suggests that *es* should be viewed as an independent unit.

Nevertheless, we observe certain problems with the aforementioned tests. Firstly, *es* is restricted with regard to its movement possibilities as (19a) and (20b) show.

- (19) a. *dass jetzt es regnet
 that now it rains
 Intended: ‘that it is raining now’
 b. dass jetzt keiner klatscht
 that now nobody claps
 ‘that nobody is clapping now’
 (20) a. Er sah es regnen.
 he saw it raining
 ‘He saw that it was raining.’

1 Introduction and basic terms

- b. * Es sah er regnen.
he saw him raining
Intended: 'He saw that it was raining.'
- c. Er sah einen Mann klatschen.
he saw a man clapping
'He saw a man clapping.'
- d. Einen Mann sah er klatschen.
a man saw he clapping
'A man, he saw clapping.'

Unlike the accusative object *einen Mann* 'a man' in (20c,d), the expletive in (20b) cannot be fronted.

Secondly, substitution and question tests also fail:

- (21) a. * Der Mann / er regnet.
the man he rains
Intended: 'The man/he rains.'
- b. * Wer/was regnet?
who/what rains
Intended: 'Who/what rains?'

Similarly, the coordination test cannot apply either:

- (22) * Es und der Mann regnet/regnen.
it and the man rains/rain
Intended: 'It and the man is/are raining.'

The failure of these tests can be easily explained: weakly stressed pronouns such as *es* are preferably placed before other arguments, directly after the conjunction (*dass* in (18c)) and directly after the finite verb in (20a) (see Abraham 1995: 570). If an element is placed in front of the expletive, as in (19a), then the sentence is rendered ungrammatical. The reason for the ungrammaticality of (20b) is the general ban on accusative *es* appearing in clause-initial position. Although such cases exist, they are only possible if *es* is referential (Lenerz 1994: 162; Gärtner & Steinbach 1997: 4).

The fact that we could not apply the substitution and question tests is also no longer mysterious as *es* is not referential in these cases. We can only replace *es* with another expletive such as *das*. If we replace the expletive with a referential expression, we derive a different semantic interpretation. It does not make sense to ask about something semantically empty or to refer to it with a pronoun.

It follows from this that not all of the tests must deliver a positive result for a sequence of words to count as a constituent, that is, the tests are therefore not a necessary requirement for constituent status.

1.3.2.2 Movement

The movement test is problematic for languages with relatively free constituent order, since it is not always possible to tell what exactly has been moved. For example, the string *gestern dem Mann* occupies different positions in the examples in (23).

- (23) a. weil keiner gestern dem Mann geholfen hat
 because nobody yesterday the man helped has
 ‘because nobody helped the man yesterday’
 b. weil gestern dem Mann keiner geholfen hat
 because yesterday the man nobody helped has
 ‘because nobody helped the man yesterday’

One could therefore assume that *gestern* ‘yesterday’ and *dem Mann* ‘the man’, which of course do not form a constituent, have been moved together. An alternative explanation for the ordering variants in (23) is that adverbs can occur in various positions in the clause and that only *dem Mann* has been moved in front of *keiner* in (23b). In any case, it is clear that *gestern* and *dem Mann* have no semantic relation and that it is impossible to refer to them both with a pronoun. Although it may seem at first glance that this material has been moved as a unit, we have seen that it is in fact not tenable to assume that *gestern dem Mann* ‘yesterday the man’ forms a constituent.

1.3.2.3 Fronting

As mentioned in the discussion of (15), the position in front of the finite verb is normally occupied by a single constituent. The possibility for a given word group to be placed in front of the finite verb is sometimes even used as a clear indicator of constituent status, and even used in the definition of *Satzglied* (a special term used in grammars of German referring to a constituent on the clause level) (Duden 2005: 783). An example of this is taken from Bußmann (1983) and is no longer present in Bußmann (1990):³

Satzglied test A process based on → topicalization used to analyse complex constituents. Since topicalization only allows a single constituent to be moved to the beginning of the sentence, complex sequences of constituents, for example adverb phrases, can be shown to actually consist of one or more constituents. In the example *Ein Taxi quält sich im Schrittempo durch den Verkehr* ‘A taxi was struggling at walking speed through the traffic’, *im Schrittempo* ‘at walking speed’ and *durch den Verkehr* ‘through the traffic’ are each constituents as both can be fronted independently of each other. (Bußmann 1983: 446)

³ **Satzgliedtest** [Auch: Konstituententest]. Auf der → Topikalisierung beruhendes Verfahren zur Analyse komplexer Konstituenten. Da bei Topikalisierung jeweils nur eine Konstituente bzw. ein → Satzglied an den Anfang gerückt werden kann, lassen sich komplexe Abfolgen von Konstituenten (z. B. Adverbialphrasen) als ein oder mehrere Satzglieder ausweisen; in *Ein Taxi quält sich im Schrittempo durch den Verkehr* sind *im Schrittempo* und *durch den Verkehr* zwei Satzglieder, da sie beide unabhängig voneinander in Anfangsposition gerückt werden können.

The preceding quote has the following implications:

- If some part of a piece of linguistic material can be fronted independently →
This material does not form a constituent.
- If linguistic material can be fronted together →
This material forms a constituent.

It will be shown that both of these prove to be problematic.

The first implication is cast into doubt by the data in (24):

- (24) a. Keine Einigung erreichten Schröder und Chirac über den Abbau der
no agreement reached Schröder and Chirac about the reduction of the
Agrarsubventionen.⁴
agricultural.subsidies
'Schröder and Chirac could not reach an agreement on the reduction of
agricultural subsidies.'
- b. [Über den Abbau der Agrarsubventionen] erreichten Schröder und
about the reduction of the agricultural.subsidies reached Schröder and
Chirac keine Einigung.
Chirac no agreement

Although parts of the noun phrase *keine Einigung über den Abbau der Agrarsubventionen* 'no agreement on the reduction of agricultural subsidies' can be fronted individually, we still want to analyse the entire word string as a noun phrase when it is not fronted as in (25):

- (25) Schröder und Chirac erreichten [keine Einigung über den Abbau der
Schröder and Chirac reached no agreement about the reduction of the
Agrarsubventionen].
agricultural.subsidies

The prepositional phrase *über den Abbau der Agrarsubventionen* 'on the reduction of agricultural subsidies' is semantically dependent on *Einigung* 'agreement' cf. (26):

- (26) Sie einigen sich über die Agrarsubventionen.
they agree REFL about the agricultural.subsidies
'They agree on the agricultural subsidies.'

This word sequence can also be fronted together:

- (27) [Keine Einigung über den Abbau der Agrarsubventionen] erreichten
no agreement about the reduction of the agricultural.subsidies reached
Schröder und Chirac.
Schröder and Chirac

⁴ tagesschau, 15.10.2002, 20:00.

In the theoretical literature, it is assumed that *keine Einigung über den Abbau der Agrar-subventionen* forms a constituent which can be “split up” under certain circumstances. In such cases, the individual subconstituents can be moved independently of each other (De Kuthy 2002) as we have seen in (25).

The second implication is problematic because of examples such as (28):

- (28) a. [Trocken] [durch die Stadt] kommt man am Wochenende auch mit
dry through the city comes one at.the weekend also with
der BVG.⁵
the BVG
‘With the BVG, you can be sure to get around town dry at the weekend.’
b. [Wenig] [mit Sprachgeschichte] hat der dritte Beitrag in dieser Rubrik zu
little with language.history has the third article in this section to
tun, [...]⁶
do
‘The third article in this section has little to do with language history.’

In (28), there are multiple constituents preceding the finite verb, which bear no obvious syntactic or semantic relation to each other. Exactly what is meant by a “syntactic or semantic relation” will be fully explained in the following chapters. At this point, I will just point out that in (28a) the adjective *trocken* ‘dry’ has *man* ‘one’ as its subject and furthermore says something about the action of ‘travelling through the city’, that is it refers to the action denoted by the verb. As (29b) shows, *durch die Stadt* ‘through the city’ cannot be combined with the adjective *trocken* ‘dry’.

- (29) a. Man ist / bleibt trocken.
one is stays dry
‘One stays dry.’
b. * Man ist / bleibt trocken durch die Stadt.
one is stays dry through the city
Intended: ‘One stays dry through the city.’

Therefore, the adjective *trocken* ‘dry’ does not have a syntactic or semantic relationship with the prepositional phrase *durch die Stadt* ‘through the city’. Both phrases have in common that they refer to the verb and are dependent on it.

One may simply wish to treat the examples in (28) as exceptions. This approach would, however, not be justified as I have shown in an extensive empirical study (Müller 2003a).

If one were to classify *trocken durch die Stadt* as a constituent due to it passing the fronting test, then one would have to assume that *trocken durch die Stadt* in (30) is also a constituent. In doing so, we would devalue the term *constituent* as the whole point

⁵ taz berlin, 10.07.1998, p. 22.

⁶ Zeitschrift für Dialektologie und Linguistik, LXIX, 3/2002, p. 339.

of constituent tests is to find out which word strings have some semantic or syntactic relationship.⁷

- (30) a. Man kommt am Wochenende auch mit der BVG trocken durch die
one comes at.the weekend also with the BVG dry through the
Stadt.
city
b. Der dritte Beitrag in dieser Rubrik hat wenig mit Sprachgeschichte zu tun.
the third article in this section has little with language.history to do

The possibility for a given sequence of words to be fronted is therefore not a sufficient diagnostic for constituent status.

We have also seen that it makes sense to treat expletives as constituents despite the fact that the accusative expletive cannot be fronted (cf. (20a)):

- (31) a. Er bringt es bis zum Professor.
he brings it until to.the professor
'He is bringing it to the professor.'
b. # Es bringt er bis zum Professor.
it brings he until to.the professor

There are other elements that can also not be fronted. Inherent reflexives are a good example of this:

- (32) a. Karl hat sich nicht erholt.
Karl has REFL not recovered
'Karl hasn't recovered.'
b. *Sich hat Karl nicht erholt.
REFL has Karl not recovered

It follows from this that fronting is not a necessary criterion for constituent status. Therefore, the possibility for a given word string to be fronted is neither a necessary nor sufficient condition for constituent status.

1.3.2.4 Coordination

Coordinated structures such as those in (33) also prove to be problematic:

- (33) Deshalb kaufte der Mann einen Esel und die Frau ein Pferd.
therefore bought the man a donkey and the woman a horse
'Therefore, the man bought a donkey and the woman a horse.'

⁷ These data can be explained by assuming a silent verbal head preceding the finite verb and thereby ensuring that there is in fact just one constituent in initial position in front of the finite verb (Müller 2005c). Nevertheless, this kind of data are problematic for constituent tests since these tests have been specifically designed to tease apart whether strings such as *trocken* and *durch die Stadt* or *wenig* and *mit Sprachgeschichte* in (30) form a constituent.

At first glance, *der Mann einen Esel* ‘the man a donkey’ and *die Frau ein Pferd* ‘the woman a horse’ in (33) seem to be coordinated. Does this mean that *der Mann einen Esel* and *die Frau ein Pferd* each form a constituent?

As other constituent tests show, this assumption is not plausible. This grouping of words cannot be moved together as a unit:⁸

(34) * *Der Mann einen Esel kaufte deshalb.*

Replacing the supposed constituent is also not possible without ellipsis:

(35) a. # *Deshalb kaufte er.*
therefore bought he

b. * *Deshalb kaufte ihn.*
therefore bought him

The pronouns do not stand in for the two logical arguments of *kaufen*, which are realized by *der Mann* and *einen Esel* in (33), but rather for one in each. There are analyses that have been proposed for examples such as (33) in which two verbs *kauft* occur, where only one is overt, however (Crysmann 2008). The example in (36) would therefore correspond to:

(36) *Deshalb kaufte der Mann einen Esel und kaufte die Frau ein Pferd.*
therefore bought the man a donkey and bought the woman a horse

This means that although it seems as though *der Mann einen Esel* and *die Frau ein Pferd* are coordinated, it is actually *kauft der Mann einen Esel* and *(kauft) die Frau ein Pferd* which are conjoined.

We should take the following from the previous discussion: even when a given word sequence passes certain constituent tests, this does not mean that one can automatically infer from this that we are dealing with a constituent, that is, the tests we have seen are not a sufficient condition for constituent status.

Summing up, it has been shown that these tests are neither sufficient nor necessary for attributing constituent status to a given sequence of words. However, as long as one keeps the problematic cases in mind, the previous discussion should be enough to get an initial idea about what should be treated as a constituent.

1.4 Parts of speech

The words in (37) differ not only in their meaning but also in other respects.

⁸ The area in front of the finite verb is also referred to as the *Vorfeld* ‘prefield’ (see Section 1.8). Putative multiple fronting is possible under certain circumstances in German. See the previous section, especially the discussion of the examples in (28) on page 15. The example in (34) is created in such a way so that the subject is present in the prefield which is not normally possible with verbs such as *kaufen* ‘to buy’ for reasons which have to do with the information-structural properties of these kinds of fronting constructions (see De Kuthy & Meurers (2003b)).

1 Introduction and basic terms

- (37) Der dicke Mann lacht jetzt.
The fat man laughs now
'The fat man is laughing now.'

Each of the words is subject to certain restrictions when forming sentences. It is common practice to group words into classes with other words which share certain salient properties. For example, *der* 'the' is an article, *Mann* 'man' is a noun, *lacht* 'laugh' is a verb and *jetzt* 'now' is an adverb. As can be seen in (38), it is possible to replace all the words in (37) with words from the same word class.

- (38) Die dünne Frau lächelt immer.
the thin woman smiles always
'The thin woman is always smiling.'

This is not always the case, however. For example, it is not possible to use a reflexive verb such as *erholt* 'recovers' or the second-person form *lächelst* in (38). The categorization of words into parts of speech is not this rudimentary. We will have to say a lot more about the properties of a given word. In this section, we will discuss various words classes/parts of speech and in the following sections we will go into further detail about the various properties which characterize a given word class.

The most important parts of speech are *verbs*, *nouns*, *adjectives*, *prepositions* and *adverbs*. In earlier decades, it was common among researchers working on German (see also Section 11.6.1 on Tesnière's category system) to speak of *action words*, *describing words*, and *naming words*. These descriptions prove problematic however, as illustrated by the following examples:

- (39) a. die *Idee*
the idea
b. die *Stunde*
the hour
c. das laute *Sprechen*
the loud speaking
'(the act of) speaking loudly'
d. Die *Erörterung* der Lage dauerte mehrere Stunden.
the discussion of the situation lasted several hours
'The discussion of the situation lasted several hours.'

(39a) does not describe a concrete entity, (39b) describes a time interval and (39c) and (39d) describe actions. It is clear that *Idee*, *Stunde*, *Sprechen* and *Erörterung* differ greatly in terms of their meaning. Nevertheless, these words still behave like *Mann* and *Frau* in many respects and are therefore still classed as nouns.

The term *action word* is not used any more in scientific grammars as verbs do not always need to denote actions:

- (40) a. Ihm gefällt das Buch.
him pleases the book
'He likes the book.'
- b. Das Eis schmilzt.
the ice melts
'The ice is melting.'
- c. Es regnet.
it rains
'It is raining.'

One would also have to class *Erörterung* 'discussion' as a action word.

Adjectives do not always describe properties of objects. In the following examples, the opposite is in fact true: The characteristic of being a murderer is expressed as being possible or probable.

- (41) a. der mutmaßliche Mörder
the suspected murderer
- b. Soldaten sind potenzielle Mörder.
soldiers are potential murderers

The adjectives themselves in (41) do not actually provide any information about the characteristics of the entities described. One may also wish to classify *lachende* 'laughing' in (42) as an adjective.

- (42) der lachende Mann
the laughing man

If, however, we are using properties and actions as our criteria for classification, *lachend* should technically be a action word.

Rather than semantic criteria, it is usually formal criteria which are used to determine word classes. The various forms a word can take are also taken into account. So *lacht* 'laughs', for example, has the forms given in (43).

- (43) a. Ich lache.
I laugh
- b. Du lachst.
you laugh
- c. Er lacht.
he laughs
- d. Wir lachen.
we laugh
- e. Ihr lacht.
you.PL laugh

1 Introduction and basic terms

- f. Sie lachen.
you laugh

In German, there are also forms for the preterite, imperative, subjunctive (I and II) and infinitive forms (participles and infinitives with or without *zu* 'to'). All of these forms constitute the inflectional paradigm of a verb. Tense (present, preterite, future), mood (indicative, subjunctive, imperative), person (1., 2., 3.) and number (singular, plural) all play a role in the inflectional paradigm. Certain forms can be collapsed together in a paradigm, as (43) shows.

Parallel to verbs, nouns also have an inflectional paradigm:

- (44) a. der Mann
the.NOM man
b. des Mannes
the.GEN man.GEN
c. dem Mann
the.DAT man
d. den Mann
the.ACC man
e. die Männer
the.NOM men
f. der Männer
the.GEN men
g. den Männern
the.DAT men
h. die Männer
the.ACC men

We can differentiate between nouns on the basis of gender (feminine, masculine, neuter). The choice of gender is often purely formal in nature and is only partially influenced by biological sex or the fact that we are describing a particular object:

- (45) a. die Tüte
the.F bag(F)
'the bag'
b. der Krampf
the.M cramp(M)
'cramp'
c. das Kind
the.N child(N)
'the child'

One should avoid using terms which refer to biological gender such as *männlich* ‘male’, *weiblich* ‘female’ and *sächlich* ‘inanimate’ in German. Gender or *genus* really means ‘kind’. Bantu languages can, for example, have between 7 and 10 genders (Corbett 2008).

As well as gender, case (nominative, genitive, dative, accusative) and number are also important for nominal paradigms.

Like nouns, adjectives inflect for gender, case and number. They differ from nouns however, in that gender marking is variable. Adjectives can be used with all three genders:

- (46) a. eine kluge Frau
a.F clever.F woman
b. ein kluger Mann
a clever.M man
c. ein kluges Kind
a clever.N child

In addition to gender, case and number, we can identify several other inflectional classes. Traditionally, we distinguish between strong, mixed and weak inflection of adjectives. The inflectional class that we have to choose is dependent on the form or presence of the article:⁹

- (47) a. ein alter Wein
an old wine
b. der alte Wein
the old wine
c. alter Wein
old wine

Furthermore, adjectives have comparative and superlative forms:

- (48) a. klug
clever
b. klüg-er
clever-er
c. am klüg-sten
at.the clever-est

This is not always the case. Adjectives which make reference to some end point are normally not able to be used in any other form than the positive. If a particular solution is optimal, for example, then no better one exists. Therefore, it does not make sense to speak of a “more optimal” solution. In a similar vein, it is not possible to be ‘deader’ than dead.

⁹ Dieter Wunderlich has shown in an unpublished article that one can get by with just strong and weak inflectional classes. For details see Pollard & Sag (1994: Section 2.2.5) or Müller (2007b: Section 13.2).

1 Introduction and basic terms

There are some special cases such as colour adjectives ending in *-a* in German *lila* ‘purple’ and *rosa* ‘pink’. These inflect optionally (49a), and the uninflected form is also possible:

- (49) a. eine lilan-e Blume
 a purple-r flower
 b. eine lila Blume
 a purple flower

In both cases, *lila* is classed an adjective. We can motivate this classification by appealing to the fact that both words occur at the same positions as other adjectives that clearly behave like adjectives with regard to inflection.

The parts of speech discussed thus far can all be differentiated in terms of their inflectional properties. For words which do not inflect, we have to use additional criteria. For example, we can classify words by the syntactic context in which they occur (as we did for the inflecting adjectives above). We can identify prepositions, adverbs, conjunctions, interjections and sometimes also particles. Prepositions are words which occur with a noun phrase whose case they determine:

- (50) a. in diesen Raum
 in this.ACC room
 b. in diesem Raum
 in this.DAT room

wegen ‘because’ is often classed as a preposition although it can also occur after the noun and in these cases would technically be a postposition:

- (51) des Geldes wegen
 the money.GEN because
 ‘because of the money’

It is also possible to speak of adpositions if one wishes to remain neutral about the exact position of the word.

Unlike prepositions, adverbs do not require a noun phrase.

- (52) a. Er schläft in diesem Raum.
 he sleeps in this room
 b. Er schläft dort.
 he sleeps there

Sometimes adverbs are simply treated as a special variant of prepositions (see page 96). The explanation for this is that a prepositional phrase such as *in diesem Raum* ‘in this room’ behaves exactly like the corresponding adverbs. *in* differs from *dort* ‘there’ in that it needs an additional noun phrase. These differences are parallel to what we have seen with other parts of speech. For instance, the verb *schlafen* ‘sleep’ requires only a noun phrase, whereas *erkennen* ‘recognize’ requires two.

- (53) a. Er schläft.
he sleeps
b. Peter erkennt ihn.
Peter recognizes him

Conjunctions can be subdivided into subordinating and coordinating conjunctions. Coordinating conjunctions include *und* ‘and’ and *oder* ‘or’. In coordinate structures, two word groups with the same syntactic properties are combined. They occur adjacent to one another. *dass* ‘that’ and *weil* ‘because’ are subordinating conjunctions because the clauses that they introduce are part of a larger clause.

- (54) a. Klaus glaubt, dass er lügt.
Klaus believes that he lies
‘Klaus believes that he is lying.’
b. Klaus glaubt ihm nicht, weil er lügt.
Klaus believes him not because he lies
‘Klaus doesn’t believe him because he is lying.’

Interjections are clause-like expressions such as *Ja!* ‘Yes!’, *Bitte!* ‘Please!’ *Hallo!* ‘Hello!’, *Hurra!* ‘Hooray!’, *Bravo!*, *Pst!*, *Plumps!* ‘Plop!’.

In the case that adverbs and prepositions cannot be assigned to a particular class, then adverbs are normally used as a kind of “left over” category in the sense that all non-inflecting words, which are neither prepositions, conjunctions nor interjections, are classed as adverbs. Sometimes this category for “left overs” is subdivided: only words which can appear in front of the finite verb when used as a constituent are referred to as adverbs. Those words which cannot be fronted are dubbed *particles*. Particles themselves can be subdivided into various classes based on their function, e. g. degree particles and illocutionary particles. Since these functionally defined classes also contain adjectives, I will not make this distinction and simply speak of *adverbs*.

We have already sorted a considerable number of inflectional words into word classes. When one is faced with the task of classifying a particular word, one can use the decision diagram in Figure 1.2 on the next page, which is taken from the Duden grammar (2005: 133) of German.

If a word inflects for tense, then it is a verb. If it displays different case forms, then one has to check if it has a fixed gender. If this is indeed the case, then we know that we are dealing with a noun. Words with variable gender have to be checked to see if they have comparative forms. A positive result will be a clear indication of an adjective. All other words are placed into a residual category, which the Duden refers to as pronouns/article words. Like in the class of non-inflectional elements, the elements in this remnant category are subdivided according to their syntactic behavior. The Duden grammar makes a distinction between pronouns and article words. According to this classification, pronouns are words which can replace a noun phrase such as *der Mann* ‘the man’, whereas article words normally combine with a noun. In Latin grammars, the notion of ‘pronoun’ includes both pronouns in the above sense and articles, since the

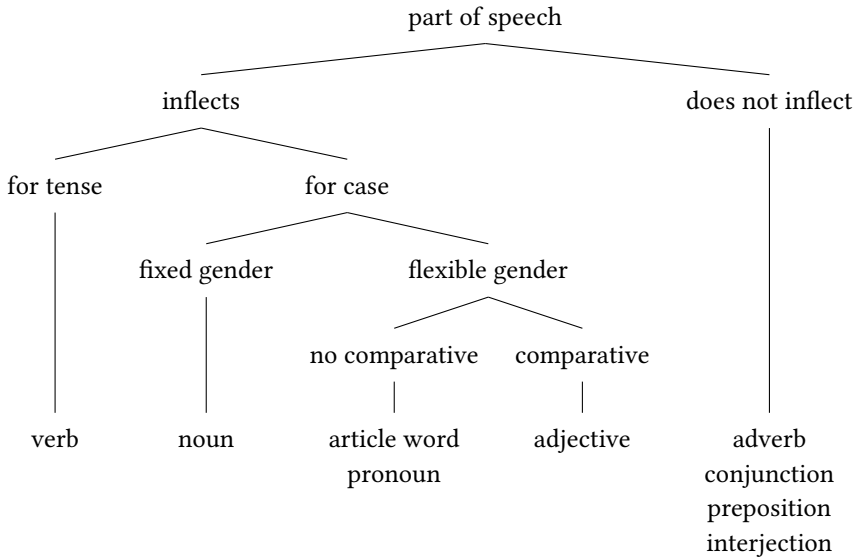


Figure 1.2: Decision tree for determining parts of speech following Duden (2005: 133)

forms with and without the noun are identical. Over the past centuries, the forms have undergone split development to the point where it is now common in contemporary Romance languages to distinguish between words which replace a noun phrase and those which must occur with one. Elements which belong to the latter class are also referred to as *determiners*.

If we follow the decision tree in Figure 1.2, the personal pronouns *ich* ‘I’, *du* ‘you’, *er* ‘he’, *sie* ‘her’, *es* ‘it’, *wir* ‘we’, *ihr* ‘you’, and *sie* ‘they’, for example, would be grouped together with the possessive pronouns *mein* ‘mine’, *dein* ‘your’, *sein* ‘his’/‘its’, *ihr* ‘her’/‘their’, and *unser* ‘our’. The corresponding reflexive pronouns, *mich* ‘myself’, *dich* ‘yourself’, *sich* ‘himself’, ‘herself’, ‘itself’, ‘themselves’, *uns* ‘ourselves’, *euch* ‘yourselves’, and the reciprocal pronoun *einander* ‘each other’ have to be viewed as a special case in German as there are no differing gender forms of *sich* ‘himself’/‘herself’, ‘itself’ and *einander* ‘each other’. Case is not expressed morphologically by reciprocal pronouns. By replacing genitive, dative and accusative pronouns with *einander*, it is possible to see that there must be variants of *einander* ‘each other’ in these cases, but these variants all share the same form:

- (55) a. Sie gedenken seiner / einander.
they commemorate him.GEN each.other
b. Sie helfen ihm / einander.
they help him.DAT each.other

- c. Sie lieben ihn / einander.
they love him.ACC each.other

So-called pronominal adverbs such as *darauf* ‘on there’, *darin* ‘in there’, *worauf* ‘on where’, *worin* ‘in where’ also prove problematic. These forms consist of a preposition (e. g. *auf* ‘on’) and the elements *da* ‘there’ and *wo* ‘where’. As the name suggests, *pronominal adverbs* contain something pronominal and this can only be *da* ‘there’ and *wo* ‘where’. However, *da* ‘there’ and *wo* ‘where’ do not inflect and would therefore, following the decision tree, not be classed as pronouns.

The same is true of relative pronouns such as *wo* in (56):

- (56) a. Ich komme eben aus der Stadt, wo ich Zeuge eines Unglücks gewesen
I come PRT from the city where I witness of.an accident been
bin.¹⁰
am
‘I come from the city where I was witness to an accident.’
- b. Studien haben gezeigt, daß mehr Unfälle in Städten passieren, wo die
studies have shown that more accidents in cities happen where the
Zebrastreifen abgebaut werden, weil die Autofahrer unaufmerksam
zebra.crossings removed become because the drivers unattentive
werden.¹¹
become
‘Studies have shown that there are more accidents in cities where they do
away with zebra crossings, because drivers become unattentive.’
- c. Zufällig war ich in dem Augenblick zugegen, wo der Steppenwolf
coincidentally was I in the moment present where the Steppenwolf
zum erstenmal unser Haus betrat und bei meiner Tante sich
to.the.first.time our house entered and by my aunt REFL
einmietete.¹²
took.lodgings
‘Coincidentally, I was present at the exact moment in which Steppenwolf
entered our house for the first time and took lodgings with my aunt.’

If they are uninflected, then they cannot belong to the class of pronouns according to the decision tree above. Eisenberg (2004: 233) notes that *wo* is a kind of uninflected relative pronoun and remarks that this description runs contrary to the use of this term for nominal, that is inflectional, elements. He therefore uses the term *relative adverb* for them (also see Duden (2005: §856, §857)).

There are also usages of the relatives *dessen* ‘whose’ and *wessen* ‘whose’ in combination with a noun:

¹⁰ Duden (1984: 672).

¹¹ taz berlin, 03.11.1997, p. 23.

¹² Herman Hesse, *Der Steppenwolf*. Berlin und Weimar: Aufbau-Verlag. 1986, p. 6.

1 Introduction and basic terms

- (57) a. der Mann, dessen Schwester ich kenne
the man whose sister I know
b. Ich möchte wissen, wessen Schwester du kennst.
I would like know whose sister you know
'I would like to know whose sister you know.'

According to the classification in the Duden, these should be covered by the terms *Relativartikelwort* 'relative article word' and *Interrogativartikelwort* 'interrogative article word'. They are mostly counted as part of the relative pronouns and question pronouns (see for instance Eisenberg (2004: 229)). Using Eisenberg's terminology, this is unproblematic as he does not make a distinction between articles, pronouns and nouns, but rather assigns them all to the class of nouns. But authors who do make a distinction between articles and pronouns sometimes also speak of interrogative pronouns when discussing words which can function as articles or indeed replace an entire noun phrase.

One should be prepared for the fact that the term *pronoun* is often simply used for words which refer to other entities and importantly not in the way that nouns such as *book* and *John* do, but rather dependent on context. The personal pronoun *er* 'he' can, for example, refer to either a table or a man. This usage of the term *pronoun* runs contrary to the decision tree in 1.2 and excludes uninflected elements such as *da* 'there' and *wo* 'where'.

Expletive pronouns such as *es* 'it' and *das* 'that', as well as the *sich* 'him/her/itself' belonging to inherently reflexive verbs do not make reference to actual objects. They are considered pronouns because of the similarity in form. Even if we were to assume a narrow definition of pronouns, we would still get the wrong results as expletive forms do not vary with regard to case, gender and number. If one does everything by the book, expletives would belong to the class of uninflected elements. If we assume that *es* 'it' as well as the personal pronouns have a nominative and accusative variant with the same form, then they would be placed in with the nominals. We would then have to admit that the assumption that *es* has gender would not make sense, that is we would have to count *es* as a noun by assuming neuter gender, analogous to personal pronouns.

We have not yet discussed how we would deal with the italicized words in (58):

- (58) a. das *geliebte* Spielzeug
the beloved toy
b. das *schlafende* Kind
the sleeping child
c. die Frage des *Sprechens* und *Schreibens* über Gefühle
the question of the talking and writing about feelings
'The question of talking and writing about one's feelings'
d. Auf dem Europa-Parteitag fordern die *Grünen* einen ökosozialen
on the Europe-party.conference demand the Greens a eco-social

- Politikwechsel.
political.change
'At the meeting European party conference, the Greens demanded eco-social political change.'
- e. Max lacht *laut*.
Max laughs loudly
- f. Max würde *wahrscheinlich* lachen.
Max would probably laugh

geliebte and *schlafende* are the present participle forms of *lieben* 'to love' and *schlafen* 'to sleep'. These forms are traditionally treated as part of the verbal paradigm. In this sense, *geliebte* and *schlafende* are verbs. This is referred to as lexical word class. The term *lexeme* is relevant in this case. All forms in a given inflectional paradigm belong to the relevant lexeme. In the classic sense, this term also includes the regularly derived forms, that is participle forms and nominalized infinitives also belong to a verbal lexeme. Not all linguists share this view, however. Particularly problematic is the fact that we are mixing verbal with nominal and adjectival paradigms. For example, *Sprechens* 'speaking.GEN' is in the genitive and adjectival participles also inflect for case, number and gender. Furthermore, it is unclear as to why *schlafende* 'sleeping' should be classed as a verbal lexeme and a noun such as *Störung* 'disturbance' is its own lexeme and does not belong to the lexeme *stören* 'to disturb'. I subscribe to the more modern view of grammar and assume that processes in which a word class is changed results in a new lexeme being created. Consequently, *schlafende* 'sleeping' does not belong to the lexeme *schlafen* 'to sleep', but is a form of the lexeme *schlafend*. This lexeme belongs to the word class 'adjective' and inflects accordingly.

As we have seen, it is still controversial as to where to draw the line between inflection and derivation (creation of a new lexeme). Sag, Wasow & Bender (2003: 263–264) view the formation of the present participle (*standing*) and the past participle (*eaten*) in English as derivation as these forms inflect for gender and number in French.

Adjectives such as *Grünen* 'the Greens' in (58d) are nominalised adjectives and are written with a capital like other nouns in German when there is no other noun present in the immediate context:

- (59) A: Willst du den roten Ball haben?
want you the red ball have
'Do you want the red ball?'
- B: Nein, gib mir bitte den grünen.
no give me please the green
'No, give me the green one please.'

In the answer to (59), the noun *Ball* has been omitted. This kind of omission is not present in (58d). One could also assume here that a word class change has taken place. If a word changes its class without combination with a visible affix, we refer to this as *conversion*. Conversion has been treated as a sub-case of derivation by some linguists.

1 Introduction and basic terms

The problem is, however, that *Grüne* ‘greens’ inflects just like an adjective and the gender varies depending on the object it is referring to:

- (60) a. ein Grüner hat vorgeschlagen, ...
a green.M has suggested
‘a (male) member of the Green Party suggested ...’
b. eine Grüne hat vorgeschlagen, ...
a green.F has suggested
‘a (female) member of the Green Party suggested ...’

We also have the situation where a word has two properties. We can make life easier for ourselves by talking about *nominalized adjectives*. The lexical category of *Grüne* is adjective and its syntactic category is noun.

The word in (58e) can inflect like an adjective and should therefore be classed as an adjective following our tests. Sometimes, these kinds of adjectives are also classed as adverbs. The reason for this is that the uninflected forms of these adjectives behave like adverbs:

- (61) Max lacht immer/oft/laut.
Max laughs always/often/loud
‘Max (always/often) laughs (loudly).’

We can conclude therefore that its lexical category is that of an adjective and the syntactic category to which it belongs is ‘adverb’. The classification of adjectives such as *laut* ‘loud(ly)’ in (61) as adverbs is not assumed by all authors. Instead, some speak of adverbial usage of an adjective, that is one assumes that the syntactic category is still adjective but it can be used in a different way so that it behaves like an adverb (see Eisenberg (2004: Section 7.3), for example). This is parallel to prepositions, which can occur in a variety of syntactic contexts:

- (62) a. Peter schläft im Büro.
Peter sleeps in.the office
‘Peter sleeps in the office.’
b. der Tisch im Büro
the table in.the office
‘The table in the office.’

We have prepositional phrases in both examples in (62), however in (62a) *im Büro* ‘in the office’ acts like an adverb in that it modifies the verb *schläft* ‘sleeps’ and in (62b) *im Büro* modifies the noun *Tisch* ‘table’. In the same way, *laut* ‘loud’ can modify a noun (63) or a verb (61).

- (63) die laute Musik
the loud music

Lastly, I would like to discuss (58f) as a particularly difficult case. Words such as *wahrscheinlich* ‘probably’, *hoffentlich* ‘hopefully’ and *glücklicherweise* ‘fortunately’ are referred to as sentential adverbs. These modify the entire utterance and give an indication of speaker attitude. Inflectional elements such as *vermutlich* ‘supposed(ly)’ and *wahrscheinlich* ‘probable/-ly’ also belong to this semantically-motivated word class. If we want to refer to all these words as adverbs, then we would have to assume that conversion has taken place in cases such as *wahrscheinlich* ‘probably’ – that is that it belongs to the lexical category ‘adjective’ but the syntactic category ‘adverb’.

1.5 Heads

The head of a constituent/phrase is the element which determines the most important properties of the constituent/phrase. At the same time, the head also determines the composition of the phrase, that is, the head requires certain other elements to be present in the phrase. The heads in the following examples have been marked in *italics*:

- (64) a. *Träumt* dieser Mann?
dreams this.NOM man
‘Does this man dream?’
b. *Erwartet* er diesen Mann?
expects he this.ACC man
‘Is he expecting this man?’
c. *Hilft* er diesem Mann?
helps he this.DAT man
‘Is he helping this man?’
d. *in* diesem Haus
in this.DAT house
e. ein *Mann*
a man

Verbs determine the case of their arguments (subjects and objects). In (64d), the preposition determines which case the noun phrase *diesem Haus* ‘this house’ bears (dative) and also determines the semantic contribution of the phrase (it describes a location). (64e) is controversial: there are linguists who believe that the determiner is the head (Hellan 1986; Abney 1987; Netter 1994; 1998), as well as those who assume that the noun forms the head of the phrase (Van Langendonck 1994; Pollard & Sag 1994: 49; Demske 2001; Müller 2007b: Section 6.6.1; Hudson 2004).

The combination of a head with another constituent is called the *projection of the head*. A projection which contains all the necessary parts to create a well-formed phrase of that type is a *maximal projection*. A sentence is the maximal projection of a finite verb.

Figure 1.3 on the following page shows the structure of (65) in box representation.

- (65) Der Mann liest einen Aufsatz.
 the man reads an essay
 ‘The man is reading an essay.’

Unlike Figure 1.1, the boxes have been labelled here.

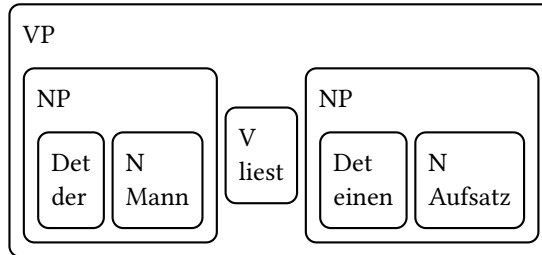


Figure 1.3: Words and phrases in annotated boxes

The annotation includes the category of the most important element in the box. VP stands for *verb phrase* and NP for *noun phrase*. VP and NP are maximal projections of their respective heads.

Anyone who has ever faced the hopeless task of trying to find particular photos of their sister’s wedding in a jumbled, unsorted cupboard can vouch for the fact that it is most definitely a good idea to mark the boxes based on their content and also mark the albums based on the kinds of photos they contain.

An interesting point is that the exact content of the box with linguistic material does not play a role when the box is put into a larger box. It is possible, for example, to replace the noun phrase *der Mann* ‘the man’ with *er*, or indeed the more complex *der Mann aus Stuttgart, der das Seminar zur Entwicklung der Zebrafinken besucht* ‘the man from Stuttgart who went to the seminar on the development of zebra finches’. However, it is not possible to use *die Männer* ‘the men’ or *des Mannes* ‘of the man’ in this position:

- (66) a. *Die Männer liest einen Aufsatz.
 the men reads an essay
 b. *Des Mannes liest einen Aufsatz.
 of.the man.GEN reads an essay

The reason for this is that *die Männer* ‘the men’ is in plural and the verb *liest* ‘reads’ is in singular. The noun phrase bearing genitive case *des Mannes* can also not occur, only nouns in the nominative case. It is therefore important to mark all boxes with the information that is important for placing these boxes into larger boxes. Figure 1.4 on the next page shows our example with more detailed annotation.

The features of a head which are relevant for determining in which contexts a phrase can occur are called *head features*. The features of a head are said to be *projected*.

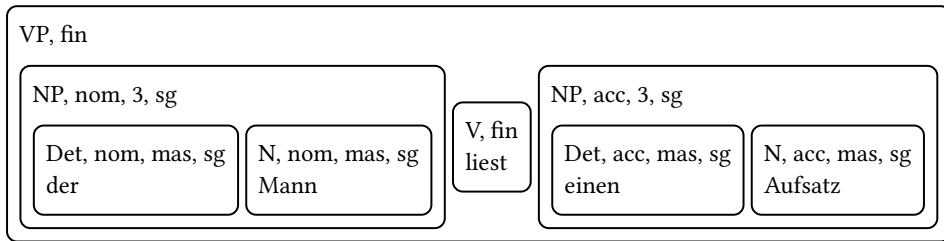


Figure 1.4: Words and word strings in annotated boxes

1.6 Arguments and adjuncts

The constituents of a given clause have different relations to their head. It is typical to distinguish between arguments and adjuncts. The syntactic arguments of a head correspond for the most part to their logical arguments. We can represent the meaning of (67a) as (67b) using predicate logic.

- (67) a. Peter helps Maria.
 b. *help'*(*peter'*, *maria'*)

The logical representation of (67b) resembles what is expressed in (67a), however it abstracts away from constituent order and inflection. *Peter* and *Maria* are syntactic arguments of the verb *helping* and their respective meanings (*Peter'* and *Maria'*) are arguments of the logical relation expressed by *help'*. One could also say that *help* assigns semantic roles to its arguments. Semantic roles include agent (the person carrying out an action), patient (the affected person or thing), beneficiary (the person who receives something) and experiencer (the person experiencing a psychological state). The subject of *help* is an agent and the direct object is a beneficiary. Arguments which fulfil a semantic role are also called *actants*. This term is also used for inanimate objects.

This kind of relation between a head and its arguments is covered by the terms *selection* and *valency*. Valency is a term borrowed from chemistry. Atoms can combine with other atoms to form molecules with varying levels of stability. The way in which the electron shells are occupied plays an important role for this stability. If an atom combines with others atoms so that its electron shell is fully occupied, then this will lead to a stable connection. Valency tells us something about the number of hydrogen atoms, which an atom of a certain element can be combined with. In forming H₂O, oxygen has a valency of 2. We can divide elements into valency classes. Following Mendelev, elements with a particular valency are listed in the same column in the periodic table.

The concept of valency was applied to linguistics by Tesnière (1959): a head needs certain arguments in order to form a stable compound. Words with the same valency – that is which require the same number and type of arguments – are divided into valency classes. Figure 1.5 on the following page shows examples from chemistry as well as linguistics.



Figure 1.5: Combination of hydrogen and oxygen and the combination of a verb with its arguments

We used (67) to explain logical valency. Logical valency can, however, sometimes differ from syntactic valency. This is the case with verbs like *rain*, which require an expletive pronoun as an argument. Inherently reflexive verbs such as *sich erholen* ‘to recover’ in German are another example.

- (68) a. Es regnet.
 it rains
 ‘It is raining.’
 b. Klaus erholt sich.
 Klaus recover REFL
 ‘Klaus is recovering.’

The expletive *es* ‘it’ with weather verbs and the *sich* of so-called inherent reflexives such as *erholen* ‘to recover’ have to be present in the sentence. Germanic languages have expletive elements that are used to fill the position preceding the finite verb. These positional expletives are not realized in embedded clauses in German, since embedded clauses have a structure that differs from canonical unembedded declarative clauses, which have the finite verb in second position. (69a) shows that *es* cannot be omitted in *dass*-clauses.

- (69) a. *Ich glaube, dass regnet.
 I think that rains
 Intended: ‘I think that it is raining.’
 b. *Ich glaube, dass Klaus erholt.
 I believe that Klaus recovers
 Intended: ‘I believe that Klaus is recovering.’

Neither the expletive nor the reflexive pronoun contribute anything semantically to the sentence. They must, however, be present to derive a complete, well-formed sentence. They therefore form part of the valency of the verb.

Constituents which do not contribute to the central meaning of their head, but rather provide additional information are called *adjuncts*. An example is the adverb *deeply* in (70):

- (70) John loves Mary deeply.

This says something about the intensity of the relation described by the verb. Further examples of adjuncts are attributive adjectives (71a) and relative clauses (71b):

- (71) a. a *beautiful* woman
 b. the man *who Mary loves*

Adjuncts have the following syntactic/semantic properties:

- (72) a. Adjuncts do not fulfil a semantic role.
 b. Adjuncts are optional.
 c. Adjuncts can be iterated.

The phrase in (71a) can be extended by adding another adjunct:

- (73) a beautiful clever woman

If one puts processing problems aside for a moment, this kind of extension by adding adjectives could proceed infinitely. (see the discussion of (37) on page 67). Arguments, on the other hand, cannot be realised more than once:

- (74) *Der Mann der Junge schläft.
 the man the boy sleeps

If the entity carrying out the sleeping action has already been mentioned, then it is not possible to have another noun phrase which refers to a sleeping individual. If one wants to express the fact that more than one individual is sleeping, then this must be done by means of coordination as in (75):

- (75) The man and the boy are sleeping.

One should note that the criteria for identifying adjuncts proposed in (72) is not sufficient, since there are also syntactic arguments, which do not fill semantic roles (e. g. *es* 'it' in (68a) and *sich* (REFL) in (68b)) or are optional (*pizza* in (76)).

- (76) Tony is eating (pizza).

Heads normally determine the syntactic properties of their arguments in a relatively fixed way. A verb is responsible for the case which its arguments bear.

- (77) a. Er gedenkt des Opfers.
 he remembers the.GEN victim.GEN
 'He remembers the victim.'
 b. *Er gedenkt dem Opfer.
 he remembers the.DAT victim
 c. Er hilft dem Opfer.
 he helps the.DAT victim
 'He helps the victim.'
 d. *Er hilft des Opfers.
 he helps the.GEN victim.GEN

The verb *governs* the case of its arguments.

The preposition and the case of the noun phrase in the prepositional phrase are both determined by the verb:¹³

- (78) a. Er denkt an seine Modelleisenbahn.
he thinks on his.ACC model.railway
'He is thinking of his model railway.'
- b. # Er denkt an seiner Modelleisenbahn.
He thinks on his.DAT model.railway
- c. Er hängt an seiner Modelleisenbahn.
He hangs on his.DAT model.railway
'He is spending time with his model railway.'
- d. * Er hängt an seine Modelleisenbahn.
he hangs on his.ACC model.railway

The case of noun phrases in modifying prepositional phrases, on the other hand, depends on their meaning. In German, directional prepositional phrases normally require a noun phrase bearing accusative case (79a), whereas local PPs (denoting a fixed location) appear in the dative case (79b):

- (79) a. Er geht in die Schule / auf den Weihnachtsmarkt / unter die
he goes in the.ACC school on the.ACC christmas.market under the.ACC
Brücke.
bridge
'He is going to school/to the Christmas market/under the bridge.'
- b. Er schläft in der Schule / auf dem Weihnachtsmarkt / unter der
he sleeps in the.DAT school on the.DAT Christmas.market under the.DAT
Brücke.
bridge
'He is sleeping at school/at the Christmas market/under the bridge.'

An interesting case is the verb *sich befinden*, which expresses the location of something (similar to 'to find oneself somewhere'. This cannot occur without some information about the location pertaining to the verb:

- (80) * Wir befinden uns.
they are REFL

The exact form of this information is not fixed – neither the syntactic category nor the preposition inside of prepositional phrases is restricted:

- (81) Wir befinden uns hier / unter der Brücke / neben dem Eingang / im Bett.
we are REFL here under the bridge next.to the entrance in bed
'We are here/under the bridge/next to the entrance/in bed.'

¹³ For similar examples, see Eisenberg (1994b: 78).

Local modifiers such as *hier* ‘here’ or *unter der Brücke* ‘under the bridge’ are analysed with the regard to other verbs (e.g. *schlafen* ‘sleep’) as adjuncts. For verbs such as *sich befinden* ‘to be (located)’, we will most likely have to assume that information about location forms an obligatory syntactic argument of the verb.¹⁴

¹⁴ The verb *wohnen* ‘to live’ is also discussed in a similar context. The prepositional phrase in (i.b) is assumed to form part of the valency of the verb (See Steinitz (1969: Chapter 2), Helbig & Schenkel (1973: 127), Engel (1994: 99), Kaufmann (1995: 119), Abraham (2005: 21)). Simple sentences with *wohnen* ‘to live’ without information about a location or situation are mostly deviant.

- (i) a. ?Er wohnt.
he lives
‘He lives.’
- b. Er wohnt in Bremen.
he lives in Bremen
‘He lives in Bremen.’
- c. Er wohnt allein.
he lives alone
‘He lives alone.’

As (ii) shows, it is not possible in general to rule out cases of *wohnen* without information about location:

- (ii) a. Das Landgericht Bad Kreuznach wies die Vermieterklage als unbegründet zurück, die the state.court Bad Kreuznach rejected the landlord.lawsuit as unfounded back the Mieterfamilie kann wohnen bleiben. (Mieterzeitung 6/2001, p. 14)
renting.family can living stay
‘The state court of Bad Kreuznach rejected the landlord’s lawsuit as unfounded and the family renting the property can carry on living (there).’
- b. Die Bevölkerungszahl explodiert. Damit immer mehr Menschen wohnen können, wächst the population exploded So.that always more people live can grows Hongkong, die Stadt, und nimmt sich ihr Terrain ohne zu fragen. (taz, 31.07.2002, p. 25)
Hongkong the city and takes REFL her terrain without to ask
‘The total population has exploded. In order for more people to be able to live (there), the city of Hongkong has been growing and simply occupying more terrain without asking.’
- c. Selbst wenn die Hochschulen genug Studienplätze für alle schaffen, müssen die Studenten even if the universities enough study.places for all create must the students auch wohnen und essen. (taz, 16.02.2011, p. 7)
also live and eat
‘Even if universities can manage to create enough places for everyone, the students still need to finance food and lodgings.’
- d. Wohnst Du noch, oder lebst Du schon?
Live you still or live you already
‘Are you just living somewhere, or are you at home?’ (strassen|feger, Obdachlosenzeitung Berlin, 01/2008, p. 3)
- e. Wer wohnt, verbraucht Energie – zumindest normalerweise.
who lives uses energy at.least normally
‘Everyone who lives (somewhere), uses energy – at least that is normally true.’ (taz, berlin, 15.12.2009, p. 23)

If we do not want to completely rule out sentences without a modifier, then the preposition in (i.b) would be an optional modifier which is still somehow part of the valency of the verb. This does not seem to make sense. We should therefore view *wohnen* as an intransitive verb.

The verb selects a phrase with information about location, but does not place any syntactic restrictions on its type. This specification of location behaves semantically like the other adjuncts we have previously seen. If I just consider the semantic aspects of the combination of a head and adjunct, then I also refer to the adjunct as a *modifier*.¹⁵ It is also possible to include arguments specifying location with verbs such as *sich befinden* ‘to be located’. Modifiers are normally adjuncts, and therefore optional, whereas in the case of *sich befinden* they seem to be (obligatory) arguments.

In conclusion, we can say that constituents that are required to occur with a certain head are arguments of that head. Furthermore, constituents which fulfil a semantic role with regard to the head are also arguments. These kinds of arguments can, however, sometimes be optional.

Arguments are normally divided into subjects and complements.¹⁶ Not all heads require a subject (see Müller 2007b: Section 3.2). The number of arguments of a head can therefore also correspond to the number of complements of a head.

1.7 Grammatical functions

In some theories, grammatical functions such as subject and object form part of the formal description of language (see Chapter 7, for example). This is not the case for the majority of the theories discussed here, but these terms are used for the informal description of certain phenomena. For this reason, I will briefly discuss them in what follows.

1.7.1 Subjects

Although I assume that the reader has a clear intuition about what a subject is, it is by no means a trivial matter to arrive at a definition of the word *subject* which can be used cross-linguistically. For German, Reis (1982) suggested the following syntactic properties as definitional for subjects:

- agreement of the finite verb with it
- nominative case in non-copular clauses

(i.a) should also be deviant since this particular expression is not very informative (see also Welke (1988: 28, 38–40) on this point), since a person normally lives *somewhere* (even if it is, regrettably, for some under a bridge). In (ii.a) it is explicit that the family lives in a rented property. The location does there not have to be repeated as an explicit argument of *wohnen*. It is only the question of whether the family can continue to live there or not that is relevant in (ii.a). Similarly, it is the fact of living somewhere in general and not the exact place which is important in (ii.b).

See Goldberg & Ackerman (2001) for more on modifiers which are obligatory in certain contexts due to pragmatic reasons.

¹⁵ See Section 1.7.2 for more on the grammatical function of adverbials. The term adverbial is normally used in conjunction with verbs. *modifier* is a more general term, which normally includes attributive adjectives.

¹⁶ In some schools the term complement is understood to include the subject, that is, the term complement is equivalent to the term argument (see for instance Groß 2003: 342). Some researchers treat some subjects, e.g. those of finite verbs, as complements (Pollard 1996b; Eisenberg 1994a: 376).

- omitted in infinitival clauses (control)
- optional in imperatives

I have already discussed agreement in conjunction with the examples in (4). Reis (1982) argued that the second bullet point is a suitable criterion for German. She formulates a restriction to non-copular clause because there can be more than one nominative argument in sentences with predicate nominals such as (82):

- (82) a. Er ist ein Lügner.
 he.NOM ist a liar.NOM
 ‘He is a liar.’
 b. Er wurde ein Lügner genannt.
 he.NOM was a liar.NOM called
 ‘He was called a liar.’

Following this criterion, arguments in the dative case such as *den Männern* ‘the men’ cannot be classed as subjects in German:

- (83) a. Er hilft den Männern.
 he helps the.DAT men.DAT
 ‘He is helping the men.’
 b. Den Männern wurde geholfen.
 the.DAT men.DAT were.3SG helped
 ‘The men were helped.’

Following the other criteria, datives should also not be classed as subjects – as Reis (1982) has shown. In (83b), *wurde*, which is the 3rd person singular form, does not agree with *den Männern*. The third of the aforementioned criteria deals with infinitive constructions such as those in (84):

- (84) a. Klaus behauptet, den Männern zu helfen.
 Klaus claims the.DAT men.DAT to help
 ‘Klaus claims to be helping the men.’
 b. Klaus behauptet, dass er den Männern hilft.
 Klaus claims that he the.DAT men.DAT helps
 ‘Klaus claims that he is helping the men.’
 c. *Die Männer behaupten, geholfen zu werden.
 the men claim helped to become
 Intended: ‘The men are claiming to be helped.’
 d. *Die Männer behaupten, elegant getanzt zu werden.
 the men claim elegantly danced to become
 Intended: ‘The men claim to be elegantly danced.’

In the first sentence, the argument of the verb *helfen* ‘to help’ has been omitted. If one wishes to explicitly mention it, then one would have to use the subordinate clause beginning with *dass* as in (84b). Examples (84c,d) show that infinitives which do not require a nominative argument cannot be embedded under verbs such as *behaupten* ‘to claim’. If the dative noun phrase *den Männern* ‘the men’ were the subject, we would expect to be able to derive the control construction (84c) from (84b). This is, however, not possible. Instead of (84c), it is necessary to use (85):

- (85) Die Männer behaupten, dass ihnen geholfen wird.
 the men.NOM claim that them.DAT helped becomes
 ‘The men claim that they are being helped.’

In the same way, imperatives are not possible with verbs that do not require a nominative.

(86) shows some examples from Reis (1982: 186).

- (86) a. Fürchte dich nicht!
 be.sacred REFL not
 ‘Don’t be scared!’
 b. *Graue nicht!
 dread not
 ‘Don’t dread it!’
 c. Werd einmal unterstützt und ...
 be once supported and
 ‘Let someone support you for once and ...’
 d. *Werd einmal geholfen und ...
 be once helped and
 ‘Let someone help you and ...’

The verb *sich fürchten* ‘to be scared’ in (86a) obligatorily requires a nominative argument as its subject (87a). The similar verb *grauen* ‘to dread’ in (86b) takes a dative argument (87b).

- (87) a. Ich fürchte mich vor Spinnen.
 I.NOM be.sacred REFL before spiders
 ‘I am scared of spiders.’
 b. Mir graut vor Spinnen
 me.DAT scares before spiders
 ‘I am dreading spiders.’

Interestingly, dative arguments in Icelandic behave differently. Zaenen, Maling & Thráins-son (1985) discuss various characteristics of subjects in Icelandic and show that it makes sense to describe dative arguments as subjects in passive sentences even if the finite verb does not agree with them (Section 3.1) or they do not bear nominative case. An example of this are infinitive constructions with an omitted dative argument (p. 457):

- (88) a. Ég vonast til að verða hjálpað.
I hope for to be helped
'I hope that I will be helped.'
- b. Að vera hjálpað í prófinu er óleyfilegt.
to be helped on the.exam is not.allowed
'It is not allowed for one to be helped during the exam.'

In a number of grammars, clausal arguments such as those in (89) are classed as subjects as they can be replaced by a noun phrase in the nominative (90) (see e. g. Eisenberg (2004: 63, 289)).

- (89) a. Dass er schon um sieben kommen wollte, stimmt nicht.
that he already at seven come wanted is.true not
'It's not true that he wanted to come as soon as seven.'
- b. Dass er Maria geheiratet hat, gefällt mir.
that he Maria married has pleases me
'I'm glad that he married Maria.'
- (90) a. Das stimmt nicht.
that is.true not
'That isn't true.'
- b. Das gefällt mir.
that pleases me
'I like that.'

We cannot take the inflection of the finite verb as evidence of subjecthood: the verb in (89) is in 3rd person singular and this form is also used when there is no subject present:

- (91) dass gelacht wurde
that laughed was
'that there was laughing'

The *dass*-clauses in (89) could also be objects and the entire sentence a subjectless construction.

It is not possible to form imperatives, but this does not necessarily tell us anything about the subjecthood of the clausal argument since imperatives are aimed at an animate addressee or a machine, whereas clausal arguments refer to situations.

Eisenberg (1994b: 285) offers the following examples, which supposedly show that sentences can take the place of subjects in a subordinate infinitival clause:

- (92) a. Daß du zu Hause bleibst, hilft nicht, die Startbahn zu verhindern.
that you at home stays helps not the runway to hinder
'The fact that your staying at home won't do anything to stop it happening.'

- b. Daß du sprichst, verdient erwähnt zu werden.
that you speak deserves mentioned to become
'The fact that you're talking deserves to be mentioned.'

The infinitives in (92) correspond to the sentences with the finite verb in (93):

- (93) a. Daß du zu Hause bleibst, verhindert die Startbahn.
that you at home stay blocks the runway
'The fact that you're staying at home, stops anything happening.'
- b. Daß du sprichst, wird erwähnt.
that you speak becomes mentioned
'The fact that you're talking is being mentioned.'

Things are not that simple, however, as it is possible to use the demonstrative pronoun *das* in place of the *dass*-clauses in (93). If we assume that the unexpressed subject of an infinitive corresponds to a pronoun which refers to an argument in the matrix clause, then the subject of the infinitive in (92) should correspond to a pronoun such as *das* and therefore be nominal (Reis 1982: 194).

We have seen that, for German, we can equate subjects with non-predicative nominatives. As was shown in the discussion of the Icelandic data, this is not appropriate for all languages.

It should be noted that the discussion of how to deal with clausal arguments is by no means over. As recent publications show, there is still some discussion in Lexical Function Grammar (see Chapter 7) (Dalrymple & Lødrup 2000; Berman 2003b; 2007; Alsina, Mohanan & Mohanan 2005; Forst 2006).

If we can be clear about what we want to view as a subject, then the definition of object is no longer difficult: objects are all other elements whose form is directly determined by a given head. As well as clausal objects, German has genitive, dative, accusative and prepositional objects:

- (94) a. Sie gedenken des Mannes.
they remember the.GEN man.GEN
'They remember the man.'
- b. Sie helfen dem Mann.
they help the.DAT man.DAT
'They are helping the man.'
- c. Sie kennen den Mann.
they know the.ACC man.ACC
'They know the man.'
- d. Sie denken an den Mann.
they think on the man
'They are thinking of the man.'

As well as defining objects by their case, it is commonplace to talk of *direct objects* and *indirect objects*. The direct object gets its name from the fact that – unlike the indirect object – the referent of a direct object is directly affected by the action denoted by the verb. With ditransitives such as the German *geben* ‘to give’, the accusative object is the direct object and the dative is the indirect object.

- (95) dass er dem Mann den Aufsatz gibt
 that he the.DAT man.DAT the.ACC essay.ACC gives
 ‘that he gives the man the essay’

For trivalent verbs (verbs taking three arguments), we see that the verb can take either an object in the genitive case (96a) or, for verbs with a direct objects in the accusative, a second accusative object (96b):

- (96) a. dass er den Mann des Mordes bezichtigte
 that he the.ACC man.ACC the.GEN murder.GEN accused
 ‘that he accused the man of murder’
 b. dass er den Mann den Vers lehrte
 that he the.ACC man.ACC the.ACC verse.ACC taught
 ‘He taught the man the verse.’

These kinds of objects are sometimes also referred to as indirect objects.

Normally, only those objects which can be passivized using *werden* ‘to be’ are classed as direct objects. This is important for theories such as LFG (see Chapter 7) since passivization is defined with reference to grammatical function. With two-place verbal predicates, the dative is not normally classed as a direct object (Cook 2006).

- (97) dass er dem Mann hilft
 that he the.DAT man.DAT helps
 ‘that he helps the man’

In many theories, grammatical function does not form a primitive component of the theory, but rather corresponds to positions in a tree structure. The direct object in German is therefore the object which is first combined with the verb in a configuration that is assumed to be the underlying structure of German sentences. The indirect object is the second object to be combined with the verb. On this view, the dative subject of *helfen* ‘to help’ would have to be viewed as a direct object.

In the following, I will simply refer to the case of objects and avoid using the terms direct object and indirect object.

In the same way as with subjects, we consider whether there are object clauses which match a certain case or direct/indirect object. If we assume that *dass du sprichst* ‘that you are speaking’ in (93b) is a subject, then the subordinate clause must be a direct object in (98):

- (98) Er erwähnt, dass du sprichst.
 he mentioned that you speak
 ‘He mentioned that you are speaking.’

In this case, we cannot really view the subordinate clause as the accusative object since it does not bear case. However, we can replace the sentence with an accusative-marked noun phrase:

- (99) Er erwähnt diesen Sachverhalt.
he mentions this.ACC matter
'He mentions this matter.'

If we want to avoid this discussion, we can simply call these arguments clausal objects.

1.7.2 Adverbials

Adverbials differ semantically from subjects and objects. They tell us something about the conditions under which an action or process takes place, or the way in which a certain state persists. In the majority of cases, adverbials are adjuncts, but there are – as we have already seen – a number of heads which also require adverbials. Examples of these are verbs such as *to be located* or *to make one's way*. For *to be located*, it is necessary to specify a location and for *to make one's way* a direction is needed. These kinds of adverbials are therefore regarded as arguments of the verb.

The term *adverbial* comes from that fact that adverbials are often adverbs. This is not the only possibility, however. Adjectives, participles, prepositional phrases, noun phrases and even sentences can be adverbials:

- (100) a. Er arbeitet sorgfältig.
he works carefully
b. Er arbeitet vergleichend.
he works comparatively
'He does comparative work.'
c. Er arbeitet in der Universität.
he works in the university
'He works at the university.'
d. Er arbeitet den ganzen Tag.
he works the whole day.ACC
'He works all day.'
e. Er arbeitet, weil es ihm Spaß macht.
he works because it him.DAT fun makes
'He works because he enjoys it.'

Although the noun phrase in (100d) bears accusative case, it is not an accusative object. *den ganzen Tag* 'the whole day' is a so-called temporal accusative. The occurrence of accusative in this case has to do with the syntactic and semantic function of the noun phrase, it is not determined by the verb. These kinds of accusatives can occur with a variety of verbs, even with verbs that do not normally require an accusative object:

- (101) a. Er schläft den ganzen Tag.
 he sleeps the whole day
 ‘He sleeps the whole day.’
 b. Er liest den ganzen Tag diesen schwierigen Aufsatz.
 he reads the.ACC whole.ACC day this.ACC difficult.ACC essay
 ‘He spends the whole day reading this difficult essay.’
 c. Er gibt den Armen den ganzen Tag Suppe.
 he gives the.DAT poor.DAT the.ACC whole.ACC day soup
 ‘He spends the whole day giving soup to the poor.’

The case of adverbials does not change under passivization:

- (102) a. weil den ganzen Tag gearbeitet wurde
 because the.ACC whole.ACC day worked was
 ‘because someone worked all day’
 b. *weil der ganze Tag gearbeitet wurde
 because the.NOM whole.NOM day worked was

1.7.3 Predicatives

Adjectives like those in (103a,b) as well as noun phrases such as *ein Lügner* ‘a liar’ in (103c) are counted as predicatives.

- (103) a. Klaus ist *klug*.
 Klaus is clever
 b. Er isst den Fisch *roh*.
 he eats the fish raw
 c. Er ist *ein Lügner*.
 he is a liar

In the copula construction in (103a,c), the adjective *klug* ‘clever’ and the noun phrase *ein Lügner* ‘a liar’ is an argument of the copula *sein* ‘to be’ and the depictive adjective in (103b) is an adjunct to *isst* ‘eats’.

For predicative noun phrases, case is not determined by the head but rather by some other element.¹⁷ In this way, the accusative in (104a) becomes nominative under passivization (104b):

¹⁷ There is some dialectal variation with regard to copula constructions: in standard German the case of the noun phrase with *sein* ‘to be’ is always nominative and does not change when embedded under *lassen* ‘to let’. According to Duden (1995: § 1259), in Switzerland the accusative form is common which one finds in examples such as (ii.a).

- (i) a. Ich bin dein Tanzpartner.
 I am your.NOM dancing.partner
 b. Der wüste Kerl ist ihr Komplize.
 the ugly guy is her.NOM accomplice

- (104) a. Sie nannte ihn einen Lügner.
 she called him.ACC a.ACC liar
 ‘She called him a liar.’
 b. Er wurde ein Lügner genannt.
 he.NOM was a.NOM liar called
 ‘He was called a liar.’

Only *ihn* ‘him’ can be described as an object in (104a). In (104b), *ihn* becomes the subject and therefore bears nominative case. *einen Lügner* ‘a liar’ refers to *ihn* ‘him’ or *er* ‘he’ and agrees in case with the noun over which it predicates. This is also referred to as *case agreement*.

For more on predicative constructions see Duden (2005: § 1206).

1.7.4 Valency classes

It is possible to divide verbs into subclasses depending on how many arguments they require and the properties these arguments are required to have. The classic division describes all verbs which have an object which becomes the subject under passivization as *transitive*. Examples of this are verbs such as *love* or *hit*. Intransitive verbs, on the other hand, are verbs which have either no object, or one that does not become the subject in passive sentences. Examples of this type of verb are *schlafen* ‘to sleep’, *helfen* ‘to help’, *gedenken* ‘to remember’. A subclass of transitive verbs are ditransitive verbs such as *geben* ‘to give’ and *zeigen* ‘to show’.

Unfortunately, this terminology is not always used consistently. Sometimes, two-place verbs with dative and genitive objects are also classed as transitive verbs. In this naming tradition, the terms intransitive, transitive and ditransitive are synonymous with one-place, two-place and three-place verbs.

The fact that this terminological confusion can lead to misunderstandings between even established linguistics was shown by Culicover and Jackendoff’s (2005: 59) criticism of Chomsky. Chomsky states that the combination of the English auxiliary *be* + verb with passive morphology can only be used for transitive verbs. Culicover and Jackendoff claim

-
- c. Laß den wüsten Kerl [...] meinetwegen ihr Komplize sein.
 Let the.ACC ugly.ACC guy for.all.I.care her.NOM accomplice be
 ‘Let’s assume that the ugly guy is her accomplice, for all I care.’ (Duden 1966: § 6925)

- d. Baby, laß mich dein Tanzpartner sein.
 Baby let me.ACC your.NOM dancing.partner be
 ‘Baby, let me be your dancing partner!’ (Funny van Dannen, Benno-Ohnesorg-Theater, Berlin, Volksbühne, 11.10.1995)

- (ii) a. Er läßt den lieben Gott ‘n frommen Mann sein.
 he lets the.ACC dear.ACC god a pious.ACC man be
 b. *Er läßt den lieben Gott ‘n frommer Mann sein.
 he lets the.ACC dear.ACC god a pious.NOM man be

that this cannot be true because there are transitive verbs such as *weigh* and *cost*, which cannot undergo passivization:

- (105) a. This book weighs ten pounds / costs ten dollars.
 b. * Ten pounds are weighed / ten dollar are cost by this book.

Culicover and Jackendoff use *transitive* in the sense of a verb requiring two arguments. If we only view those verbs whose object becomes the subject of a passive clause as transitive, then *weigh* and *cost* no longer count as transitive verbs and Culicover and Jackendoff's criticism no longer holds.¹⁸ The fact that noun phrases such as those in (105) are no ordinary objects can also be seen by the fact they cannot be replaced by pronouns. It is therefore not possible to ascertain which case they bear since case distinctions are only realized on pronouns in English. If we translate the English examples into German, we find accusative objects:

- (106) a. Das Buch kostete einen Dollar.
 the book costs one.ACC dollar
 'The book costs one dollar.'
 b. Das Buch wiegt einen Zentner.
 the book weighs one.ACC centner

In the following, I will use *transitive* in the former sense, that is for verbs with an object that becomes the subject when passivized (e. g. with *werden* in German). When I talk about the class of verbs that includes *helfen* 'to help', which takes a nominative and dative argument, and *schlagen* 'to hit', which takes a nominative and accusative argument, I will use the term two-place or bivalent verbs.

1.8 A topological model of the German clause

In this section, I will introduce the concept of so-called *topological fields* (*topologische Felder*). These will be used frequently in later chapters to discuss different parts of the German clause. One can find further, more detailed introductions to topology in Reis (1980), Höhle (1986b) and Askedal (1986). Wöllstein (2010) is a textbook about the topological field model.

1.8.1 The position of the verb

It is common practice to divide German sentences into three types pertaining to the position of the finite verb:

- verb-final clauses

¹⁸ Their criticism also turn out to be unjust even if one views transitives as being two-place predicates. If one claims that a verb must take at least two arguments to be able to undergo passivization, one is not necessarily claiming that all verbs taking two or more arguments have to allow passivization. The property of taking multiple arguments is a condition which must be fulfilled, but it is by no means the only one.

- verb-first (initial) clauses
- verb-second (V2) clauses

The following examples illustrate these possibilities:

- (107) a. (Peter hat erzählt,) dass er das Eis gegessen hat.
Peter has told that he the ice.cream eaten has
'Peter said that he has eaten the ice cream.'
- b. Hat Peter das Eis gegessen?
has Peter the ice.cream eaten
'Has Peter eaten the ice cream?'
- c. Peter hat das Eis gegessen.
Peter has the ice.cream eaten
'Peter has eaten the ice cream.'

1.8.2 The sentence bracket, prefield, middle field and postfield

We observe that the finite verb *hat* 'has' is only adjacent to its complement *gegessen* 'eaten' in (107a). In (107b) and (107c), the verb and its complement are separated, that is discontinuous. We can then divide the German clause into various sub-parts on the basis of these distinctions. In (107b) and (107c), the verb and the auxiliary form a "bracket" around the clause. For this reason, we call this the *sentence bracket* (*Satzklammer*). The finite verbs in (107b) and (107c) form the left bracket and the non-finite verbs form the right bracket. Clauses with verb-final order are usually introduced by conjunctions such as *weil* 'because', *dass* 'that' and *ob* 'whether'. These conjunctions occupy the same position as the finite verb in verb-initial or verb-final clauses. We therefore also assume that these conjunctions form the left bracket in these cases. Using the notion of the sentence bracket, it is possible to divide the structure of the German clause into the prefield (*Vorfeld*), middle field (*Mittelfeld*) and postfield (*Nachfeld*). The prefield describes everything preceding the left sentence bracket, the middlefield is the section between the left and right bracket and the postfield describes the position after the right bracket. Table 1.1 on the next page gives some examples of this.

The right bracket can contain multiple verbs and is often referred to as a *verbal complex* or *verb cluster*.

The exact position of question words and relative pronouns in the prefield will be discussed in the following section.

1.8.3 Assigning elements to fields

As the examples in Table 1.1 show, it is not required that all fields are always occupied. Even the left bracket can be empty if one opts to leave out the copula *sein* 'to be' such as in the following examples:

Table 1.1: Examples of how topological fields can be occupied

Prefield	Left bracket	Middle field	Right bracket	Postfield
Karl	schläft.			
Karl	hat		geschlafen.	
Karl	erkennt	Maria.		
Karl	färbt	den Mantel	um	den Maria kennt.
Karl	hat	Maria	erkannt.	
Karl	hat	Maria als sie aus dem Zug stieg sofort	erkannt.	als sie aus dem Zug stieg.
Karl	hat	Maria sofort	erkannt	
Karl	hat	Maria zu erkennen	behauptet.	
Karl	hat		behauptet	Maria zu erkennen.
	Schläft	Karl?		
	Schlaf!		auf!	
	Iss	jetzt dein Eis	gegessen.	
	Hat	er doch das ganze Eis alleine		
	weil	er das ganze Eis alleine	gegessen hat	ohne mit der Wimper zu zucken.
	weil	er das ganze Eis alleine	essen können will	ohne gestört zu werden.
wer		das ganze Eis alleine	gegessen hat	
der		das ganze Eis alleine	gegessen hat	
mit wem		du	geredet hast	
mit dem		du	geredet hast	

1 Introduction and basic terms

- (108) a. [...] egal, was noch passiert, der Norddeutsche
regardless what still happens the north.German
Rundfunk steht schon jetzt als Gewinner fest.¹⁹
broadcasting.company stands already now as winner PRT
'Regardless of what still may happen, the North German broadcasting
company is already the clear winner.'
- b. Interessant, zu erwähnen, daß ihre Seele völlig in Ordnung war.²⁰
interesting to mention that her soul completely in order was
'It is interesting to note that her soul was entirely fine.'
- c. Ein Treppenwitz der Musikgeschichte, daß die Kollegen von
an afterwit of.the.history.of.music that the colleagues of
Rammstein vor fünf Jahren noch im Vorprogramm von Sandow
Rammstein before five years still in.the.pre-programme of sandow
spielten.²¹
played
'One of the little ironies of music history is that five years ago their
colleagues of Rammstein were still an opening act for Sandow.'

The examples in (108) correspond to those with the copula in (109):

- (109) a. Was noch passiert, ist egal, ...
what still happens ist regardless
'It is not important what happens next.'
- b. Interessant ist zu erwähnen, dass ihre Seele völlig in Ordnung war.
interesting is to mention that her soul completely in order was
'It is interesting to note that her soul was completely fine.'
- c. Ein Treppenwitz der Musikgeschichte ist, dass die Kollegen von
an afterwit of.the.music.history is that the colleagues of
Rammstein vor fünf Jahren noch im Vorprogramm von Sandow
Rammstein before five years still in pre.programme of Sandow
spielten.
played
'It is one of the little ironies of music history that five years ago the
Rammstein boys were still an opening act for Sandow.'

When certain fields are empty, it is sometimes not clear which fields are occupied by certain constituents. For the examples in (108), one would have to insert the copula to be able to ascertain whether in fact a single constituent is in the prefield and, furthermore, which fields the rest of the constituents occupy.

¹⁹ Spiegel, 12/1999, p. 258.

²⁰ Michail Bulgakow, *Der Meister und Margarita*. München: Deutscher Taschenbuch Verlag. 1997, p. 422.

²¹ Flüstern & Schweigen, taz, 12.07.1999, p. 14.

In the following example taken from Paul (1919: 13), inserting the copula obtains a different result:

- (110) a. Niemand da?
 nobody there
 b. Ist niemand da?
 is nobody there
 ‘Is nobody there?’

Here we are dealing with a question and *niemand* ‘nobody’ in (110a) should therefore not be analyzed as in the prefield but rather the middle field.

In (111), there are elements in the prefield, the left bracket and the middle field. The right bracket is empty.

- (111) Er gibt der Frau das Buch, die er kennt.
 he gives the woman the book that he knows
 ‘He gives the book to the woman that he knows.’

How should we analyse relative clauses such as *die er kennt* ‘that he knows’? Do they form part of the middle field or the postfield? This can be tested using a test discussed in (Bech 1955: 72) (*Rangprobe*): First, we modify the example in (111) so that it is in the perfect. Since non-finite verb forms occupy the right bracket, we can clearly see the border between the middle field and postfield. The examples in (112) show that the relative clause cannot occur in the *Mittelfeld* unless it is part of a complex constituent with the head noun *Frau* ‘woman’.

- (112) a. Er hat [der Frau] das Buch gegeben, [die er kennt].
 he has the woman the book given that he knows
 b. *Er hat [der Frau] das Buch, [die er kennt,] gegeben.
 he has the woman the book that he knows given
 c. Er hat [der Frau, die er kennt,] das Buch gegeben.
 he has the woman that he knows the book given

This test does not help if the relative clause is realized together with its head noun at the end of the sentence as in (113):

- (113) Er gibt das Buch der Frau, die er kennt.
 he gives the book the woman that he knows
 ‘He gives the book to the woman that he knows.’

If we put the example in (113) in the perfect, then we observe that the lexical verb can occur before or after the relative clause:

- (114) a. Er hat das Buch [der Frau] gegeben, [die er kennt].
 he has the book the woman given that he knows

- b. Er hat das Buch [der Frau, die er kennt,] gegeben.
 he has the book the woman that he knows given
 ‘He has given the book to the woman he knows.’

In (114a), the relative clause has been extraposed. In (114b) it forms part of the noun phrase *der Frau, die er kennt* ‘the woman that he knows’ and therefore occurs inside the NP in the middle field. It is therefore not possible to rely on this test for (113). We assume that the relative clause in (113) also belongs to the NP since this is the most simple structure. If the relative clause were in the postfield, we would have to assume that it has undergone extraposition from its position inside the NP, that is, we would have to assume the NP-structure anyway and then extraposition follows this.

We have a similar problem with interrogative and relative pronouns. Depending on the author, these are assumed to be in the left bracket (Kathol 2001; Eisenberg 2004: 403) or the prefield (Duden 2005: §1345; Wöllstein 2010: 29–30, Section 3.1) or even in the middle field (Altmann & Hofman 2004: 75). In Standard German interrogative or relative clauses, both fields are never simultaneously occupied. For this reason, it is not immediately clear to which field an element belongs. Nevertheless, we can draw parallels to main clauses: the pronouns in interrogative and relative clauses can be contained inside complex phrases:

- (115) a. der Mann, [mit dem] du gesprochen hast
 the man with that you spoken have
 ‘The man you spoke to.’
 b. Ich möchte wissen, [mit wem] du gesprochen hast
 I would.like know with whom you spoken have
 ‘I want to know who spoke to.’

Normally, only individual words (conjunctions or verbs) can occupy the left bracket,²² whereas words and phrases can appear in the prefield. It therefore makes sense to assume that interrogative and relative pronouns (and phrases containing them) also occur in this position.

Furthermore it can be observed that the dependency between the elements in the *Vorfeld* of declarative clauses and the remaining sentence is of the same kind as the dependency between the phrase that contains the relative pronoun and the remaining sentence. For instance, *über dieses Thema* in (116a) ‘about this topic’ depends on *Vortrag* ‘talk’, which is deeply embedded in the sentence: *einen Vortrag* ‘a talk’ is an argument of *zu halten* ‘to hold’, which in turn is an argument of *gebeten* ‘asked’.

²² Coordination is an exception to this:

- (i) Er [kennt und liebt] diese Schallplatte.
 he knows and loves this record
 ‘He knows and loves this record.’

- (116) a. Über dieses Thema habe ich ihn gebeten, einen Vortrag zu halten.
 about this topic have I him asked a talk to hold
 ‘I asked him to give a talk about this topic.’
 b. das Thema, über das ich ihn gebeten habe, einen Vortrag zu halten
 the topic about which I him asked have a talk to hold
 ‘the topic about which I asked him to give a talk’

The situation is similar in (116b): The relative phrase *über das* ‘about which’ is a dependent of *Vortrag* ‘talk’ which is realized far away from it. Thus, if the relative phrase is assigned to the *Vorfeld*, it is possible to say that such nonlocal frontings always target the *Vorfeld*.

Finally, the Duden (2005: §1347) provides the following examples from non-standard German (mainly southern dialects):

- (117) a. Kommt drauf an, mit wem dass sie zu tun haben.
 comes there.upon PART with whom that you to do have
 ‘It depends on to whom you talk to.’
 (118) a. Lotti, die wo eine tolle Sekretärin ist, hat ein paar merkwürdige
 Lotti who where a great secretary is has a few strange
 Herren empfangen.
 gentlemen welcomed
 ‘Lotti, who is a great secretary welcomed a few strange gentlemen.’
 b. Du bist der beste Sänger, den wo ich kenn.
 you are the best singer who where I know
 ‘You are the best singer who I know.’

These examples of interrogative and relative clauses show that the left sentence bracket is filled with a conjunction (*dass* ‘that’ or *wo* ‘where’ in the respective dialects. So if one wants to have a model that treats Standard German and the dialectal forms uniformly, it is reasonable to assume the relative phrases and interrogative phrases are located in the *Vorfeld*.

1.8.4 Recursion

As already noted by Reis (1980: 82), when occupied by a complex constituent, the prefield can be subdivided into further fields including a postfield, for example. The constituents *für lange lange Zeit* ‘for a long, long time’ in (119b) and *daß du kommst* ‘that you are coming’ in (119d) are inside the prefield but occur to the right of the right bracket *verschüttet* ‘spilled’ / *gewußt* ‘knew’, that is they are in the postfield of the prefield.

- (119) a. Die Möglichkeit, etwas zu verändern, ist damit verschüttet für
 the possibility something to change is there.with spilled for
 lange lange Zeit.
 long long time

1 Introduction and basic terms

- b. [Verschüttet für lange lange Zeit] ist damit die Möglichkeit, etwas
spilled for long long time ist there.with the possibility something
zu verändern.
to change
'The possibility to change something will now be gone for a long, long time.'
- c. Wir haben schon seit langem gewußt, daß du kommst.
we have PRT since long known that you come
- d. [Gewußt, daß du kommst,] haben wir schon seit langem.
known that you come have we PRT since long
'We've known for a while that you are coming.'

Like constituents in the prefield, elements in the middle field and postfield can also have an internal structure and be divided into subfields accordingly. For example, *daß* 'that' is the left bracket of the subordinate clause *daß du kommst* in (119c), whereas *du* 'you' occupies the middle field and *kommst* 'come' the right bracket.

Comprehension questions

1. How does the head of phrase differ from other non-heads?
2. What is the head in the examples in (120)?
(120) a. he
b. Go!
c. quick
3. How do arguments differ from adjuncts?
4. Identify the heads, arguments and adjuncts in the following sentence (121) and in the subparts of the sentence:
(121) Er hilft den kleinen Kindern in der Schule.
he helps the small children in the school
'He helps small children at school'
5. How can we define the terms prefield (*Vorfeld*), middle field (*Mittelfeld*), postfield (*Vorfeld*) and the left and right sentence brackets (*Satzklammer*)?

Exercises

1. Identify the sentence brackets, prefield, middle field and postfield in the following sentences. Do the same for the embedded clauses!

- (122) a. Karl isst.
Karl eats
'Karl is eating.'
- b. Der Mann liebt eine Frau, den Peter kennt.
the man loves a woman that Peter knows
'The man, who Peter knows, loves a woman'
- c. Der Mann liebt eine Frau, die Peter kennt.
the man loves a woman, that Peter knows
'The man loves a woman, who Peter knows'
- d. Die Studenten haben behauptet, nur wegen der Hitze
the students have claimed only because.of the heat
einzuschlafen.
to.fall.asleep
'The students claimed that they were only falling asleep because of the heat.'
- e. Dass Peter nicht kommt, ärgert Klaus.
that Peter not comes annoys Klaus
'(The fact) that Peter isn't coming annoys Klaus.'
- f. Einen Mann küssen, der ihr nicht gefällt, würde sie nie.
a man kiss that her not pleases would she never
'She would never kiss a man she doesn't like.'

Further reading

Reis (1980) gives reasons for why field theory is important for the description of the position of constituents in German.

Höhle (1986a) discusses fields to the left of the prefield, which are needed for left-dislocation structures such as with *der Mittwoch* in (123), *aber* in (124a) and *denn* in (124b):

- (123) Der Mittwoch, der passt mir gut.
the Wednesday that fits me good
'Wednesday, that suits me.'
- (124) a. Aber würde denn jemand den Hund füttern morgen abend?
but would PRT anybody the dog feed tomorrow evening
'But would anyone feed the dog tomorrow evening?'
- b. Denn dass es regnet, damit rechnet keiner.
because that it rains there.with reckons nobody
'Because no-one expects that it will rain.'

1 Introduction and basic terms

Höhle discusses the historical development of field theory.

2 Phrase Structure Grammar

This chapter deals with phrase structure grammars, which play an important role in several of the theories we will encounter in later chapters.

2.1 Symbols and rewrite rules

Words can be assigned to a particular part of speech on the basis of their inflectional properties and syntactic distribution. Thus, *weil* ‘because’ in (1) is a conjunction, whereas *das* ‘the’ and *dem* ‘the’ are articles and therefore classed as determiners. Furthermore, *Buch* ‘book’ and *Mann* ‘man’ are nouns and *gibt* ‘gives’ is a verb.

- (1) weil er das Buch dem Mann gibt
because he the book the man gives
‘because he gives the man the book’

Using the constituency tests we introduced in Section 1.3, we can show that individual words as well as the strings *das Buch* ‘the book’ and *dem Mann* ‘the man’, form constituents. These are then assigned to certain symbols. Since nouns form an important part of the phrases *das Buch* and *dem Mann*, these are referred to as *noun phrases* or NPs, for short. The pronoun *er* ‘he’ can occur in the same positions as full NPs and is therefore also assigned to the category NP.

Phrase structure grammars come with rules specifying which symbols are assigned to certain kinds of words and how these are combined to create more complex units. A simple phrase structure grammar which can be used to analyze (1) is given in (2):^{1,2}

- | | | |
|----------------|---------|----------|
| (2) NP → D N | NP → er | N → Buch |
| S → NP NP NP V | D → das | N → Mann |
| | D → dem | V → gibt |

We can therefore interpret a rule such as NP → D N as meaning that a noun phrase, that is, something which is assigned the symbol NP, can consist of a determiner (D) and a noun (N).

¹ We will ignore the conjunction *weil* ‘because’ for now. Since the exact analysis for German verb-first and verb-second clauses requires a number of additional assumptions, we will restrict ourselves to verb-final clauses in this chapter.

² The rule NP → er may seem odd. We could assume the rule PersPron → er instead but then would have to posit a further rule which would specify that personal pronouns can replace full NPs: NP → PersPron. The rule in (2) combines the two aforementioned rules and states that *er* ‘he’ can occur in positions where noun phrases can.

We can analyze the sentence in (1) using the grammar in (2) in the following way: First, we take the first word in the sentence and check if there is a rule in which this word occurs on the right-hand side of the rule. If this is the case, then we replace the word with the symbol on the left-hand side of the rule. This happens in lines 2–4, 6–7 and 9 of the derivation in (3). For instance, in line 2 *er* is replaced by NP. If there are two or more symbols which occur together on the right-hand side of a rule, then we replace all these words with the symbol on the left. This happens in lines 5, 8 and 10. For instance, the D and the N are combined into NP. In (3), we began with a string of words

(3)	words and symbols						rules that are applied
1	er	das	Buch	dem	Mann	gibt	
2	NP	das	Buch	dem	Mann	gibt	NP → er
3	NP	D	Buch	dem	Mann	gibt	D → das
4	NP	D	N	dem	Mann	gibt	N → Buch
5	NP		NP	dem	Mann	gibt	NP → D N
6	NP		NP	D	Mann	gibt	D → dem
7	NP		NP	D	N	gibt	N → Mann
8	NP		NP		NP	gibt	NP → D N
9	NP		NP		NP	V	V → gibt
10						S	S → NP NP NP V

and it was shown that we can derive the structure of a sentence by applying the rules of a given phrase structure grammar. We could have applied the same steps in reverse order: starting with the sentence symbol S, we would have applied the steps 9–1 and arrived at the string of words. Selecting different rules from the grammar for rewriting symbols, we could use the grammar in (2) to get from S to the string *er dem Mann das Buch gibt* ‘he the man the book gives’. We can say that this grammar licenses (or generates) a set of sentences.

The derivation in (3) can also be represented as a tree. This is shown by Figure 2.1. The

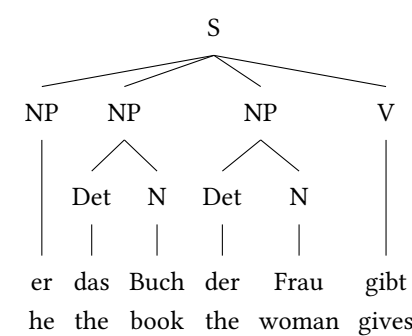


Figure 2.1: Analysis of *er das Buch dem Mann gibt* ‘he the book the woman gives’

symbols in the tree are called *nodes*. We say that S immediately dominates the NP nodes and the V node. The other nodes in the tree are also dominated, but not immediately dominated, by S. If we want to talk about the relationship between nodes, it is common to use kinship terms. In Figure 2.1, S is the mother node of the three NP nodes and the V node. The NP node and V are sisters since they have the same mother node. If a node has two daughters, then we have a binary branching structure. If there is exactly one daughter, then we have a unary branching structure. Two constituents are said to be *adjacent* if there are directly next to each other.

Phrase structure rules are often omitted in linguistic publications. Instead, authors opt for tree diagrams or the compact equivalent bracket notation such as (4).

- (4) [S [NP er] [NP [D das] [N Buch]] [NP [D dem] [N Mann]] [V gibt]]
 he the book the man gives

Nevertheless, it is the grammatical rules which are actually important since these represent grammatical knowledge which is independent of specific structures. In this way, we can use the grammar in (2) to parse or generate the sentence in (5), which differs from (1) in the order of objects:

- (5) [weil] er dem Mann das Buch gibt
 because he.NOM the man.DAT the book.ACC gives
 ‘because he gives the man the book’

The rules for replacing determiners and nouns are simply applied in a different order than in (1). Rather than replacing the first Det with *das* ‘the’ and the first noun with *Buch* ‘book’, the first Det is replaced with *dem* ‘the’ and the first noun with *Mann*.

At this juncture, I should point out that the grammar in (2) is not the only possible grammar for the example sentence in (1). There is an infinite number of possible grammars which could be used to analyze these kinds of sentences (see exercise 1). Another possible grammar is given in (6):

- (6) NP → D N NP → er N → Buch
 V → NP V D → das N → Mann
 D → dem V → gibt

This grammar licenses binary branching structures as shown in Figure 2.2 on the following page.

Both the grammar in (6) and (2) are too imprecise. If we adopt additional lexical entries for *ich* ‘I’ and *den* ‘the’ (accusative) in our grammar, then we would incorrectly license the ungrammatical sentences in (7b–d):³

³ With the grammar in (6), we also have the additional problem that we cannot determine when an utterance is complete since the symbol V is used for all combinations of V and NP. Therefore, we can also analyze the sentence in (i) with this grammar:

- (i) a. * der Mann erwartet
 the man expects

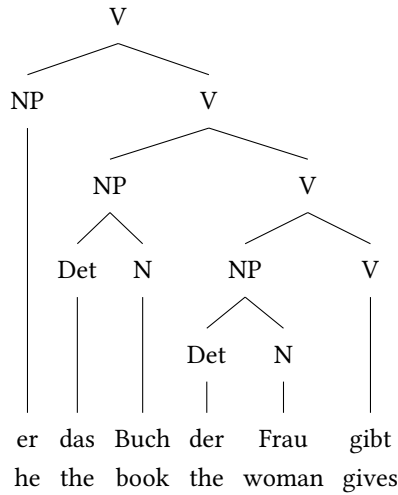


Figure 2.2: Analysis of *er das Buch dem Mann gibt* with a binary branching structure

- (7) a. er das Buch dem Mann gibt
he.NOM the book.ACC the man.DAT gives
'He gives the book to the man.'
- b. *ich das Buch dem Mann gibt
I.NOM the book.ACC the man.DAT gives
- c. *er das Buch den Mann gibt
he.NOM the book.ACC the.ACC man gives
- d. *er den Buch dem Mann gibt
he.NOM the.M book.N the man gives

In (7b), subject-verb agreement has been violated, in other words: *ich* 'I' and *gibt* 'gives' do not fit together. (7c) is ungrammatical because the case requirements of the verb have not been satisfied: *gibt* 'gives' requires a dative object. Finally, (7d) is ungrammatical because there is a lack of agreement between the determiner and the noun. It is not possible to combine *den* 'the', which is masculine and bears accusative case, and *Buch* 'book' because *Buch* is neuter gender. For this reason, the gender properties of these two elements are not the same and can therefore not be combined.

-
- b. *des Mannes er das Buch dem Mann gibt
the.GEN man.GEN he the.DAT book.DAT the man gives

The number of arguments required by a verb must be somehow represented in the grammar. In the following chapters, we will see exactly how the selection of arguments by a verb (valency) can be captured in various grammatical theories.

In the following, we will consider how we would have to change our grammar to stop it from licensing the sentences in (7b–d). If we want to capture subject-verb agreement, then we have to cover the following six cases since in German, as the verb has to agree with the subject in both person (1, 2, 3) and number (sg, pl):

- (8) a. Ich schlafe. (1, sg)
 I sleep
 b. Du schläfst. (2, sg)
 you sleep
 c. Er schläft. (3, sg)
 he sleeps
 d. Wir schlafen. (1, pl)
 we sleep
 e. Ihr schlaft. (2, pl)
 you sleep
 f. Sie schlafen. (3, pl)
 they sleep

It is possible to capture these relations with grammatical rules by increasing the number of symbols we use. Instead of the rule $S \rightarrow NP\ NP\ NP\ V$, we can use the following:

- (9) $S \rightarrow NP_1_sg\ NP\ NP\ V_1_sg$
 $S \rightarrow NP_2_sg\ NP\ NP\ V_2_sg$
 $S \rightarrow NP_3_sg\ NP\ NP\ V_3_sg$
 $S \rightarrow NP_1_pl\ NP\ NP\ V_1_pl$
 $S \rightarrow NP_2_pl\ NP\ NP\ V_2_pl$
 $S \rightarrow NP_3_pl\ NP\ NP\ V_3_pl$

This would mean that we need six different symbols for noun phrases and verbs respectively, as well as six rules rather than one.

In order to account for case assignment by the verb, we can incorporate case information into the symbols in an analogous way. We would then get rules such as the following:

- (10) $S \rightarrow NP_1_sg_nom\ NP_dat\ NP_acc\ V_1_sg_nom_dat_acc$
 $S \rightarrow NP_2_sg_nom\ NP_dat\ NP_acc\ V_2_sg_nom_dat_acc$
 $S \rightarrow NP_3_sg_nom\ NP_dat\ NP_acc\ V_3_sg_nom_dat_acc$
 $S \rightarrow NP_1_pl_nom\ NP_dat\ NP_acc\ V_1_pl_nom_dat_acc$
 $S \rightarrow NP_2_pl_nom\ NP_dat\ NP_acc\ V_2_pl_nom_dat_acc$
 $S \rightarrow NP_3_pl_nom\ NP_dat\ NP_acc\ V_3_pl_nom_dat_acc$

Since it is necessary to differentiate between noun phrases in four cases, we have a total of six symbols for NPs in the nominative and three symbols for NPs with other cases. Since verbs have to match the NPs, that is, we have to differentiate between verbs which

select three arguments and those selecting only one or two (11), we have to increase the number of symbols we assume for verbs.

- (11) a. Er schläft.
 he sleeps
 ‘He is sleeping.’
 b. * Er schläft das Buch.
 he sleeps the book
 c. Er kennt das Buch.
 he knows the book
 ‘He knows the book.’
 d. * Er kennt.
 he knows

In the rules above, the information about the number of arguments required is included in the marking ‘nom_dat_acc’.

In order to capture the determiner-noun agreement in (12), we have to incorporate information about gender (fem, mas, neu), number (sg, pl), case (nom, gen, dat, acc) and the inflectional classes (strong, weak⁴).

- (12) a. der Mann, die Frau, das Buch (gender)
 the man the woman the book
 b. das Buch, die Bücher (number)
 the book the books
 c. des Buches, dem Buch (case)
 the.GEN book the.DAT book
 d. ein Beamter, der Beamte (inflectional class)
 a civil.servant the civil.servant

Instead of the rule $NP \rightarrow D N$, we will have to use rules such as those in (13):

- (13) $NP_3_sg_nom \rightarrow D_fem_sg_nom_weak N_fem_sg_nom_weak$
 $NP_3_sg_nom \rightarrow D_mas_sg_nom_weak N_mas_sg_nom_weak$
 $NP_3_sg_nom \rightarrow D_neu_sg_nom_weak N_neu_sg_nom_weak$
 $NP_3_pl_nom \rightarrow D_fem_pl_nom_weak N_fem_pl_nom_weak$
 $NP_3_pl_nom \rightarrow D_mas_pl_nom_weak N_mas_pl_nom_weak$
 $NP_3_pl_nom \rightarrow D_neu_pl_nom_weak N_neu_pl_nom_weak$

⁴ These are inflectional classes for adjectives which are also relevant for some nouns such as *Beamter* ‘civil servant’, *Verwandter* ‘relative’, *Gesandter* ‘envoy’, ...). For more on adjective classes see page 21.

NP_3_sg_nom → D_fem_sg_nom_stark N_fem_sg_nom_stark
 NP_3_sg_nom → D_mas_sg_nom_stark N_mas_sg_nom_stark
 NP_3_sg_nom → D_neu_sg_nom_stark N_neu_sg_nom_stark
 NP_3_pl_nom → D_fem_pl_nom_stark N_fem_pl_nom_stark
 NP_3_pl_nom → D_mas_pl_nom_stark N_mas_pl_nom_stark
 NP_3_pl_nom → D_neu_pl_nom_stark N_neu_pl_nom_stark

(13) shows the rules for nominative noun phrases. We would need analogous rules for genitive, dative, and accusative. We would then require 48 symbols for determiners ($3 \cdot 2 \cdot 4 \cdot 2$), 48 symbols for nouns and 48 rules rather than one.

2.2 Expanding PSG with features

Phase structure grammars which only use atomic symbols are problematic as they cannot capture certain generalizations. We as linguists can recognize that NP_3_sg_nom stands for a noun phrase because it contains the letters NP. However, in formal terms this symbol is just like any other symbol in the grammar and we cannot capture the commonalities of all the symbols used for NPs. Furthermore, unstructured symbols do not capture the fact that the rules in (13) all have something in common. In formal terms, the only thing that the rules have in common is that there is one symbol on the left-hand side of the rule and two on the right.

We can solve this problem by introducing features which are assigned to category symbols and therefore allow for the values of such features to be included in our rules. For example, we can assume the features person, number and case for the category symbol NP. For determiners and nouns, we would adopt an additional feature for gender and one for inflectional class.

(14) NP(3,sg,nom) → D(fem,sg,nom,strong) N(fem,sg,nom,strong)
 NP(3,sg,nom) → D(mas,sg,nom,strong) N(mas,sg,nom,strong)

If we were to use variables rather than the values in (14), we would get the following rules as in (15):

(15) NP(3,Num,Case) → D(Gen,Num,Case,Infl) N(Gen,Num,Case,Infl)

The values of the variables here are not important. What is important is that they match. The value of the person feature (the first position in the NP(3,Num,Case)) is fixed at '3' by the rule. These kind of restrictions on the values can, of course, be determined in the lexicon:

(16) NP(3,sg,nom) → es
 D(mas,sg,nom,strong) → des

The rules in (10) can be summarized as in (17):

- (17) $S \rightarrow NP(\text{Per1}, \text{Num1}, \text{nom})$
 $NP(\text{Per2}, \text{Num2}, \text{dat})$
 $NP(\text{Per3}, \text{Num3}, \text{acc})$
 $V(\text{Per1}, \text{Num1}, \text{ditransitive})$

Due to Per1 and Num1 on the verb and subject, we will therefore ensure that there is subject-verb agreement. For the other NPs, the values of these features are irrelevant since the case of these NPs is explicitly determined.

2.3 Semantics

In the introductory chapter and the previous sections, we have been dealing with syntactic aspects of language and the focus will remain very much on syntax for the remainder of this book. It is, however, important to remember that we use language to communicate, that is, to transfer information about certain situations, topics or opinions. If we want to accurately explain our capacity for language, then we also have to explain the meanings that our utterances have. To this end, it is necessary to understand their syntactic structure, but this alone is not enough. Furthermore, theories of language acquisition that only concern themselves with the acquisition of syntactic constructions are also inadequate. The syntax-semantics interface is therefore important and every grammatical theory has to say something about how syntax and semantics interact. In the following, I will show how we can combine phrase structure rules with semantic information. To represent meanings, I will use first-order predicate logic and λ -calculus. Unfortunately, it is not possible to provide a detailed discussion of the basics of logic so that even readers without prior knowledge can follow all the details, however, the simple examples discussed here should be enough to provide some initial insights into how syntax and semantics interact and furthermore, how we can develop a linguistic theory to account for this.

To show how the meaning of a sentence is derived from the sum of its parts, we will consider (18a). We assign the meaning in (18b) to the sentence in (18a).

- (18) a. Max schläft.
 Max sleeps
 ‘Max is sleeping.’
 b. $\text{schlafen}'(\text{max}')$

Here, we are assuming $\text{schlafen}'$ to be the meaning of schläft ‘sleeps’. We use prime symbols to indicate that we are dealing with word meanings and not actual words. At first glance, it may not seem that we have really gained anything by using $\text{schlafen}'$ to represent the meaning of (18a), since it is just another form of the verb schläft ‘sleeps’. It is, however, important to concentrate on a single verb form as inflection is irrelevant when it comes to meaning. We can see this by comparing the examples in (19a) and (19b):

- (19) a. Jeder Junge schläft.
 every boy sleeps
 ‘Every boy sleeps.’
 b. Alle Jungen schlafen.
 all boys sleep
 ‘All boys sleep.’

When looking at the meaning in (18b), we can consider which part of the meaning comes from each word. It seems relatively intuitive that *max'* comes from *Max*, but the trickier question is what exactly *schläft* ‘sleeps’ contributes in terms of meaning. If we think about what characterizes a ‘sleeping’ event, we know that there is typically an individual who is sleeping. This information is part of the meaning of the verb *schlafen* ‘to sleep’. The verb meaning does not contain information about the sleeping individual, however, as this verb can be used with various subjects:

- (20) a. Paul schläft.
 Paul sleeps
 ‘Paul is sleeping.’
 b. Mio schläft.
 Mio sleeps
 ‘Mio is sleeping.’
 c. Xaver schläft.
 Xaver sleeps
 ‘Xaver is sleeping.’

We can therefore abstract away from any specific use of *schlafen'* and instead of, for example, *max'* in (18b), we use a variable (e. g. *x*). This *x* can then be replaced by *paul'*, *mio'* or *xaver'* in a given sentence. To allow us to access these variables in a given meaning, we can write them with a λ in front. Accordingly, *schläft* ‘sleeps’ will have the following meaning:

- (21) $\lambda x \text{ sleep}'(x)$

The step from (18b) to (21) is referred to as *lambda abstraction*. The combination of the expression (21) with the meaning of its arguments happens in the following way: we remove the λ and the corresponding variable and then replace all instances of the variable with the meaning of the argument. If we combine (21) and *max'* as in (22), we arrive at the meaning in (18b).

- (22) $\lambda x \text{ sleep}'(x) \text{ max}'$

The process is called β -reduction or λ -conversion. To show this further, let us consider an example with a transitive verb. The sentence in (23a) has the meaning given in (23b):

- (23) a. Max mag Lotte.
 Max likes Lotte
 ‘Max likes Lotte.’
 b. $like'(max', lotte')$

The λ -abstraction of *mag* ‘likes’ is shown in (24):

- (24) $\lambda y \lambda x like'(x, y)$

Note that it is always the first λ that has to be used first. The variable y corresponds to the object of *mögen*. For languages like English it is assumed that the object forms a verb phrase (VP) together with the verb and this VP is combined with the subject. German differs from English in allowing more freedom in constituent order. The problems that result for form meaning mappings are solved in different ways by different theories. The respective solutions will be addressed in the following chapters.

If we combine the representation in (24) with that of the object *Lotte*, we arrive at (25a), and following β -reduction, (25b):

- (25) a. $\lambda y \lambda x like'(x, y) lotte'$
 b. $\lambda x like'(x, lotte')$

This meaning can in turn be combined with the subject and we then get (26a) and (26b) after β -reduction:

- (26) a. $\lambda x like'(x, lotte') max'$
 b. $like'(max', lotte')$

After introducing lambda calculus, integrating the composition of meaning into our phrase structure rules is simple. A rule for the combination of a verb with its subject has to be expanded to include positions for the semantic contribution of the verb, the semantic contribution of the subject and then the meaning of the combination of these two (the entire sentence). The complete meaning is the combination of the individual meanings in the correct order. We can therefore take the simple rule in (27a) and turn it into (27b):

- (27) a. $S \rightarrow NP(nom) V$
 b. $S(V' NP') \rightarrow NP(nom, NP') V(V')$

V' stands for the meaning of V and NP' the meaning of the $NP(nom)$. $V' NP'$ stands for the combination of V' and NP' . When analyzing (18a), the meaning of V' is $\lambda x schlafen(x)$ and the meaning of NP' is max' . The combination of $V' NP'$ corresponds to (28a) or after β -reduction to (28b) – repeated here as (28b):

- (28) a. $\lambda x sleep'(x) max'$
 b. $sleep'(max')$

For the example with a transitive verb in (23a), the rule in (29) can be proposed:

(29) $S(V' NP2' NP1') \rightarrow NP(nom, NP1') V(V') NP(acc, NP2')$

The meaning of the verb (V') is first combined with the meaning of the object ($NP2'$) and then with the meaning of the subject ($NP1'$).

At this point, we can see that there are several distinct semantic rules for the phrase structure rules above. The hypothesis that we should analyze language in this way is called the 'rule-by-rule hypothesis' (Bach 1976). A more general process for deriving the meaning of linguistic expression will be presented in Section 5.1.4.

2.4 Phase structure rules for some aspects of German syntax

Whereas determining the direct constituents of a sentence is relative easy, since we can very much rely on the movement test due to the somewhat flexible order of constituents in German, it is more difficult to identify the parts of the noun phrase. This is the problem we will focus on in this section. To help motivate assumptions about \bar{X} syntax to be discussed in Section 2.5, we will also discuss prepositional phrases.

2.4.1 Noun phrases

Up to now, we have assumed a relatively simple structure for noun phrases: Our rules state that a noun phrase consists of a determiner and a noun. Noun phrases can have a distinctly more complex structure than (30a). This is shown by the following examples in (30):

- (30)
- a. eine Frau
a woman
 - b. eine Frau, die wir kennen
a woman who we know
 - c. eine Frau aus Stuttgart
a woman from Stuttgart
 - d. eine kluge Frau
a smart woman
 - e. eine Frau aus Stuttgart, die wir kennen
a woman from Stuttgart who we know
 - f. eine kluge Frau aus Stuttgart
a smart woman from Stuttgart
 - g. eine kluge Frau, die wir kennen
a smart woman who we know
 - h. eine kluge Frau aus Stuttgart, die wir kennen
a smart woman from Stuttgart who we know

As well as determiners and nouns, noun phrases can also contain adjectives, prepositional phrases and relative clauses. The additional elements in (30) are adjuncts. They restrict the set of objects which the noun phrase refers to. Whereas (30a) refers to a being which has the property of being a woman, the referent of (30b) must also have the property of being known to us.

Our previous rules for noun phrases simply combined a noun and a determiner and can therefore only be used to analyze (30a). The question facing us now is how we can modify this rule or which additional rules we would have to assume in order to analyze the other noun phrases in (30). In addition to rule (31a), one could propose a rule such as the one in (31b).^{5,6}

- (31) a. $NP \rightarrow Det\ N$
 b. $NP \rightarrow Det\ A\ N$

However, this rule would still not allow us to analyze noun phrases such as (32):

- (32) alle weiteren schlagkräftigen Argumente
 all further strong arguments
 ‘all other strong arguments’

In order to be able to analyze (32), we require a rule such as (33):

- (33) $NP \rightarrow Det\ A\ A\ N$

It is always possible to increase the number of adjectives in a noun phrase and setting an upper limit for adjectives would be entirely arbitrary. Even if we opt for the following abbreviation, there are still problems:

- (34) $NP \rightarrow Det\ A^*\ N$

The asterisk in (34) stands for any number of iterations. Therefore, (34) encompasses rules with no adjectives as well as those with one, two or more.

The problem is that according to the rule in (34) adjectives and nouns do not form a constituent and we can therefore not explain why coordination is still possible in (35):

- (35) alle [[geschickten Kinder] und [klugen Frauen]]
 all smart children and smart women
 ‘all the smart children and smart women’

If we assume that coordination involves the combination of two or more word strings with the same syntactic properties, then we would have to assume that the adjective and noun form a unit.

The noun phrases with adjectives discussed thus far can be explained by the following rules:

⁵ See Eisenberg (2004: 238) for the assumption of flat structures in noun phrases.

⁶ There are, of course, other features such as gender and number, which should be part all the rules discussed in this section. I have omitted these in the following due for ease of exposition.

- (36) a. $NP \rightarrow Det \bar{N}$
 b. $\bar{N} \rightarrow A \bar{N}$
 c. $\bar{N} \rightarrow N$

These rules state the following: A noun phrase consists of a determiner and a nominal element (\bar{N}). This nominal element can consist of an adjective and a nominal element (36b), or just a noun (36c). Since \bar{N} is also on the right-hand side of the rule in (36b), we can apply this rule multiple times and therefore account for noun phrases with multiple adjectives such as (32). Figure 2.3 shows the structure of a noun phrase without an adjective and for a noun phrase with one or two adjectives. The adjective *klug* ‘smart’

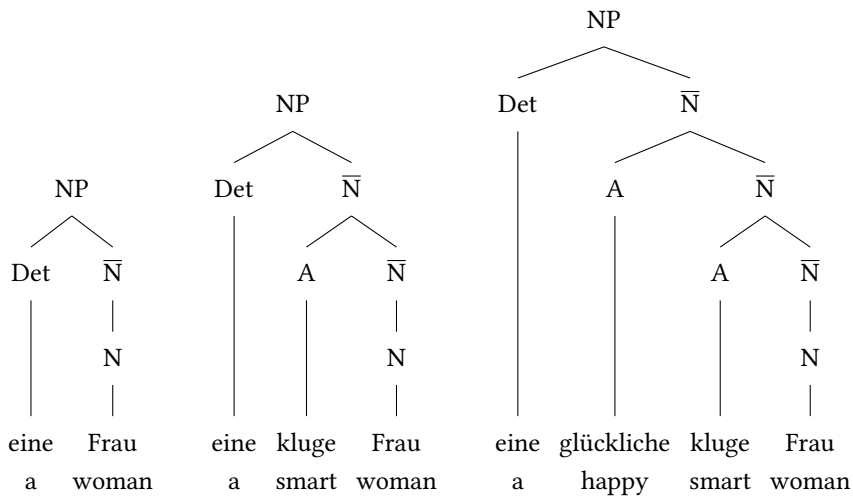


Figure 2.3: Noun phrases with differing numbers of adjectives

restricts the set of referents for the noun phrase. If we assume an additional adjective such as *glücklich* ‘happy’, then it only refers to those women who are happy as well as smart. These kinds of noun phrases can be used in contexts such as the following:

- (37) A: Alle klugen Frauen sind unglücklich.
 all smart women are unhappy
 B: Nein, ich kenne eine glückliche kluge Frau.
 no I know a happy smart woman

We observe that this discourse can be continued with *Aber alle glücklichen klugen Frauen sind schön* ‘but all happy, smart women are beautiful’ and a corresponding answer. The possibility to add adjectives to noun phrases such as *eine glückliche kluge Frau* ‘a happy, smart woman’ is accounted for in our rule system in (36). In the rule (36b), \bar{N} occurs on the left as well as the right-hand side of the rule. This kind of rule is referred to as *recursive*.

We have now developed a nifty little grammar that can be used to analyze noun phrases modified by adjectives. As a result, the combination of an adjective and noun is given constituent status. One may wonder at this point if it would not also make sense to also assume that determiners and adjectives form a constituent as we also have the following kind of noun phrases:

- (38) diese schlauen und diese neugierigen Frauen
these smart and these curious women

Here, we are dealing with a different structure, however. Two full NPs have been conjoined and part of the first conjunct has been deleted.

- (39) diese schlauen ~~Frauen~~ und diese neugierigen Frauen
these smart women and these curious women

One can find similar phenomena at the sentence and even word level:

- (40) a. dass Peter dem Mann das Buch gibt und Maria der Frau die Schallplatte
that Peter the man the book gives and Maria the woman the record
gibt
gives
'that Peters gives the book to the man and Maria the record to the woman'
b. be- und ent-laden
PRT and PRT-load
'load and unload'

Thus far, we have discussed how we can ideally integrate adjectives into our rules for the structure of noun phrases. Other adjuncts such as prepositional phrases or relative clauses can be combined with NPs in an analogous way to adjectives with \bar{N} :

- (41) a. $\bar{N} \rightarrow \bar{N}$ PP
b. $\bar{N} \rightarrow \bar{N}$ relative clause

With these rules and those in (36), it is possible – assuming the corresponding rules for PPs and relative clauses – to analyze all the examples in (30).

(36c) states that it possible for \bar{N} to consist of a single noun. A further important rule has not yet been discussed: we need another rule to combine nouns such as *Vater* 'father', *Sohn* 'son' or *Bild* 'picture', so-called *relational nouns*, with their arguments. Examples of these can be found in (42a–b). (42c) is an example of a nominalization of a verb with its argument:

- (42) a. der Vater von Peter
the father of Peter
'Peter's father'
b. das Bild vom Gleimtunnel
the picture of the Gleimtunnel
'The picture of the Gleim tunnel.'

- c. das Kommen des Installateurs
the coming of.the plumber
'The plumber's visit.'

The rule that we need to analyze (42a,b) is given in (43):

$$(43) \quad \bar{N} \rightarrow N \text{ PP}$$

Figure 2.4 shows two structures with PP-arguments. The tree on the right also contains an additional PP-adjunct, which is licensed by the rule in (41a).

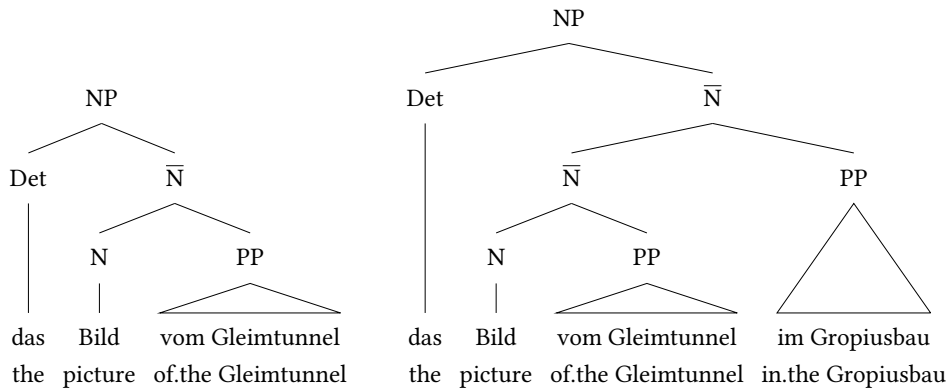


Figure 2.4: Combination of a noun with PP complement *vom Gleimtunnel* to the right with an adjunct PP

As well as the previously discussed NP structures, there are other structures where the determiner or noun is missing. Nouns can be omitted via ellipsis. (44) gives an example of noun phrases, where a noun that does not require a complement has been omitted. The examples in (45) show NPs in which only one determiner and complement of the noun has been realized, but not the noun itself:

- (44) a. eine kluge _
a smart
'a smart one'
b. eine kluge große _
a smart tall
'a smart tall one'
c. eine kluge _ aus Hamburg
a smart from Hamburg
'a smart one from Hamburg'

- d. eine kluge _, die alle kennen
 a smart who everyone knows
 'a smart one who everyone knows'
- (45) a. (Nein, nicht der Vater von Klaus), der _ von Peter war gemeint.
 no not the father of Klaus the of Peter was meant
 'No, it wasn't the father of Klaus, but rather the one of Peter that was meant.'
- b. (Nein, nicht das Bild von der Stadtautobahn), das _ vom Gleimtunnel
 no not the picture of the motorway the of the Gleimtunnel
 war beeindruckend.
 was impressive
 'No, it wasn't the picture of the motorway, but rather the one of the
 Gleimtunnel that was impressive.'
- c. (Nein, nicht das Kommen des Tischlers), das _ des Installateurs ist
 no not the coming of the carpenter the of the plumber is
 wichtig.
 important
 'No, it isn't the visit of the carpenter, but rather the visit of the plumber that
 is important.'

The underscore marks the position where the noun would normally occur. In English, the pronoun *one* must often be used in the corresponding position, but in German the noun is simply omitted. (See Fillmore, Lee-Goldmann & Rhomieux (2012: Section 4.12) for English examples without the pronoun *one*.) In phrase structure grammars, this can be described by so-called *epsilon production*. These rules replace a symbol with nothing (46a). The rule in (46b) is an equivalent variant which is responsible for the term *epsilon production*:

- (46) a. $N \rightarrow$
 b. $N \rightarrow \epsilon$

The corresponding trees are shown in Figure 2.5 on the facing page. Going back to boxes, the rules in (46) correspond to empty boxes with the same labels as the boxes of ordinary nouns. As we have considered previously, the actual content of the boxes is unimportant when considering the question of where we can incorporate them. In this way, the noun phrases in (30) can occur in the same sentences. The empty noun box also behaves like one with a genuine noun. If we do not open the empty box, we will not be able to ascertain the difference to a filled box.

It is not only possible to omit the noun from noun phrases, but the determiner can also remain unrealized in certain contexts. (47) shows noun phrases in plural:

- (47) a. Frauen
 women

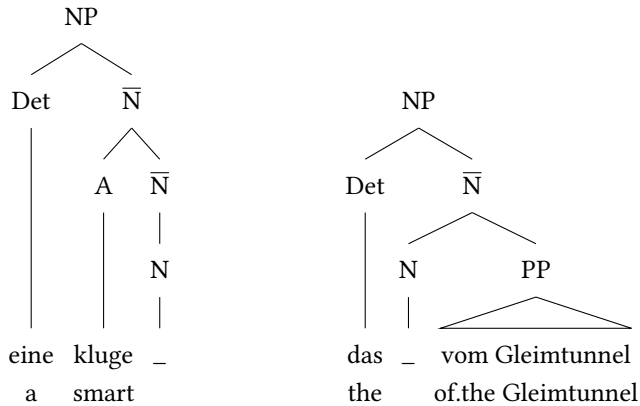


Figure 2.5: Noun phrases without an overt head

- b. Frauen, die wir kennen
women who we know
- c. kluge Frauen
smart women
- d. kluge Frauen, die wir kennen
smart women who we know

The determiner can also be omitted in singular if the noun denotes a mass noun:

- (48) a. Getreide
grain
- b. Getreide, das gerade gemahlen wurde
grain that just ground was
'Grain that has just been ground'
- c. frisches Getreide
fresh grain
- d. frisches Getreide, das gerade gemahlen wurde
fresh grain that just ground was
'Fresh grain that has just been ground'

Finally, both the determiner and the noun can be omitted:

- (49) a. Ich helfe klugen.
I help smart
'I help smart people.'

- b. Dort drüben steht frisches, das gerade gemahlen wurde.
 there overt stands fresh that just ground was
 ‘Over there is some fresh (grain) that has just been ground.’

Figure 2.6 shows the corresponding trees.

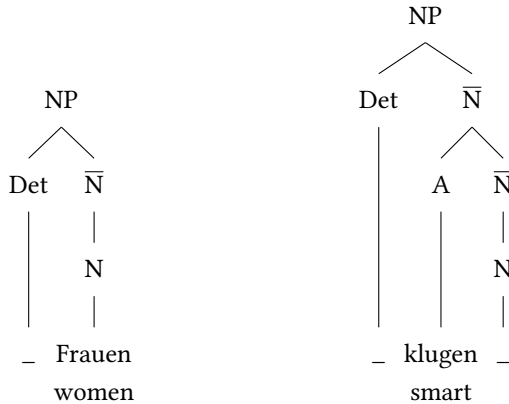


Figure 2.6: Noun phrases without overt determiner

It is necessary to add two further comments to the rules we have developed up to this point: Up to now, I have always spoken of adjectives. However, it is possible to have very complex adjective phrases in pre-nominal position. These can be adjectives with complements (50a,b) or adjectival participles (50c,d):

- (50) a. der seiner Frau treue Mann
 the his.DAT woman faithful man
 ‘the man faithful to his wife’
 b. der auf seinen Sohn stolze Mann
 the on his.ACC son proud man
 ‘the man proud of his son’
 c. der seine Frau liebende Mann
 the his.ACC woman loving man
 ‘the man who loves his wife’
 d. der von seiner Frau geliebte Mann
 the by his.DAT wife loved man
 ‘the man loved by his wife’

Taking this into account, the rule (36b) has to be modified in the following way:

- (51) $\bar{N} \rightarrow \text{AP } \bar{N}$

An adjective phrase (AP) can consist of an NP and an adjective, a PP and an adjective or just an adjective:

- (52) a. $AP \rightarrow NP\ A$
 b. $AP \rightarrow PP\ A$
 c. $AP \rightarrow A$

There are two imperfections resulting from the rules we have developed thus far. These are the rules for adjectives or nouns without complements in (52c) as well as (36c) – repeated here as (53):

- (53) $\bar{N} \rightarrow N$

If we apply these rules, then we will generate unary branching subtrees, that is trees with a mother that only has one daughter. See Figure 2.6 for an example of this. If we maintain the parallel to the boxes, this would mean that there is a box which contains another box which is the one with the relevant content.

There is in principle nothing stopping us from placing this information directly into the larger box. Instead of the rules in (54), we will simply use the rules in (55):

- (54) a. $A \rightarrow kluge$
 b. $N \rightarrow Mann$
- (55) a. $AP \rightarrow kluge$
 b. $\bar{N} \rightarrow Mann$

(55a) states that *kluge* ‘smart’ has the same properties as a full adjective phrase, in particular that it cannot be combined with a complement. This is parallel to the categorization of the pronoun *er* ‘he’ as an NP in the grammars (2) and (6).

Assigning \bar{N} to nouns which do not require a complement has the advantage that we do not have to explain why the analysis in (56b) is possible as well as (56a) despite there not being any difference in meaning.

- (56) a. $[_{NP}\ einige\ [_{\bar{N}}\ kluge\ [_{\bar{N}}\ [_{N}\ Frauen\]\ und\ [_{\bar{N}}\ [_{N}\ M\ddot{a}nner\]]]]]]$
 some smart women and men
- b. $[_{NP}\ einige\ [_{\bar{N}}\ kluge\ [_{\bar{N}}\ [_{N}\ [_{N}\ Frauen\]\ und\ [_{N}\ M\ddot{a}nner\]]]]]]$
 some smart women and men

In (56a), two nouns have projected to \bar{N} and have then been joined by coordination. The result of coordination of two constituents of the same category is always a new constituent with that category. In the case of (56a), this is also \bar{N} . This constituent is then combined with the adjective and the determiner. In (56b), the nouns themselves have been coordinated. The result of this is always another constituent which has the same category as its parts. In this case, this would be N. This N becomes \bar{N} and is then combined with the adjective. If nouns which do not require complements were categorized as \bar{N} rather than N, we would not have the problem of spurious ambiguities. The structure in (57) shows the only possible analysis.

- (57) $[_{NP} \text{ einige } [_{\bar{N}} \text{ kluge } [_{\bar{N}} [_{\bar{N}} \text{ Frauen }] \text{ und } [_{\bar{N}} \text{ Männer }]]]]$
 some smart women and men

2.4.2 Prepositional phrases

Compared to the syntax of noun phrases, the syntax of prepositional phrases (PPs) is relatively straightforward. PPs normally consist of a preposition and a noun phrase whose case is determined by that preposition. We can capture this with the following rule:

- (58) $PP \rightarrow P \text{ NP}$

This rule must, of course, also contain information about the case of the NP. I have omitted this for ease of exposition as was the case with the NP-rules and AP-rules above.

The Duden grammar (2005: §1300) offers examples such as those in (59), which show that certain prepositional phrases serve to further define the semantic contribution of the preposition by indicating some measurement, for example:

- (59) a. $[[\text{Einen Schritt}] \text{ vor } \text{dem Abgrund}] \text{ blieb er stehen.}$
 one step before the abyss stayed he standing
 ‘He stopped one step in front of the abyss.’
 b. $[[\text{Kurz}] \text{ nach dem Start}] \text{ fiel die Klimaanlage aus.}$
 shortly after the take.off fell the air.conditioning out
 ‘Shortly after take off, the air conditioning stopped working.’
 c. $[[\text{Schräg}] \text{ hinter der Scheune}] \text{ ist ein Weiher.}$
 diagonally behind the barn is a pond
 ‘There is a pond diagonally across from the barn.’
 d. $[[\text{Mitten}] \text{ im } \text{Urwald}] \text{ stießen die Forscher auf einen alten Tempel.}$
 middle in.the jungle stumbled the researchers on an old temple
 ‘In the middle of the jungle, the researches came across an old temple.’

To analyze the sentences in (59a,b), one could propose the following rules in (60):

- (60) a. $PP \rightarrow NP \text{ PP}$
 b. $PP \rightarrow AP \text{ PP}$

These rules combine a PP with an indication of measurement. The resulting constituent is another PP. It is possible to use these rules to analyze prepositional phrases in (59a,b), but it unfortunately also allows us to generate those in (61):

- (61) a. $*[_{PP} \text{ einen Schritt } [_{PP} \text{ kurz } [_{PP} \text{ vor } \text{dem Abgrund}]]]$
 one step shortly before the abyss
 b. $*[_{PP} \text{ kurz } [_{PP} \text{ einen Schritt } [_{PP} \text{ vor } \text{dem Abgrund}]]]$
 shortly one step before the abyss

The rules in (60) were used to analyze the examples in (61). Since the symbol PP occurs on both the left and right-hand of the rules, we can apply the rules in any order and as many times we like.

We can avoid this undesired side-effect by reformulating the previously assumed rules:

- (62) a. $PP \rightarrow NP \bar{P}$
 b. $PP \rightarrow AP \bar{P}$
 c. $PP \rightarrow \bar{P}$
 d. $\bar{P} \rightarrow P NP$

Rule (58) becomes (62d). The rule in (62c) states that a PP can consist of \bar{P} . Figure 2.7 shows the analysis of (63) using (62c) and (62d) as well as the analysis of an example with an adjective in the first position following the rules in (62b) and (62d):

- (63) vor dem Abgrund
 before the abyss
 ‘in front of the abyss’

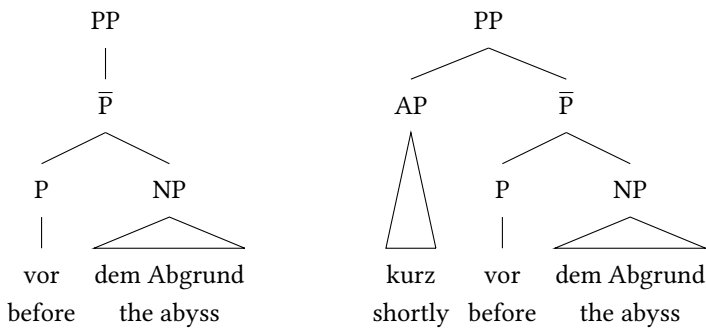


Figure 2.7: Prepositional phrases with and without measurement

At this point, the attentive reader is probably wondering why there is not an empty measurement in the left figure of Figure 2.7 analogous to the empty determiner in Figure 2.6. The reason for the empty determiner in Figure 2.6 is that the entire noun phrase without the determiner has a meaning similar to those with a determiner. The meaning normally contributed by the visible determiner has to somehow be incorporated in the structure of the noun phrase. If we did not place this meaning in the empty determiner, this would lead to more complicated assumptions about semantic combination: We only really require the mechanisms presented in Section 2.3 and these are very general in nature. The meaning is contributed by the words themselves and not by any rules. If we were to assume a unary branching rule such as that in the left tree in Figure 2.7 instead of the empty determiner, then this unary branching rule would have to provide

the semantics of the determiner. This kind of analysis has also been proposed by some researchers. See Chapter 21 for more on empty elements.

Unlike determiner-less NPs, prepositional phrases without an indication of degree or measurement do not lack any meaning component for composition. It is therefore not necessary to assume an empty indication of measurement, which somehow contributes to the meaning of the entire PP. The rule in (62c) states that a preposition consists of \bar{P} , that is, a combination of P and NP.

2.5 \bar{X} Theory

If we look again at the rules that we have formulated in the previous section, we see that heads are always combined with their complements to form a new constituent (64a,b), which can then be combined with further constituents (64c,d):

- (64) a. $\bar{N} \rightarrow N \text{ PP}$
 b. $\bar{P} \rightarrow P \text{ NP}$
 c. $\text{NP} \rightarrow \text{Det } \bar{N}$
 d. $\text{PP} \rightarrow \text{NP } \bar{P}$

Grammarians working on English noticed that parallel structures can be used for phrases which have adjectives or verbs as their head. I will discuss adjective phrases at this point and will discuss verb phrases later in Chapter 3. As in German, certain adjectives in English can take complements with the important restriction that adjective phrases with complements cannot realize these pre-nominally in English. (65) gives some examples of adjective phrases:

- (65) a. He is proud.
 b. He is very proud.
 c. He is proud of his son.
 d. He is very proud of his son.

Unlike prepositional phrases, complements of adjectives are normally optional. *proud* can be used with or without a PP. The degree expression *very* is also optional.

The rules which we need for this analysis are given in (66), with the corresponding structures in Figure 2.8 on the next page.

- (66) a. $\text{AP} \rightarrow \bar{A}$
 b. $\text{AP} \rightarrow \text{AdvP } \bar{A}$
 c. $\bar{A} \rightarrow A \text{ PP}$
 d. $\bar{A} \rightarrow A$

As was shown in Section 2.2, it is possible to generalize over very specific phrase structure rules and thereby arrive at more general rules. In this way, properties such as

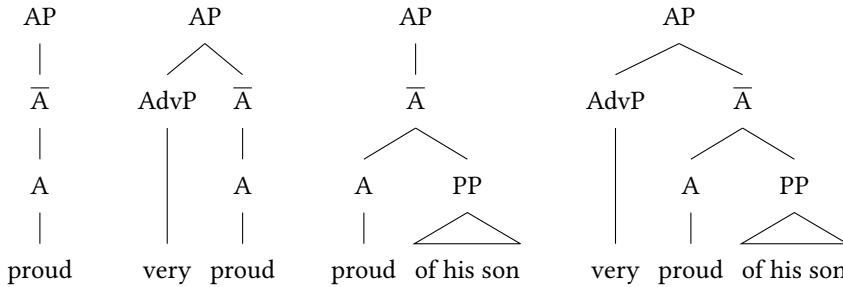


Figure 2.8: English adjective phrases

person, number and gender are no longer encoded in the category symbols, but rather only simple symbols such as NP, Det and N are used. It is only necessary to specify something about the values of a feature if it is relevant in the context of a given rule. We can take this abstraction a step further: Instead of using explicit category symbols such as N, V, P and A (or NP, VP, PP and PP), one can simply use a variable for the word class in question and speak of X and XP.

This form of abstraction can be found in so-called \bar{X} Theory (or X-bar theory, The term *bar* refers the line above the symbol.), which was developed by Chomsky (1970) and refined by Jackendoff (1977). This form of abstract rules plays an important role in many different theories. For example: Government & Binding (Chapter 3), Generalized Phrase Structure Grammar (Chapter 5) and Lexical Functional Grammar (Chapter 7). In HPSG (Chapter 9), \bar{X} Theory also plays a role, however not all restrictions of the \bar{X} schema have been adopted.

(67) shows a possible instantiation of \bar{X} rules, where the category X has been used in place of N, as well as examples of word strings which can be derived by these rules:

(67) \bar{X} rule	with specific categories	example
$\bar{X} \rightarrow \text{specifier } \bar{X}$	$\bar{N} \rightarrow \bar{\text{DET}} \bar{N}$	the [picture of Paris]
$\bar{X} \rightarrow \bar{X} \text{ adjunct}$	$\bar{N} \rightarrow \bar{N} \text{ REL_CLAUSE}$	[picture of Paris] [that everybody knows]
$\bar{X} \rightarrow \text{adjunct } \bar{X}$	$\bar{N} \rightarrow \bar{A} \bar{N}$	beautiful [picture of Paris]
$\bar{X} \rightarrow X \text{ complement}^*$	$\bar{N} \rightarrow N \bar{P}$	picture [of Paris]

Any word class can replace X (e.g. V, A or P). The X without the bar stands for a lexical item in the above rules. If one wants to make the bar level explicit, then it is possible to write X^0 . Just as with the rule in (15), where we did not specify the case value of the determiner or the noun but rather simply required that the values on the right-hand side of the rule match, the rules in (67) require that the word class of an element on the right-hand side of the rule (X or \bar{X}) matches that of the element on the left-hand side of the rule \bar{X} or $\bar{\bar{X}}$.

A lexical element can be combined with all its complements. The ‘*’ in the last rule stands for an unlimited amount of repetitions of the symbol it follows. A special case is zero-fold occurrence of complements. There is no PP complement of *Bild* ‘picture’ present in *das Bild* ‘the picture’ and thus N becomes \bar{N} . The result of the combination of a lexical element with its complements is a new projection level of X: the projection level 1, which is marked by a bar. \bar{X} can then be combined with adjuncts. These can occur to the left or right of \bar{X} . The result of this combination is still \bar{X} , that is the projection level is not changed by combining it with an adjunct. Maximal projections are marked by two bars. One can also write XP for a projection of X with two bars. An XP consists of a specifier and \bar{X} . Depending on one’s theoretical assumptions, subjects of sentences (Haider 1995; 1997a; Berman 2003a: Section 3.2.2) and determiners in NPs (Chomsky 1970: 210) are specifiers. Furthermore, degree modifiers (Chomsky 1970: 210) in adjective phrases and measurement indicators in prepositional phrases are also counted as specifiers.

Non-head positions can only host maximal projections and therefore complements, adjuncts and specifiers always have two bars. Figure 2.9 gives an overview of the minimal and maximal structure of phrases.

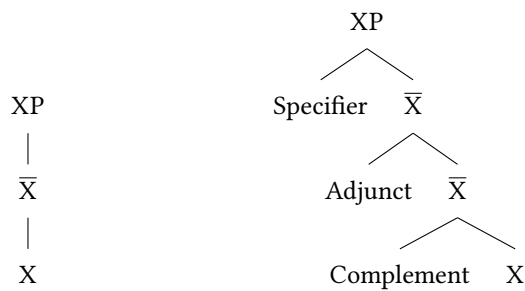


Figure 2.9: Minimal and maximal structure of phrases

Some categories do not have a specifier or have the option of having one. Adjuncts are optional and therefore not all structures have to contain an \bar{X} with an adjunct daughter. In addition to the branching shown in the right-hand figure, adjuncts to XP and head-adjuncts are sometimes possible. There is only a single rule in (67) for cases in which a head precedes the complements, however an order in which the complement precedes the head is of course also possible. This is shown in Figure 2.9.

Figure 2.10 on the next page shows the analysis of the NP structures *das Bild* ‘the picture’ and *das schöne Bild von Paris* ‘the beautiful picture of Paris’. The NP structures in Figure 2.10 and the tree for *proud* in Figure 2.8 show examples of minimally populated structures. The left tree in Figure 2.10 is also an example of a structure without an adjunct. The right-hand structure in Figure 2.10 is an example for the maximally populated structure: specifier, adjunct, and complement are present.

The analysis given in Figure 2.10 assumes that all non-heads in a rule are phrases. One therefore has to assume that there is a determiner phrase even if the determiner is

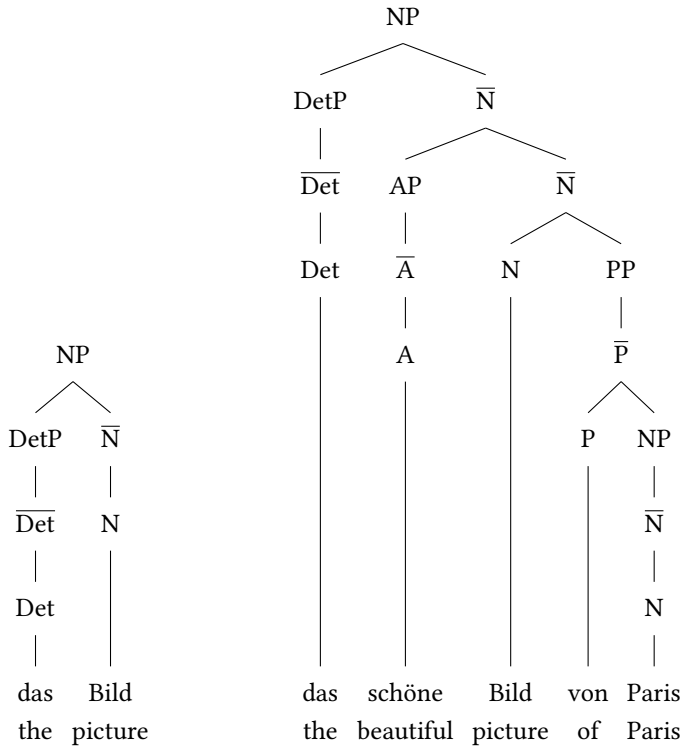


Figure 2.10: \bar{X} analysis of *das Bild* 'the picture' and *das schöne Bild von Paris* 'the beautiful picture of Paris'

not combined with other elements. The unary branching of determiners is not elegant but it is consistent.⁷ The unary branchings for the NP *Paris* in Figure 2.10 may also seem somewhat odd, but they actually become more plausible when one considers more complex noun phrases:

- (68) a. *das Paris der dreißiger Jahre*
 the Paris of.the thirty years
 '30's Paris'
- b. *die Maria aus Hamburg*
 the Maria from Hamburg
 'Maria from Hamburg'

⁷ For an alternative version of \bar{X} theory, which does not assume elaborate structure for determiners, see Muysken (1982a).

Unary projections are somewhat inelegant but this should not concern us too much here as we have already seen in the discussion of lexical entries in (55) that unary branching nodes can be avoided for the most part and that it is indeed desirable to avoid such structures as one otherwise gets spurious ambiguities. In the following chapters, we will discuss approaches such as Categorical Grammar and HPSG, which do not assume unary rules for determiners, adjectives and nouns.

Furthermore, other \bar{X} theoretical assumptions will not be shared by several theories discussed in this book. In particular, the assumption that non-heads always have to be maximal projections will be disregarded. Pullum (1985) and Kornai & Pullum (1990) have shown that the respective theories are not necessarily less restrictive than theories which adopt a strict version of the \bar{X} theory. The reader is referred to the discussion in Section 13.1.2.

Comprehension questions

1. Why are phrase structure grammars that use only atomic categories inadequate for the description of natural languages?
2. Assuming the grammar in (6), state which steps (replacing symbols) one has to take to get to the symbol V in the sentence (69).

(69) er das Buch dem Mann gibt
he the book the man gives
'He gives the book to the man.'

Your answer should resemble the table in (3).

3. Give a representation of the meaning of (70) using predicate logic:

(70) a. Ulrike kennt Hans.
Ulrike knows Hans
b. Joshi freut sich.
Joshi is.happy REFL
'Joshi is happy.'

Exercises

1. On page 57, I claimed that there is an infinite number of grammars we could use to analyze (1). Why is this claim correct?
2. Try to come up with some ways in which we can tell which of these possible grammars is or are the best?

3. A fragment for noun phrase syntax was presented in Section 2.4.1. Why is the interaction of the rules in (71) is problematic?

- (71) a. $NP \rightarrow Det \bar{N}$
 b. $\bar{N} \rightarrow N$
 c. $Det \rightarrow \epsilon$
 d. $N \rightarrow \epsilon$

4. Why is it not a good idea to mark *books* as NP in the lexicon?
5. Can you think of some reasons why it is not desirable to assume the following rule for nouns such as *books*:

- (72) $NP \rightarrow \text{Modifier}^* \text{books} \text{Modifier}^*$

The rule in (72) combines an unlimited number of modifiers with the noun *books* followed by an unlimited number of modifiers. We can use this rule to derive phrases such as those in (73):

- (73) a. *books*
 b. *interesting books*
 c. *interesting books from Stuttgart*

Make reference to coordination data in your answer. Assume that symmetric coordination requires that both coordinated phrases or words have the same syntactic category.

6. Fillmore et al. (2012) suggested treating nounless structures like those in (74) as involving a phrasal construction that combines the determiner *the* with an adjective.

- (74) a. *Examine the plight of the very poor.*
 b. *Their outfits range from the flamboyant to the functional.*
 c. *The unimaginable happened.*

(75) shows a phrase structure rule that corresponds to their construction:

- (75) $NP \rightarrow \text{the Adj}$

Adj stands for something that can be a single word like *poor* or complex like *very poor*.

Revisit the German data in (44) and (45) and explain why such an analysis and even a more general one as in (76) would not extend to German.

(76) $NP \rightarrow \text{Det Adj}$

7. Why can \bar{X} theory not account for German adjective phrases without additional assumptions?
8. Come up with a phrase structure grammar that can be used to analyze the sentence in (77), but also rules out the sentences in (78).

(77) a. Der Mann hilft der Frau.
the.NOM man helps the.DAT woman
b. Er gibt ihr das Buch.
he.NOM gives her.DAT the book
c. Er wartet auf ein Wunder.
he.NOM waits on a miracle
'He is waiting for a miracle.'

(78) a. *Der Mann hilft er.
the.NOM man helps he.NOM
b. *Er gibt ihr den Buch.
he.NOM gives her.DAT the.M book.N

9. Consider which additional rules would have to be added to the grammar you developed in the previous exercise in order to be able to analyze the sentences in (79):

(79) a. Der Mann hilft der Frau jetzt.
the.NOM man helps the.DAT woman now
b. Der Mann hilft der Frau neben dem Bushäuschen.
the.NOM man helps the.DAT woman next to.the bus.shelter
c. Er gibt ihr das Buch jetzt.
he.NOM gives her.DAT the.ACC book now
d. Er gibt ihr das Buch neben dem Bushäuschen.
he.NOM gives her.DAT the.ACC book next to.the bus.shelter
e. Er wartet jetzt auf ein Wunder.
he.NOM waits now on a miracle
'He is waiting for miracle now.'
f. Er wartet neben dem Bushäuschen auf ein Wunder.
he.NOM waits next to.the.DAT bus.shelter on a miracle
'He is waiting for a miracle next to the bus shelter.'

10. Install a prolog-system (e. g. SWI-Prolog⁸) and try out your grammar. Details for the notation can be found in the corresponding handbook under the key word Definite Clause Grammar (DCG).

Further reading

The expansion of phrase structure grammars to include features was proposed as early as Harman (1963).

The phrase structure grammar for noun phrases discussed in this chapter covers a large part of the syntax of noun phrases but cannot explain certain NP structures. Furthermore, it has the problem, which exercise 3 is designed to show. A discussion of these phenomena and a solution in the framework of HPSG can be found in Netter (1998) and Kiss (2005).

The discussion of the integration of semantic information into phrase structure grammars was very short. A detailed discussion of predicate logic and its integration into phrase structure grammars – as well as a discussion of quantifier scope – can be found in Blackburn & Bos (2005).

⁸ <http://www.swi-prolog.org/>

3 Transformational grammar – Government & Binding

Transformational Grammar and its subsequent incarnations (such as Government and Binding Theory and Minimalism) were developed by Noam Chomsky at MIT in Boston (Chomsky 1957; 1965; 1975; 1981a; 1986a; 1995). Manfred Bierwisch (1963) was the first to implement Chomsky's ideas for German. In the 60s, the decisive impulse came from the *Arbeitsstelle Strukturelle Grammatik* ('Workgroup for Structural Grammar', which was part of the Academy of Science in the GDR. See Bierwisch (1992) and Vater (2010) for an historic overview.) As well as Bierwisch's work, the following other works focussing on German, which have been written as part of this research program should also be mentioned : Fanselow (1987), Fanselow & Felix (1987), von Stechow & Sternefeld (1988), Grewendorf (1988), Haider (1993), Sternefeld (2006).

The variants of Chomskyan theories are often grouped under the heading *Generative Grammar*. This term comes from the fact that phrase structure grammars, coupled with the additional assumptions of Chomsky, can generate sets of well-formed expressions (see p. 56). It is this set of sentences that constitutes a language (in the formal sense) and one can test if a sentence forms part of a language by checking if a particular sentence is in the set of sentences generated by a given grammar or not. In this sense, simple phrase structure grammars and, with corresponding formal assumptions, GPSG, LFG, HPSG and Construction Grammar (CxG) are generative theories. In recent years, a different view of the formal basis of theories such as LFG, HPSG and CxG has emerged such that the aforementioned theories are now *model theoretic* theories rather than generative-eumerative ones¹ (See Chapter 14 for discussion). In 1957, Chomsky defined the term *Generative Grammar* in the following way (also see Chomsky 1995: 162):

A grammar of a language purports to be a description of the ideal speaker-hearer's intrinsic competence. If the grammar is, furthermore, perfectly explicit – in other words, if it does not rely on the intelligence of the understanding reader but rather provides an explicit analysis of his contribution – we may call it (somewhat redundantly) a *generative grammar*. (Chomsky 1965: 4)

In this sense, all grammatical theories discussed in this book would be viewed as generative grammars. To differentiate further, sometimes the term *Mainstream Generative Grammar* (MGG) is used (Culicover & Jackendoff 2005: 3) for Chomskyan models. In Sections 3.1–3.5, I will discuss a well-developed and very influential version of Chomskyan

¹ Model theoretic approaches are always constraint-based and the terms *model theoretic* and *constraint-based* are sometimes used synonymously.

grammar, GB theory. More recent developments known under Section 4 then deals with the proposals made as part of the Minimalist Program.

3.1 General remarks on the representational format

3.1.1 Transformations

In the previous chapter, I introduced simple phrase structure grammars. Chomsky (1957: Chapter 5) criticized these kind of rewrite grammars since – in his opinion – it is not clear how one can capture the relationship between active and passive sentences or the various ordering possibilities of constituents in a sentence. While it is of course possible to formulate different rules for active and passive sentences in a phrase structure grammar, it would not adequately capture the fact that the same phenomenon occurs in the example pairs in (1)–(3):

- (1) a. weil dort noch jemand arbeitet
because there still somebody works
'because somebody is still working there'
b. weil dort noch gearbeitet wird
because there still worked will
'because work is still being done there'
- (2) a. weil er den Weltmeister schlägt
because he the world.champion beats
'because he beats the world champion'
b. weil der Weltmeister geschlagen wird
because the world.champion beaten was
'because the world champion was beaten'
- (3) a. weil der Mann der Frau den Schlüssel stiehlt
because the man the woman the key steals
'because the man is stealing the key from the woman'
b. weil der Frau der Schlüssel gestohlen wurde
because the woman the key stolen was
'because the key was stolen from the woman'

Chomsky (1957: 43) suggests a transformation that creates a connection between active and passive sentences. In English, there is transformation similar to that in (4):

- (4) NP V NP \rightarrow 3 [AUX be] 2en [PP [P by] 1]
1 2 3

This transformational rule maps a tree with the symbols on the left-hand side of the rule onto a tree with the symbols on the right-hand side of the rule. Accordingly, 1, 2 and 3 on the right of the rule correspond to symbols, which are under the numbers on the left-hand side. *en* stands for the morpheme which forms the participle (*seen*, *been*, ..., but also *loved*). Both trees for (5a,b) are shown in Figure 3.1.

- (5) a. John loves Mary.
b. Mary is loved by John.

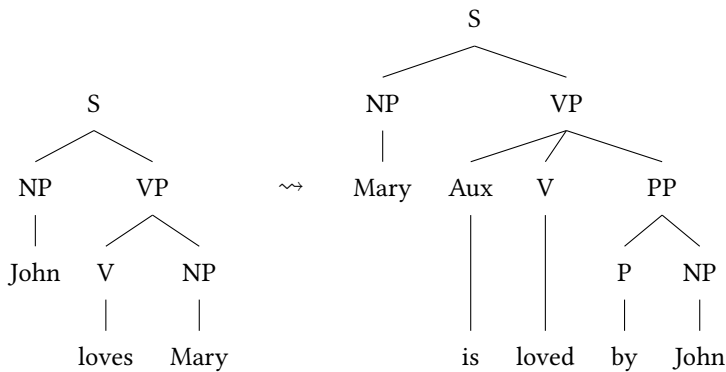


Figure 3.1: Application of passive transformation

The symbols on the left of transformational rules do not necessarily have to be in a local tree, that is, they can be daughters of different mothers as in Figure 3.1.

Rewrite grammars can be divided into four complexity classes based on the properties they have. The simplest grammars are assigned to the class 3, whereas the most complex are of type 0. The so-called context-free grammars we have dealt with thus far are of type 2. Transformational grammars which allow symbols to be replaced by arbitrary other symbols are of type 0 (Peters & Ritchie 1973). Research on the complexity of natural languages shows that the highest complexity level (type 0) is too complex for natural language. It follows from this – assuming that one wants to have a restrictive formal apparatus for the description of grammatical knowledge (Chomsky 1965: 62) – that the form and potential power of transformations has to be restricted.² Another criticism of early versions of transformational grammar was that, due to a lack of restrictions, the way in which transformations interact was not clear. Furthermore, there were problems associated with transformations which delete material (see Klenk 2003: Section 3.1.4). For this reason, new theoretical approaches such as Government & Binding (Chomsky 1981a) were developed. In this model, the form that grammatical rules can take is restricted (see Section 3.1.7). Elements moved by transformations are still represented in

² For more on the power of formal languages, see Chapter 18.

their original position, where they can be located and semantically interpreted. There are also more general principles, which serve to restrict transformations.

After some initial remarks on the model assumed for language acquisition in GB theory, we will take a closer look at phrase structure rules, transformations and constraints.

3.1.2 The hypothesis of language acquisition: Principles & Parameters

Chomsky (1965: Section I.8) assumes that language knowledge must be innate since the language system is, in his opinion, so complex that it would be impossible to learn a language from the given input using more general cognitive principles alone (also, see Section 13.8). If it is not possible to learn language solely through interaction with our environment, then at least part of our language ability must be innate. The question of exactly what is innate and if humans actually have an innate capacity for language remains controversial and the various positions on the question have changed over the course of the last decades. Some notable works on this topic are Pinker (1994), Tomasello (1995), Wunderlich (2004), Hauser, Chomsky & Fitch (2002) and Chomsky (2007). For more on this discussion, see Chapter 13.

Chomsky (1981a) also assumes that there are general, innate principles which linguistic structure cannot violate. These principles are parametrized, that is, there is more than one option for a given parameter. A parameter can be set in a different way for different languages. An example for a parametrized principle is shown in (6):

- (6) Principle: A head occurs before or after its complement(s) depending on the value of the parameter POSITION.

The Principles and Parameters model assumes that a significant part of language acquisition consists of extracting enough information from the linguistic input in order to be able to set parameters. Chomsky (2000: 8) compares the setting of parameters to flipping a switch. For a detailed discussion of the various assumptions about language acquisition in the P&P-model, see Section 16. Speakers of English have to learn that heads occur before their complements in their language, whereas a speaker of Japanese has to learn that heads follow their complements. (7) gives the respective examples:

- (7) a. be showing pictures of himself
b. zibun -no syasin-o mise-te iru
REFL from picture showing be

As one can see, the Japanese verb, noun and prepositional phrases are a mirror image of the corresponding phrases in English. (8) provides a summary and shows the parametric value for the position parameter:

- | | | | |
|-----|----------|--------------------------------|-------------------------|
| (8) | Language | Observation | Parameter: head initial |
| | English | Heads occur before complements | + |
| | Japanese | Heads occur after complements | – |

Investigating languages based on their differences with regard to certain assumed parameters has proven to be a very fruitful line of research in the last few decades and has resulted in an abundance of comparative cross-linguistic studies.

After these introductory comments on language acquisition, the following sections will discuss the basic assumptions of GB theory.

3.1.3 The T-Model

Chomsky criticized the fact that simple PSGs cannot adequately capture certain correlations. An example of this is the relationship between active and passive sentences. In phrase structure grammars, one would have to formulate active and passive rules for intransitive, transitive and ditransitive verbs. The fact that the passive can otherwise be consistently described as the suppression of the most prominent argument is not captured by phrase structure rules. Chomsky therefore assumes that there is an underlying structure, the so-called *Deep Structure*, and that other structures are derived from this. During the derivation of new structures, parts of this Deep Structure can be deleted or moved. In this way, one can explain the relationship between active and passive sentences. As the result of this kind of manipulation of structures, also called transformations, one arrives at a new structure, the *Surface Structure*, from the original Deep Structure. Since the Surface Structure does not actually mirror the actual use of words in a sentence in some versions of the theory, the term *S-Structure* is sometimes used instead as to avoid misunderstandings.

- (9) *Surface Structure* = S-Structure
Deep Structure = D-Structure

Figure 3.2 on the following page gives an overview of the GB architecture: phrase structure rules and the lexicon license the D-Structure from which the S-Structure is derived by means of transformations.

S-Structure consists of a Phonetic Form (PF) and Logical Form (LF). The model is referred to as the *T-model* (or Y-model) because D-Structure, S-Structure, PF and LF form an upside-down T (or Y). We will look at each of these individual components in more detail.

Using phrase structure rules, one can describe the relationships between individual elements. The format for these rules is \bar{X} syntax (see Section 2.5). The lexicon, together with the structure licensed by \bar{X} syntax, forms the basis for D-Structure. D-Structure is then a syntactic representation of the selectional grid (= valency classes) of individual word forms in the lexicon.

The lexicon contains a lexical entry for every word, which comprises information about morphophonological structure, syntactic features and selectional properties. This will be explained in more detail in Section 3.1.6. Depending on one's exact theoretical assumptions, morphology is viewed as part of the lexicon. Inflectional morphology is, however, mostly consigned to the realm of syntax. The lexicon is an interface for semantic interpretation of individual word forms.

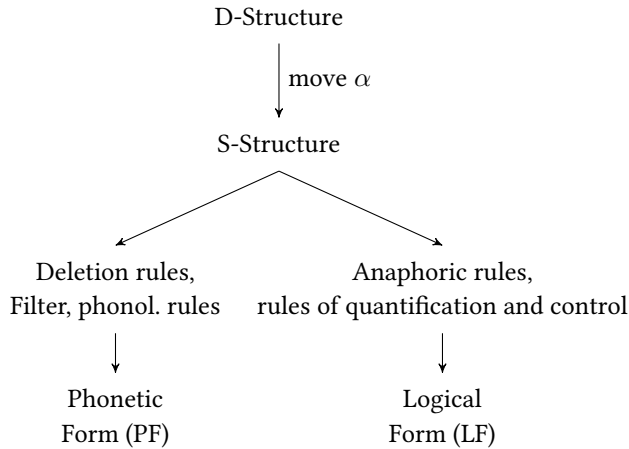


Figure 3.2: The T-Model

The surface position in which constituents are realized is not necessarily the position specified by valency information. There are therefore the following ordering variants for a sentence with a ditransitive verb in (10):

- (10) a. [dass] der Mann der Frau das Buch gibt
 that the man.NOM the woman.DAT the book.ACC gives
 ‘that the man gives the woman the book’
- b. Gibt der Mann der Frau das Buch?
 gives the man.NOM the woman.DAT the book.ACC
 ‘Does the man give the woman the book?’
- c. Der Mann gibt der Frau das Buch.
 the man.NOM gives the woman.DAT the book.ACC
 ‘The man gives the woman the book.’

The following transformational rules for the movements above are assumed: (10b) is derived from (10a) by fronting the verb, and (10c) is derived from (10b) by fronting the nominative noun phrase. In GB theory, there is only one very general transformation: Move- α = “Move anything anywhere!”. The nature of what exactly can be moved where and for which reason is determined by principles. Examples of such principles are the Theta Criterion and the Case Filter, which will be dealt with below.

The relations between a predicate and its arguments that are determined by the lexical entries have to be accessible for semantic interpretation at all representational levels. For this reason, the base position of a moved element is marked with a trace. This means that the position in which the fronted *gibt* (‘gives’) originated in (11b) is indicated. This marking is referred to as a *trace* or a *gap*. In the discussion of noun phrase syntax in

Section 2.4.1, we saw a number of examples for the fact that positions in constituents can be empty (page 70).

- (11) a. [dass] der Mann der Frau das Buch gibt
 that the man the woman the book gives
 ‘that the man gives the woman the book’
 b. Gibt_i der Mann der Frau das Buch _{-i}?
 gives the man the woman the book
 ‘Does the man give the woman the book?’
 c. [Der Mann]_j gibt_i _{-j} der Frau das Buch _{-i}.
 the man gives the woman the book
 ‘The man gives the woman the book.’

(11c) is derived from (11a) by means of two movements, which is why there are two traces in (11c). The traces are marked with indices so it is possible to distinguish the moved constituents. The corresponding indices are then present on the moved constituents. Sometimes, *e* (for *empty*) or *t* (for *trace*) is used to represent traces.

The S-Structure derived from the D-Structure is a surface-like structure but should not be equated with the structure of actual utterances.

3.1.4 Phonetic Form

Phonological operations are represented at the level of Phonetic Form (PF). PF is responsible for creating the form, which is actually pronounced. For example, so-called *wanna*-contraction takes place at PF (Chomsky 1981a: 20–21).

- (12) a. The students want to visit Paris.
 b. The students wannna visit Paris.

The contraction in (12) is licensed by the optional rule in (13):

- (13) want + to → wanna

3.1.5 Logical Form

Logical Form is the syntactic level which mediates between S-Structure and the semantic interpretation of a sentence. Some of the phenomena which are dealt with by LF are anaphoric reference of pronouns, quantification and control.

Syntactic factors play a role in resolving anaphoric dependencies. An important component of GB theory is Binding Theory, which seeks to explain what a pronoun can or must refer to and when a reflexive pronoun can or must be used. (14) gives some examples of both pronouns and reflexive pronouns:

- (14) a. Peter kauft einen Tisch. Er gefällt ihm.
 Peter buys a table.M he likes him
 ‘Peter is buying a table. He likes it.’
 b. Peter kauft eine Tasche. Er gefällt ihm.
 Peter buys a bag.F he likes him
 ‘Peter is buying a bag. He likes him.’
 c. Peter kauft eine Tasche. Er gefällt sich.
 Peter buys a bag.F he likes himself
 ‘Peter is buying a bag. He likes himself.’

In the first example, *er* (‘he’) can refer to either Peter, the table or something/someone else that was previously mentioned in the context. *ihm* (‘him’) can refer to Peter or someone in the context. Reference to the table is restricted by world knowledge. In the second example, *er* (‘he’) cannot refer to *Tasche* (‘bag’) since *Tasche* is feminine and *er* is masculine. *er* (‘he’) can refer to Peter only if *ihm* (‘him’) does not refer to Peter. *ihm* would otherwise have to refer to a person in the wider context. This is different in (14c). In (14c), *er* (‘he’) and *sich* (‘himself’) must refer to the same object. This is due to the fact that the reference of reflexives such as *sich* is restricted to a particular local domain. Binding Theory attempts to capture these restrictions.

LF is also important for quantifier scope. Sentences such as (15a) have two readings. These are given in (15b) and (15c).

- (15) a. Every man loves a woman.
 b. $\forall x \exists y (man(x) \rightarrow (woman(y) \wedge love(x, y)))$
 c. $\exists y \forall x (man(x) \rightarrow (woman(y) \wedge love(x, y)))$

The symbol \forall stands for a *universal quantifier* and \exists stands for an *existential quantifier*. The first formula corresponds to the reading that for every man, there is a woman who he loves and in fact, these can be different women. Under the second reading, there is exactly one woman such that all men love her. The question of when such an ambiguity arises and which reading is possible when depends on the syntactic properties of the given utterance. LF is the level which is important for the meaning of determiners such as *a* and *every*.

Control Theory is also specified with reference to LF. Control Theory deals with the question of how the semantic role of the infinitive subject in sentences such as (16) is filled.

- (16) a. Der Professor schlägt dem Studenten vor, die Klausur noch mal zu
 the professor suggests the student PRT the test once again to
 schreiben.
 write
 ‘The professor advises the student to take the test again.’

- b. Der Professor schlägt dem Studenten vor, die Klausur nicht zu bewerten.
 the professor suggests the student PRT the test not to grade
 ‘The professor suggests to the student not to grade the test.’
- c. Der Professor schlägt dem Studenten vor, gemeinsam ins Kino zu gehen.
 the professor suggests the student PRT together into cinema to go
 ‘The professor suggests to the student to go to the cinema together.’

3.1.6 The Lexicon

The meaning of words tells us that they have to be combined with certain roles (“acting person” or “affected thing”) when creating more complex phrases. For example, the fact that it needs two arguments belongs to the semantic contribution of *know*. The semantic representation of the contribution of the verb *know* in (17a) is given in (17b):

- (17) a. Maria knows the man.
 b. *know'*(x,y)

Dividing heads into valency classes is also referred to as *subcategorization*: *know* subcategorizes for a subject and an object. This term comes from the fact that a head is already categorized with regard to its part of speech (verb, noun, adjective, ...) and then further subclasses (e. g. intransitive or transitive verb) are formed from valency information. As well as the phrase *X subcategorizes for Y*, we can also say that *X selects Y*. *know* is referred to as the predicate since *know'* is the logical predicate. The subject and object are the arguments of the predicate. There are several terms used to describe the total selectional requirements such as *argument structure*, *valency frames*, *subcategorization frame*, *thematic grid* and *theta grid* or θ -grid.

Adjuncts modify semantic predicates and when the semantic aspect is emphasized they are also called *modifiers*. Adjuncts are not present in the argument structure of predicates.

Following GB assumptions, arguments occur in specific positions in the clause – in so-called argument positions (e. g. the sister of an X^0 element, see Section 2.5). The Theta Criterion states that elements in argument positions have to be assigned a semantic role – a so-called theta role – and each role can only be assigned once (Chomsky 1981a: 36):

Principle 1 (Theta Criterion)

- *Each theta role is assigned to exactly one argument position.*
- *Every phrase in an argument position receives exactly one theta role.*

The arguments of a head are ordered, that is, one can differentiate between higher- and lower-ranked arguments. The highest-ranked argument of verbs and adjectives has a special status. Since GB assumes that it is often (and always in some languages) realized

in a position outside of the verb or adjective phrase, it is often referred to as the *external argument*. The remaining arguments occur in positions inside of the verb or adjective phrase. These kind of arguments are dubbed *internal arguments* or *complements*. For simple sentences, this often means that the subject is the external argument.

When discussing types of arguments, one can identify three classes of theta roles:

- Class 1: agent (acting individual), the cause of an action or feeling (stimulus), holder of a certain property
- Class 2: experiencer (perceiving individual), the person profiting from something (beneficiary) (or the opposite: the person affected by some kind of damage), possessor (owner or soon-to-be owner of something, or the opposite: someone who has lost or is lacking something)
- Class 3: patient (affected person or thing), theme

If a verb has several theta roles of this kind to assign, Class 1 normally has the highest rank, whereas Class 3 has the lowest. Unfortunately, the assignment of semantic roles to actual arguments of verbs has received a rather inconsistent treatment in the literature. This problem has been discussed by Dowty (1991), who suggests using proto-roles. An argument is assigned the proto-agent role if it has sufficiently many of the properties that were identified by Dowty as prototypical properties of agents (e.g. animacy, volitionality).

The mental lexicon contains *lexical entries* with the specific properties of syntactic words needed to use that word grammatically. Some of these properties are the following:

- Form
- Meaning (Semantics)
- Grammatical features:
syntactic word class + morphosyntactic features
- Theta grid

(18) shows an example of a lexical entry:

(18)

form	<i>hilft</i> ‘helps’
semantics	<i>helfen</i> ’
grammatical features	verb, 3rd person singular indicative present active
theta grid	
theta roles	<u>agent</u> beneficiary
grammatical particularities	dative

Assigning semantic roles to specific syntactic requirements (beneficiary = dative) is also called *linking*.

Arguments are ordered according to their ranking: The highest argument is furthest left. In the case of *helfen*, the highest argument is the external argument, which is why the Agent is underlined. With so-called unaccusative verbs,³ the highest argument is not treated as the external argument. It would therefore not be underlined in the corresponding lexical entry.

3.1.7 \bar{X} Theory

In GB, it is assumed that all syntactic structures licensed by the core grammar⁴ correspond to the \bar{X} schema (see Section 2.5).⁵ In the following sections, I will comment on the syntactic categories assumed and the basic assumptions with regard to the interpretation of grammatical rules.

3.1.7.1 Syntactic categories

The categories which can be used for the variable X in the \bar{X} schema are divided into lexical and functional categories. This correlates somewhat to the difference between open and closed word classes. The following are lexical categories:

- V = verb
- N = noun
- A = adjective
- P = preposition/postposition
- Adv = adverb

Lexical categories can be represented using elementary features based on cross-classification:⁶

³ See Perlmutter (1978) for a discussion of unaccusative verbs. The term *ergative verb* is also common, albeit a misnomer. See Burzio (1981; 1986) for the earliest work on unaccusatives in the Chomskyan Framework and Grewendorf (1989) for German. Also, see Pullum (1988) on the usage of these terms and for a historical evaluation.

⁴ Chomsky (1981a: 7–8) distinguishes between a regular area of language which is determined by a grammar which can be acquired using genetically fixed language-specific knowledge and a periphery, to which irregular parts of language such as idioms (e. g. *to pull the wool over sb.'s eyes*) belong. See Section 16.3.

⁵ Chomsky (1970: 210) assumes that there can be grammatical rules which deviate from the \bar{X} schema. It is, however, common practice to assume that languages exclusively use \bar{X} structures.

⁶ See Chomsky (1970: 199) for a cross-classification of N, A and V, and Jackendoff (1977: Section 3.2) for a cross-classification of P, albeit with differing feature assignment.

Table 3.1: Representation of four lexical categories using two binary features

	–V	+V
–N	P = [–N, –V]	V = [–N, +V]
+N	N = [+N, –V]	A = [+N, +V]

Adverbs are viewed as intransitive prepositions and are therefore captured by the decomposition in the table above.

Using this cross-classification, it is possible to formulate generalizations. One can, for example, simply refer to adjectives and verbs: all lexical categories which are [+V] are either adjectives or verbs. Furthermore, one can say of [+N] categories (nouns and adjectives) that they can bear case.

Apart from this some authors have tried to associate the head position with the feature values in Table 3.1 (see z. B. Grewendorf (1988: 52); Haftka (1996: 124); G. Müller (2011: 238)). With prepositions and nouns, the head precedes the complement in German:

- (19) a. *für* Marie
for Marie
b. *Bild* von Maria
picture of Maria

With adjectives and verbs, the head is final:

- (20) a. dem König *treu*
the.DAT king loyal
'Loyal to the king'
b. der [dem Kind *helfende*] Mann
the the.DAT child helping man
'the man helping the child'
c. dem Mann *helfen*
the.DAT man help
'help the man'

With respect to the values in Table 3.1 one can conclude that the head is final with [+V] categories and initial with [–V] categories. Unfortunately, this generalization runs into the problem that there are also postpositions in German. These are, like prepositions, not verbal, but do not occur before the NP they require:

- (21) a. des Geldes *wegen*
the money.GEN because
'because of the money'

- b. die Nacht *über*
the night during
'during the night'

Therefore, one must either invent a new category, or abandon the attempt to use binary category features to describe ordering restrictions. If one were to place postpositions in a new category, it would be necessary to assume another binary feature.⁷ Since this feature can have either a negative or a positive value, one would then have four additional categories. There are then eight possible feature combinations, some of which would not correspond to any plausible category.

For functional categories, GB does not propose a cross-classification. Usually, the following categories are assumed:

- C Complementizer (subordinating conjunctions such as *dass* 'that')
- I Finiteness (as well as Tense and Mood);
also *linfl* in earlier work (inflection),
T in more recent work (Tense)
- D Determiner (article, demonstrative)

3.1.7.2 Assumptions and rules

In GB, it is assumed that all rules must follow the \bar{X} format discussed in Section 2.5. In other theories, rules which correspond to the \bar{X} format are used along other rules which do not. If the strict version of \bar{X} Theory is assumed, this comes with the assumption of *endocentricity*: Every phrase has a head and every head is part of a phrase (put more technically: every head projects to a phrase).

Furthermore, as with phrase structure grammars, it is assumed that the branches of tree structures cannot cross (*Non-Tangling Condition*). This assumption is made by the majority of theories discussed in this book. There are, however, some variants of TAG, HPSG and Construction Grammar, which allow crossing branches and therefore discontinuous constituents (Becker, Joshi & Rambow 1991; Reape 1994; Bergen & Chang 2005).

⁷ Martin Haspelmath has pointed out that one could assume a rule that moves a post-head argument into a pre-head position. This would be parallel to the realization of prepositional arguments of adjectives in German:

- (i) a. auf seinen Sohn stolz
on his son proud
'proud of his son'
- b. stolz auf seinen Sohn
proud of his son

But note that the situation is different with postpositions here, while all adjectives that take prepositional objects allow for both orders, this is not the case for prepositions. Most prepositions do not allow their object to occur before them. It is an idiosyncratic feature of some postpositions that they want to have their argument to the left.

In \bar{X} Theory, one normally assumes that there are at most two projection levels (X' and X''). However, there are some versions of Mainstream Generative Grammar and other theories which allow three or more levels (Jackendoff 1977; Uszkoreit 1987). In this chapter, we will follow the standard assumption that there are two levels, that is, phrases have at least three projection levels:

- X^0 = head
- X' = intermediate projection (X bar)
- XP = highest projection ($= X'' = \bar{\bar{X}}$), also called *maximal projection*

3.1.8 CP and IP in English

Most work in Mainstream Generative Grammar is heavily influenced by previous publications dealing with English. If one wants to understand GB analyses of German and other languages, it is important to first understand the analyses of English and, for this reason, this will be the focus of this section. The CP/IP system is also assumed in LFG grammars of English and thus the following section also provides a foundation for understanding some of the fundamentals of LFG presented in Chapter 7.

In earlier work, the rules in (22a) and (22b) were proposed for English sentences (Chomsky 1981a: 19).

- (22) a. $S \rightarrow NP VP$
 b. $S \rightarrow NP Infl VP$

Infl stands for *Inflection* as inflectional affixes are inserted at this position in the structure. The symbol AUX was also used instead of Infl in earlier work, since auxiliary verbs are treated in the same way as inflectional affixes. Figure 3.3 on the next page shows a sample analysis of a sentence with an auxiliary, which uses the rule in (22b).

Together with its complements, the verb forms a structural unit: the VP. The constituent status of the VP is supported by several constituent tests and further differences between subjects and objects regarding their positional restrictions.

The rules in (22) do not follow the \bar{X} template since there is no symbol on the right-hand side of the rule with the same category as one on the left-hand side, that is, there is no head. In order to integrate rules like (22) into the general theory, Chomsky (1986a: 3) developed a rule system with two layers above the verb phrase (VP), namely the CP/IP system. CP stands for *Complementizer Phrase*. The head of a CP can be a complementizer. Before we look at CPs in more detail, I will discuss an example of an IP in this new system. Figure 3.4 on the facing page shows an IP with an auxiliary in the I^0 position. As we can see, this corresponds to the structure of the \bar{X} template: I^0 is a head, which takes the VP as its complements and thereby forms I' . The subject is the specifier of the IP.

The sentences in (23) are analyzed as complementizer phrases (CPs), the complementizer is the head:

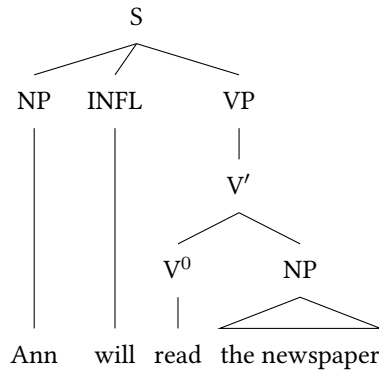


Figure 3.3: English sentence with an auxiliary verb from Chomsky (1981a: 19)

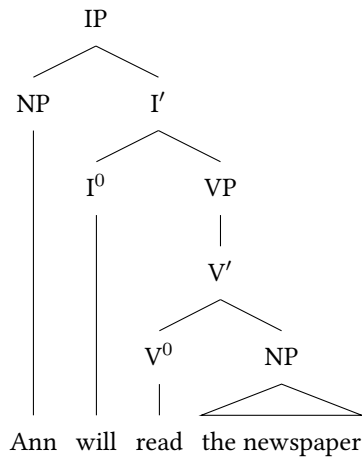


Figure 3.4: English sentence with auxiliary verb in the CP/IP system

- (23) a. that Ann will read the newspaper
b. that Ann reads the newspaper

In sentences such as (23), the CPs do not have a specifier. Figure 3.5 shows the analysis of (23a).

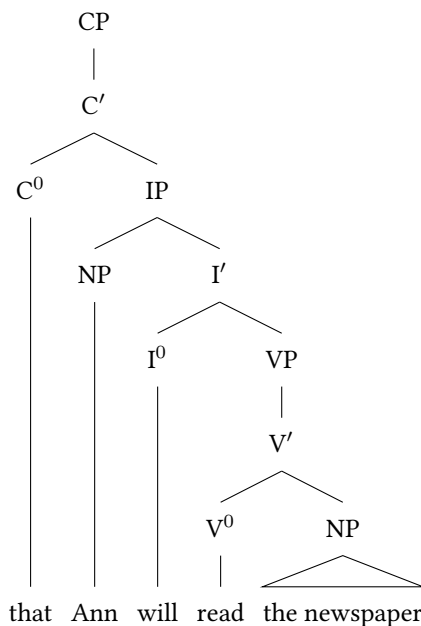


Figure 3.5: English Complementizer Phrase

Yes/no-questions in English such as those in (24) are formed by moving the auxiliary verb in front of the subject.

- (24) Will Ann read the newspaper?

Let us assume that the structure of questions corresponds to the structure of sentences with complementizers. This means that questions are also CPs. Unlike the sentences in (23), however, there is no subordinating conjunction. In the Deep Structure of questions, the C^0 position is empty and the auxiliary verb is moved to this position. Figure 3.6 on the facing page shows an analysis of (24).

The original position of the auxiliary is marked by the trace $_{-k}$, which is coindexed with the moved auxiliary.

wh-questions are formed by the additional movement of a constituent in front of the auxiliary. Figure 3.7 on page 102 shows the analysis of (25):

- (25) What will Ann read?

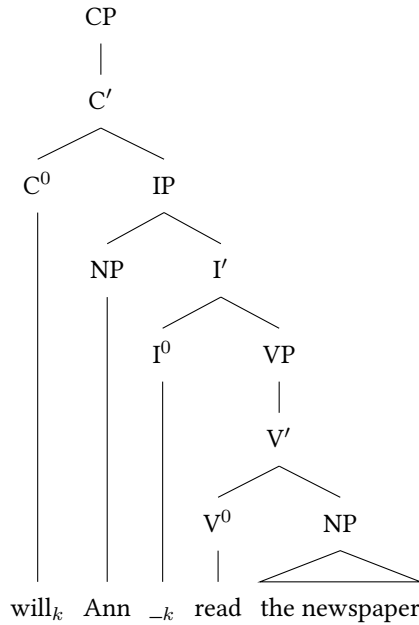


Figure 3.6: English Yes/No-Question

As before, the movement of the object of *read* is indicated by the trace. This is important when constructing the meaning of the sentence. The verb assigns some semantic role to the element in its object position. Therefore, one has to be able to “reconstruct” the fact that *what* actually originates in this position. This is ensured by coindexation of the trace with *what*.

Until now, I have not yet discussed sentences without auxiliaries such as (23b). In order to analyze these kinds of sentences, one has to assume that the inflectional affix is present in the I⁰ position. An example analysis is given in Figure 3.8 on the following page. Since the inflectional affix precedes the verb, some kind of movement operation still needs to take place. For theory-internal reasons, one does not wish to assume movement operations to positions lower in the tree, hence the verb has to move to the affix and not the other way around.

Following this excursus on the analysis of English sentences, we can now turn to German.

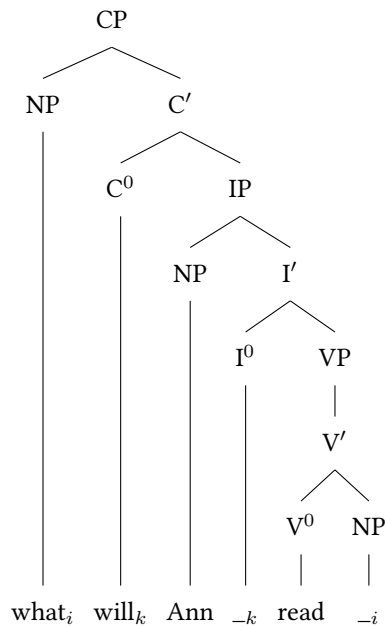


Figure 3.7: English *wh*-question

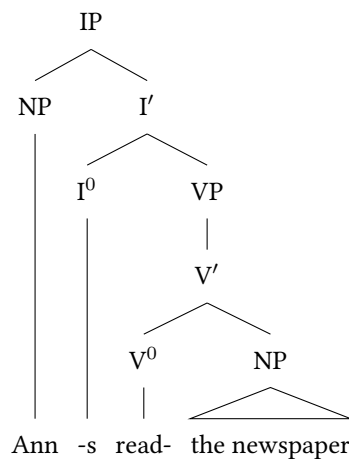


Figure 3.8: English sentence without auxiliary

3.1.9 The structure of the German clause

The CP/IP model has been adopted by many scholars for the analysis of German.⁸ The categories C, I and V, together with their specifier positions, can be linked to the topological fields as shown in Figure 3.9.

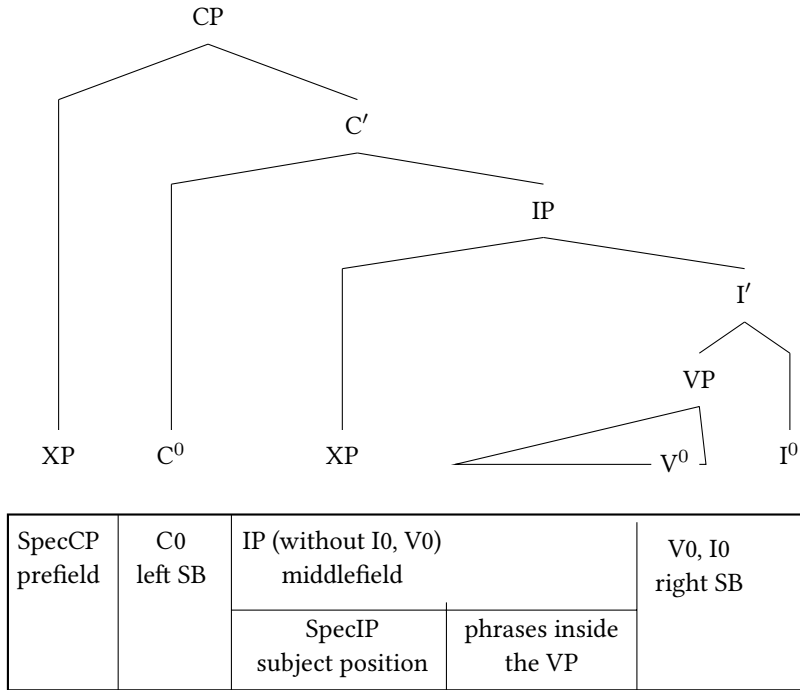


Figure 3.9: CP, IP and VP and the topological model of German

Note that SpecCP and SpecIP are not category symbols. They do not occur in grammars with rewrite rules. Instead, they simply describe positions in the tree.

As shown in Figure 3.9, it is assumed that the highest argument of the verb (the subject in simple sentences) has a special status. It is taken for granted that the subject always occurs outside of the VP, which is why it is referred to as the external argument. The VP itself does not have a specifier. In more recent work, however, the subject is generated in the specifier of the VP (Fukui & Speas 1986; Koopman & Sportiche 1991). In some languages, it is assumed that it moves to a position outside of the VP. In other languages such as German, this is the case at least under certain conditions (e. g. definiteness, see Diesing 1992). I am presenting the classical GB analysis here, where the subject is out-

⁸ For GB analyses without IP, see Bayer & Kornfilt (1989), Höhle (1991: 157), Haider (1993; 1997a) and Sternefeld (2006: Section IV.3). Haider assumes that the verb integrates the function of I. In LFG, an IP is assumed for English, but not for German. In HPSG, no IP is assumed.

side the VP. All arguments other than the subject are complements of the V, that are realized within the VP, that is, they are internal arguments. If the verb requires just one complement, then this is the sister of the head V^0 and the daughter of V' according to the \bar{X} template. The accusative object is the prototypical complement.

Following the \bar{X} template, adjuncts branch off above the complements of V' . The analysis of a VP with an adjunct is shown in Figure 3.10.

- (26) weil der Mann morgen den Jungen trifft
because the man tomorrow the boy meets
‘because the man is meeting the boy tomorrow’

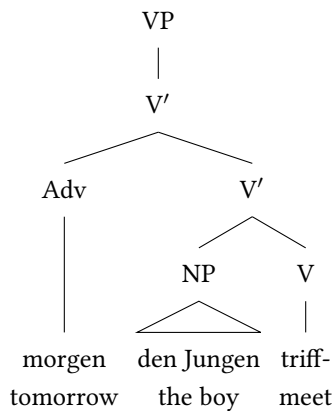


Figure 3.10: Analysis of adjuncts in GB-Theory

3.2 Verb position

In German, the position of the heads of VP and IP (V^0 and I^0) are to the right of their complements and V^0 and I^0 form part of the right sentence bracket. The subject and all other constituents (complements and adjuncts) all occur to the left of V^0 and I^0 and form the middle field. It is assumed that German – at least in terms of D-Structure – is an SOV language (= a language with the base order Subject–Object–Verb). The analysis of German as an SOV language is almost as old as Transformational Grammar itself. It was originally proposed by Bierwisch (1963: 34).⁹ Unlike German, Germanic languages

⁹ Bierwisch attributes the assumption of an underlying verb-final order to Fourquet (1957). A German translation of the French manuscript cited by Bierwisch can be found in Fourquet (1970: 117–135). For other proposals, see Bach (1962), Reis (1974), Koster (1975) and Thiersch (1978: Chapter 1). Analyses which assume that German has an underlying SOV pattern were also suggested in GPSG (Jacobs 1986: 110), LFG (Berman 1996: Section 2.1.4) and HPSG (Kiss & Wesche 1991; Oliva 1992; Netter 1992; Kiss 1993; Frank 1994; Kiss 1995; Feldhaus 1997; Meurers 2000; Müller 2005b).

like Danish, English and Romance languages like French are SVO languages, whereas Welsh and Arabic are VSO languages. Around 40 % of all languages belong to the SOV languages, around 25 % are SVO.

The assumption of verb-final order as the base order is motivated by the follow observations:¹⁰

1. Verb particles form a close unit with the verb.

- (27) a. weil er morgen an-fängt
 because he tomorrow PRT-starts
 ‘because he is starting tomorrow’
 b. Er fängt morgen an.
 he starts tomorrow PRT
 ‘He is starting tomorrow.’

This unit can only be seen in verb-final structures, which speaks for the fact that this structure reflects the base order.

Verbs which are derived from a noun by back-formation (e. g. *uraufführen* ‘to perform something for the first time’), can often not be divided into their component parts and V2 clauses are therefore ruled out (see Haider (1993: 62)):

- (28) a. weil sie das Stück heute urauf-führen
 because they the play today PRT-lead
 ‘because they are performing the play for the first time today’
 b. * Sie uraufführen heute das Stück.
 they PRT-lead today the play
 c. * Sie führen heute das Stück urauf.
 they lead today the play PRT

The examples show that there is only one possible position for the verb. This order is the one that is assumed to be the base order.

2. Verbs in non-finite clauses and in finite subordinate clauses with a conjunction are always in final position (I am ignoring the possibility of extraposing constituents):

- (29) a. Der Clown versucht, Kurt-Martin die Ware zu geben.
 the clown tries Kurt-Martin the goods to give
 ‘The clown is trying to give Kurt-Martin the goods.’
 b. dass der Clown Kurt-Martin die Ware gibt
 that the clown Kurt-Martin the goods gives
 ‘that the clown gives Kurt-Martin the goods’

¹⁰ For points 1 and 2, see Bierwisch (1963: 34–36). For point 4 see Netter (1992: Section 2.3).

3. If one compares the position of the verb in German to Danish (Danish is an SVO language like English), then one can clearly see that the verbs in German form a cluster at the end of the sentence, whereas they occur before any objects in Danish (Ørsnes 2009a):

- (30) a. dass er ihn gesehen₃ haben₂ muss₁
 that he him seen have must
 b. at han må₁ have₂ set₃ ham
 that he must have seen him
 ‘that he must have seen him’

4. The scope relations of the adverbs in (31) depend on their order: The left-most adverb has scope over the two following elements.¹¹ This was explained with the following structure:

¹¹ At this point, it should be mentioned that there seem to be exceptions from the rule that modifiers to the left take scope over those to their right. Kasper (1994: 47) discusses examples such as (i), which go back to Bartsch & Vennemann (1972: 137).

- (i) a. Peter liest gut wegen der Nachhilfestunden.
 Peter reads well because.of the tutoring
 b. Peter liest wegen der Nachhilfestunden gut.
 Peter reads because.of the tutoring well
 ‘Peter can read well thanks to the tutoring.’

As Koster (1975: Section 6) and Reis (1980: 67) have shown, these are not particularly convincing counter-examples as the right sentence bracket is not filled in these examples and it must therefore not necessarily constitute normal reordering inside of the middle field, but could instead be a case of extraposition. As noted by Koster and Reis, these examples become ungrammatical if one fills the right bracket and does not extrapose the causal adjunct:

- (ii) a. *Hans hat gut wegen der Nachhilfestunden gelesen.
 Hans has well because.of the tutoring read
 b. Hans hat gut gelesen wegen der Nachhilfestunden.
 Hans has well read because.of the tutoring
 ‘Hans has been reading well because of the tutoring.’

However, the following example from Crysmann (2004: 383) shows that, even with the right bracket occupied, one can still have an order where an adjunct to the right has scope over one to the left:

- (iii) Da muß es schon erhebliche Probleme mit der Ausrüstung gegeben haben, da wegen
 there must it already serious problems with the equipment given have since because.of
 schlechten Wetters ein Reinhold Messmer niemals aufgab.
 bad weather a Reinhold Messmer never would.give.up
 ‘There really must have been some serious problems with the equipment because someone like Reinhold
 Messmer would never give up just because of some bad weather.’

Nevertheless, this does not change anything regarding the fact that the corresponding cases in (31) and (32) have the same meaning regardless of the position of the verb. The general means of semantic composition may well have to be implemented in the same way as in Crysmann’s analysis.

- (31) a. weil er [absichtlich [nicht lacht]]
 because he intentionally not laughs
 ‘because he is intentionally not laughing’
 b. weil er [nicht [absichtlich lacht]]
 because he not intentionally laughs
 ‘because he is not laughing intentionally’

One can now see that scope relations are not affected by verb position. If one assumes that sentences with verb-second order have the underlying structure in (31), then this fact requires no further explanation. (32) shows the derived S-Structure for (31):

- (32) a. Er lacht_i [absichtlich [nicht _{-i}]].
 he laughs intentionally not
 ‘He is intentionally not laughing.’
 b. Er lacht_i [nicht [absichtlich _{-i}]].
 he laughs not intentionally
 ‘He is not laughing intentionally.’

After motivating and briefly sketching the analysis of verb-final order, I will now look at the CP/IP analysis of German in more detail. C^0 corresponds to the left sentence bracket and can be filled in two different ways: In subordinate clauses introduced by a conjunction, the subordinating conjunction (the complementizer) occupies C^0 as in English. The verb remains in the right sentence bracket, as illustrated by (33).

- (33) dass jeder diesen Mann kennt
 that everybody this man knows
 ‘that everybody knows this man’

Figure 3.11 on the next page gives an analysis of (33). In verb-first and verb-second clauses, the finite verb is moved to C^0 via the I^0 position: $V^0 \rightarrow I^0 \rightarrow C^0$. Figure 3.12 on page 109 shows the analysis of (34):

- (34) Kennt jeder diesen Mann?
 knows everybody this man
 ‘Does everybody know this man?’

The C^0 position is empty in the D-Structure of (34). Since it is not occupied by a complementizer, the verb can move there.

3.3 Long-distance dependencies

The SpecCP position corresponds to the prefield and can be filled by any XP in declarative clauses in German. In this way, one can derive the sentences in (36) from (35) by moving a constituent in front of the verb:

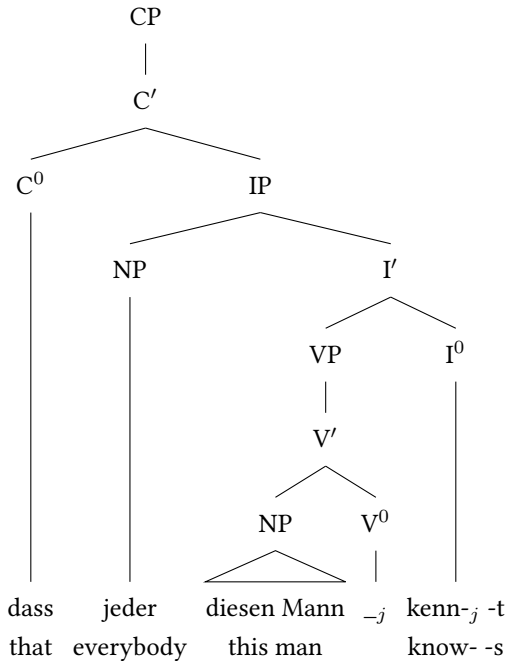


Figure 3.11: Sentence with a complementizer in C⁰

- (35) Gibt der Mann dem Kind jetzt den Mantel?
gives the man the.DAT child now the.ACC coat
'Is the man going to give the child the coat now?'
- (36) a. Der Mann gibt dem Kind jetzt den Mantel.
the man gives the.DAT child now the.ACC coat
'The man is giving the child the coat now.'
b. Dem Kind gibt der Mann jetzt den Mantel.
the.DAT child gives the man now the.ACC coat
c. Den Mantel gibt der Mann dem Kind jetzt.
the.ACC coat gives the man the.DAT child now
d. Jetzt gibt der Mann dem Kind den Mantel.
now gives the man the.DAT child the.ACC coat

Since any constituent can be placed in front of the finite verb, German is treated typologically as one of the verb-second languages (V2). Thus, it is a verb-second language with SOV base order. English, on the other hand, is an SVO language without the V2 property, whereas Danish is a V2 language with SVO as its base order (see Ørsnes (2009a) for Danish).

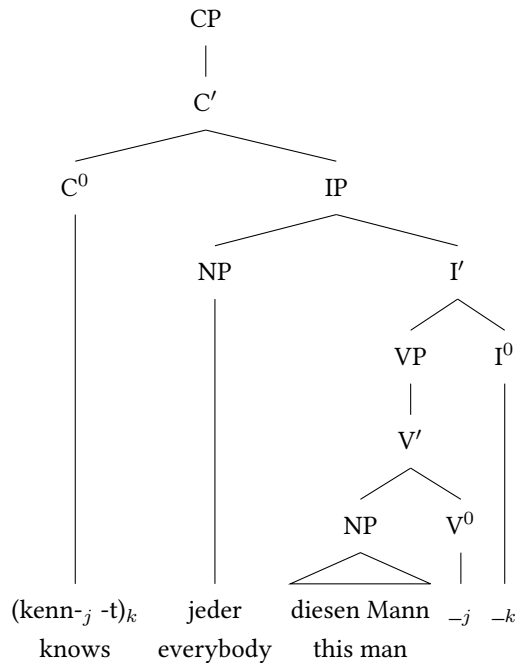


Figure 3.12: Verb position in GB

Figure 3.13 on the next page shows the structure derived from Figure 3.12. The crucial factor for deciding which phrase to move is the *information structure* of the sentence, that is, material connected to previously mentioned or otherwise-known information is placed further left (preferably in the prefield) and new information tends to occur to the right. Fronting to the prefield in declarative clauses is often referred to as *topicalization*. Focus (informally: the constituent being asked for) can also occur in the prefield. Furthermore, expletive pronouns can occur there and these are non-referential and as such cannot be linked to preceding or known information.

Transformation-based analyses also work for so-called *long-distance dependencies*, that is, dependencies over several phrase boundaries:

- (37) a. [Um zwei Millionen Mark]_i soll er versucht haben, [eine
 around two million Deutsche.Marks should he tried have an
 Versicherung _{-i} zu betrügen].¹²
 insurance.company to deceive
 ‘He apparently tried to cheat an insurance company out of two million
 Deutsche Marks.’

¹² taz, 04.05.2001, p. 20.

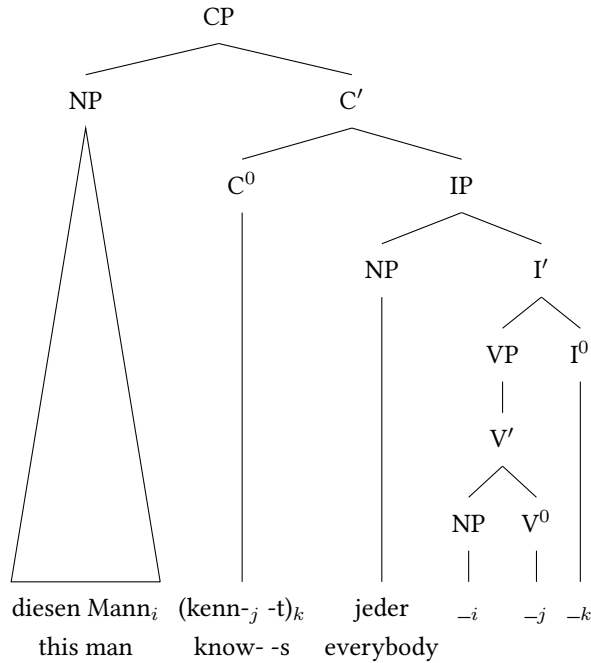


Figure 3.13: Fronting in GB theory

- b. „Wer_i, glaubt er, daß er _i ist?“ erregte sich ein Politiker vom Nil.¹³
 who believes he that he is retort REFL a politician from.the Nile
 ‘Who does he think he is?’¹³, a politician from the Nile exclaimed.’
- c. Wen_i glaubst du, daß ich _i gesehen habe.¹⁴
 who believe you that I seen have
 ‘Who do you think I saw?’
- d. [Gegen ihn]_i falle es den Republikanern hingegen schwerer,
 against him fall it the Republicans however more.difficult
 [[Angriffe _i] zu lancieren].¹⁵
 attacks to launch
 ‘It is, however, more difficult for the Republicans to launch attacks against him.’

The elements in the prefield in the examples in (37) all originate from more deeply embedded phrases. In GB, it is assumed that long-distance dependencies across sentence

¹³ Spiegel, 8/1999, p. 18.

¹⁴ Scherpenisse (1986: 84).

¹⁵ taz, 08.02.2008, p. 9.

boundaries are derived in steps (Grewendorf 1988: 75–79), that is, in the analysis of (37c), the interrogative pronoun is moved to the specifier position of the *dass*-clause and is moved from there to the specifier of the matrix clause. The reason for this is that there are certain restrictions on movement which must be checked locally.

3.4 Passive

Before I turn to the analysis of the passive in Section 3.4.2, the first subsection will elaborate on the differences between structural and lexical case.

3.4.1 Structural and lexical case

The case of many case-marked arguments is dependent on the syntactic environment in which the head of the argument is realized. These arguments are referred to as arguments with *structural case*. Case-marked arguments, which do not bear structural case, are said to have *lexical case*.¹⁶

The following are examples of structural case:¹⁷

- (38) a. Der Installateur kommt.
the.NOM plumber comes
'The plumber is coming.'
- b. Der Mann lässt den Installateur kommen.
the man lets the.ACC plumber come
'The man is getting the plumber to come.'
- c. das Kommen des Installateurs
the coming of.the plumber
'the plumber's visit'

In the first example, the subject is in the nominative case, whereas *Installateur* ('plumber') is in accusative in the second example and even in the genitive in the third following nominalization. The accusative case of objects is normally structural case. This case becomes nominative under passivization:

¹⁶ Furthermore, there is a so-called *agreeing case* (see page 43) and *semantic case*. Agreeing case is found in predicatives. This case also changes depending on the structure involved, but the change is due to the antecedent element changing its case. Semantic case depends on the function of certain phrases (e.g. temporal accusative adverbials). Furthermore, as with lexical case of objects, semantic case does not change depending on the syntactic environment. For the analysis of the passive, which will be discussed in this section, only structural and lexical case will be relevant.

¹⁷ Compare Heinz & Matiaszek (1994: 200).

(38b) is a so-called Acl construction. Acl stands for *Accusativus cum infinitivo*, which means "accusative with infinitive". The logical subject of the embedded verb (*kommen* in this case) becomes the accusative object of the matrix verb *lassen*. Examples for Acl-verbs are perception verbs such as *hören* ('to hear') and *sehen* ('to see') as well as *lassen* ('to let').

- (39) a. Karl schlägt den Weltmeister.
Karl beats the.ACC world.champion
'Karl beats the world champion.'
b. Der Weltmeister wird geschlagen.
the.NOM world.champion is beaten
'The world champion is being beaten.'

Unlike the accusative, the genitive governed by a verb is a lexical case. The case of a genitive object does not change when passivized.

- (40) a. Wir gedenken der Opfer.
we remember the.GEN victims
b. Der Opfer wird gedacht.
the.GEN victims are remembered
'The victims are being remembered.'

(40b) is an example of the so-called *impersonal passive*. Unlike example (39b), where the accusative object became the subject, there is no subject in (40b). See Section 1.7.1.

Similarly, there is no change in case with dative objects:

- (41) a. Der Mann hat ihm geholfen.
the man has him.DAT helped
'The man has helped him.'
b. Ihm wird geholfen.
him.DAT is helped
'He is being helped.'

It still remains controversial as to whether some or all of the datives in verbal environments should be treated as instances of structural case. For reasons of space, I will not recount this discussion but instead refer the interested reader to Chapter 14 of Müller (2007b). In what follows, I assume – like Haider (1986a: 20) – that the dative is in fact a lexical case.

3.4.2 Case assignment and the Case Filter

In GB, it is assumed that the subject receives case from (finite) I and that the case of the remaining arguments comes from V (Chomsky 1981a: 50; Haider 1984: 26; Fanselow & Felix 1987: 71–73).

Principle 2 (Case Principle)

- V assigns objective case (accusative) to its complement if it bears structural case.
- When finite, INFL assigns case to the subject.

The Case Filter rules out structures where case has not been assigned to an NP.

Figure 3.14 shows the Case Principle in action with the example in (42a).¹⁸

- (42) a. [dass] der Mann der Frau den Jungen zeigt
 that the man the.DAT woman the.ACC boy shows
 ‘that the man shows the boy to the woman’
 b. [dass] der Junge der Frau gezeigt wird
 that the boy.NOM the.DAT woman shown is
 ‘that the boy is shown to the woman’

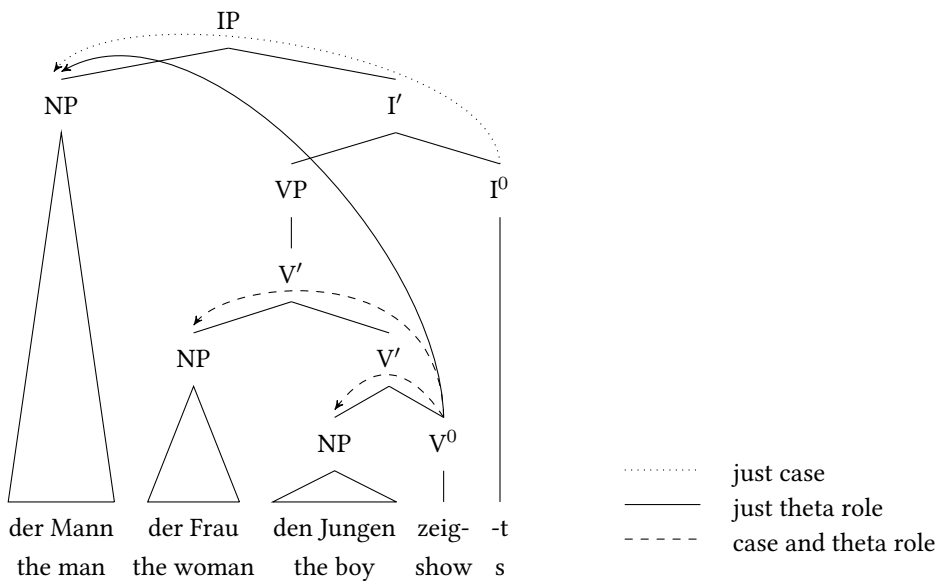


Figure 3.14: Case and theta-role assignment in active clauses

The passive morphology blocks the subject. The object that would get accusative in the active receives a semantic role in passives, but it does not get case. Therefore, it has to move to a position where case can be assigned to it (Chomsky 1981a: 124). Figure 3.15 on the next page shows how this works for example (42b): This movement-based analysis works well for English since the underlying object always has to be moved:

- (43) a. The mother gave [the girl] [a cookie].

¹⁸ The figure does not correspond to \bar{X} theory in its classic form, since *der Frau* ('the woman') is a complement which is combined with V' . In classical \bar{X} theory, all complements have to be combined with V^0 . Furthermore, in the following figures the verb has been left in V^0 for reasons of clarity. In order to create a well-formed S-Structure, the verb would have to move to its affix in I^0 .

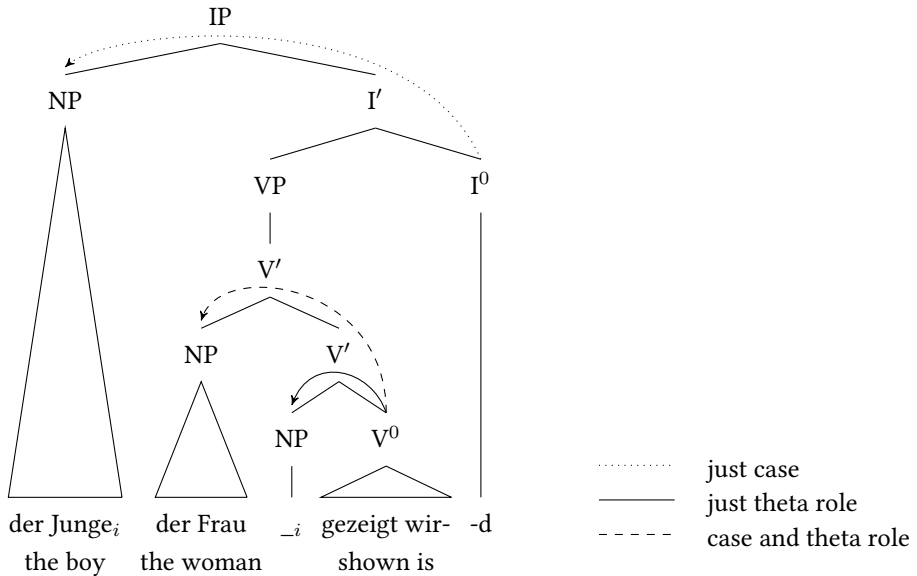


Figure 3.15: Case and theta-role assignment in passive clauses

- b. [The girl] was given [a cookie] (by the mother).
- c. * It was given [the girl] [a cookie].

(43c) shows that the subject position cannot be filled by an expletive.

Lenerz (1977: Section 4.4.3) showed that such a movement is not obligatory in German:

- (44) a. weil das Mädchen dem Jungen den Ball schenkt
because the girl the.DAT boy the.ACC ball gives
‘because the girl gives the ball to the boy’
- b. weil dem Jungen der Ball geschenkt wurde
because the.DAT boy the.NOM ball given was
- c. weil der Ball dem Jungen geschenkt wurde
because the.NOM ball the.DAT boy given was
‘because the ball was given to the boy’

In comparison to (44c), (44b) is the unmarked order. *der Ball* (‘the ball’) in (44b) occurs in the same position as *den Ball* in (44a), that is, no movement is necessary. Only the case differs. (44c) is, however, somewhat marked in comparison to (44b). The analysis which has been proposed for cases such as (44b) involves abstract movement: The elements stay in their positions, but are connected to the subject position and receive their case information from there. Grewendorf (1993: 1311) assumes that there is an empty expletive

pronoun in the subject position of sentences such as (44b) as well the the subject position of sentences with an impersonal passive such as (45):¹⁹

- (45) weil heute nicht gearbeitet wird
 because today not worked is
 ‘because there will be no work done today’

A silent expletive pronoun is something that one cannot see or hear and that does not carry any meaning. For discussion of these kind of elements, see Section 13.1.3 and Chapter 21.

In the following chapters, I describe alternative treatments of the passive, which do without mechanisms such as merging empty elements in argument positions and seek to describe the passive in a more general, cross-linguistic manner as the consistent suppression of the most prominent argument.

A further question which needs to be answered is why the accusative object does not receive case from the verb. This is captured by a constraint, which goes back to Burzio (1986: 178–185) and is therefore referred to as *Burzio’s Generalization*.²⁰

- (46) Burzio’s Generalization (modified):
 If V does not have an external argument, then it does not assign (structural) accusative case.

Koster (1986: 12) has pointed out that the passive in English cannot be derived by Case Theory since if one allows empty explicit subjects for English as well as German and

¹⁹ See Koster (1986: 11–12) for a parallel analysis for Dutch as well as Lohnstein (2014), who also assumes an empty expletive, for a movement-based account of the passive.

²⁰ Burzio’s original formulation was equivalent to the following: A verb assigns accusative, if and only if it assigns a semantic role to its subject.

This claim is problematic from both sides. In (i), the verb does not assign a semantic role to the subject, however there is nevertheless accusative case:

- (i) Mich friert.
 me.ACC freezes
 ‘I am freezing.’

One therefore has to differentiate between structural and lexical accusative and modify Burzio’s Generalization accordingly. The existence of verbs like *begegnen* (‘to bump into’) is problematic for the other side of the implication. *begegnen* has a subject but still does not assign accusative but rather dative:

- (ii) Peter begegnete einem Mann.
 Peter met a.DAT man
 ‘Peter met a man.’

Burzio (1986: 185) assumes that one-place intransitive verbs have the potential to assign accusative. He discusses resultative constructions such as (ii):

- (iii) He talked my head off.

There are also verbs such as *verschwinden* (‘to disappear’), which cannot assign accusative in such constructions.

See Haider (1999) and Webelhuth (1995: 89) as well as the references cited there for further problems with Burzio’s Generalization.

Dutch, then it would be possible to have analyses such as the following in (47) where np is an empty expletive:

- (47) np was read the book.

Koster also assumes that subjects in English have to be filled either by a non-expletive or lexical item, that is, by visible material. Therefore, the structure in (47) would be ruled out and it would be ensured that *the book* would have to be placed in front of the finite verb so that the subject position is filled.

3.5 Local reordering

Arguments in the middle field can, in principle, occur in an almost arbitrary order. (48) exemplifies this:

- (48) a. [weil] der Mann der Frau das Buch gibt
because the man the woman the book gives
'because the man gives the book to the woman'
b. [weil] der Mann das Buch der Frau gibt
because the man the book the woman gives
c. [weil] das Buch der Mann der Frau gibt
because the book the man the woman gives
d. [weil] das Buch der Frau der Mann gibt
because the book the woman the man gives
e. [weil] der Frau der Mann das Buch gibt
because the woman the man the book gives
f. [weil] der Frau das Buch der Mann gibt
because the woman the book the man gives

In (48b–f), the constituents receive different stress and the number of contexts in which each sentence can be uttered is greatly more restricted than (48a) (Höhle 1982). The order in (48a) is therefore referred to as the *neutral order* or *unmarked order*.

Two proposals have been made for analysing these orders: The first suggestion assumes that the five orderings in (48) are derived from a single underlying order by means of move α (Frey 1993). An argument that has often been used to support this analysis is the fact that even after the supposed movement of the arguments, there are scope ambiguities which are not present in the base order. The explanation of ambiguities comes from the assumption that the scope of quantifiers can be derived from their position in the surface structure as well as their position in the deep structure. If the position in both the surface and deep structure are the same, that is when there has not been any movement, then there is only one reading possible. If movement has taken place, however, then there are two possible readings (Frey 1993):

- (49) a. Es ist nicht der Fall, daß er mindestens einem Verleger fast jedes
 it is not the case that he at.least one publisher almost every
 Gedicht anbot.
 poem offered
 'It is not the case that he offered at least one publisher almost every poem.'
- b. Es ist nicht der Fall, daß er fast jedes Gedicht_i mindestens einem
 it is not the case that he almost every poem at.least one
 Verleger _i anbot.
 publisher offered
 'It is not the case that he offered almost every poem to at least one publisher.'

It turns out that approaches assuming traces run into problems as they predict certain readings for sentences with multiple traces, which do not exist (see Kiss (2001: 146) and Fanselow (2001: Section 2.6)). For instance in an example such as (50), it should be possible to interpret *mindestens einem Verleger* ('at least one publisher') at the position of _i, which would lead to a reading where *fast jedes Gedicht* ('almost every poem') has scope over *mindestens einem Verleger* ('at least one publisher').

- (50) Ich glaube, dass mindestens einem Verleger_i fast jedes Gedicht_j nur dieser
 I believe that at.least one publisher almost every poem only this
 Dichter _i _j angeboten hat.
 poet offered has
 'I think that only this poet offered almost every poem to at least one publisher.'

This reading does not exist, however.

Sauerland & Elbourne (2002: 308) discuss analogous examples from Japanese, which they credit to Kazuko Yatsushiro. They develop an analysis where the first step is to move the accusative object in front of the subject. Then, the dative object is placed in front of that and then in a third movement, the accusative is then moved once more. The last movement can take place to construct either the S-Structure²¹ or as a movement to construct the phonological form. In the latter case, this movement will not have any semantic effects. While this analysis can predict the correct available readings, it does require a number of additional movement operations with intermediate steps.

The alternative to a movement analysis is so-called *base generation*: The starting structure generated by phrase structure rules is referred to as the *base*. A variant of the base generation is the assumption that the verb is combined with an argument and then assigns it a θ -role. The order in which they are combined is not specified, which means that all of the orders in (48) can be generated directly without any transformations (compare this to the grammar in (6) on page 57). This kind of analysis has been proposed for GB by Fanselow (2001).²² For the discussion of different approaches to describing constituent position, see Fanselow (1993).

²¹ The authors are working in the Minimalist framework. This means there is no longer S-Structure strictly speaking. I have simply translated the analysis into the terms used here.

²² The base generation analysis is the natural analysis in the HPSG framework. It has already been developed by Gunji in 1986 for Japanese and will be discussed in more detail in Section 9.4. Sauerland & Elbourne

3.6 Summary and classification

Works in GB some contributions to the Minimalist Program (see Chapter 4) have led to a number of new discoveries in both language-specific and cross-linguistic research. In the following, I will focus on some aspects of German syntax.

The analysis of verb movement developed in Transformational Grammar by Bierwisch (1963: 34), Reis (1974), Koster (1975), Thiersch (1978: Chapter 1) and den Besten (1983) has become the standard analysis in almost all grammar models (possibly with the exception of Construction Grammar).

The work by Lenerz on constituent order (1977) has influenced analyses in other frameworks (the linearization rules in GPSG and HPSG go back to Lenerz' descriptions). Haider's work on constituent order, case and passive (1984; 1985b; 1985a; 1986a; 1990b; 1993) has had a significant influence on LFG and HPSG analyses of German.

The entire configurationality discussion, that is, whether it better to assume that the subject of finite verbs in German is inside or outside the VP, was important (for instance Haider (1982); Grewendorf (1983); Kratzer (1984); Webelhuth (1985); Sternefeld (1985b); Scherpenisse (1986); Fanselow (1987); Grewendorf (1988); Dürscheid (1989); Webelhuth (1990); Oppenrieder (1991); Wilder (1991); Haider (1993); Grewendorf (1993); Frey (1993); Lenerz (1994); Meinunger (2000)) and German unaccusative verbs received their first detailed discussion in GB circles (Grewendorf 1989; Fanselow 1992a). The works by Fanselow and Frey on constituent order, in particular with regard to information structure, have advanced German syntax quite considerably (Fanselow 1988; 1990; 1993; 2000a; 2001; 2003b;c; 2004a; Frey 2000; 2001; 2004a; 2005). Infinitive constructions, complex predicates and partial fronting have also received detailed and successful treatments in the GB/MP frameworks (Bierwisch (1963); Evers (1975); Haider (1982; 1986b; 1990a; 1991; 1993); Grewendorf (1983; 1987; 1988); den Besten (1985); Sternefeld (1985b); Fanselow (1987; 2002); von Stechow & Sternefeld (1988); Bayer & Kornfilt (1989), G. Müller (1996a; 1998); Vogel & Steinbach (1998)). In the area of secondary predication, the work by Susanne Winkler (1997) is particularly noteworthy.

This list of works from subdisciplines of grammar is somewhat arbitrary (it corresponds more or less to my own research interests) and is very much focussed on German. There are, of course, a wealth of other articles on other languages and phenomena, which should be recognized without having to be individually listed here.

In this section, I will critically discuss two points: the model of language acquisition of the Principles & Parameters model and the degree of formalization inside Chomskyan linguistics (in particular the last few decades and the consequences this has). Some of these points will be mentioned again in Part II.

(2002: 313–314) claim that they show that syntax has to be derivational, that is, a sequence of words must come from a syntactic tree. I am of the opinion that this cannot generally be shown to be the case. There is, for example, an analysis by Kiss (2001) which shows that scope phenomena can be explained well by constraint-based approaches.

3.6.1 Explaining language acquisition

One of the aims of Chomskyan research on grammar is to explain language acquisition. In GB, one assumed a very simple set of rules, which was the same for all languages (\bar{X} theory) as well as general principles that held for all languages and others which were parameterized for individual languages or language classes. It was assumed that a parameter was relevant for multiple phenomena. The Principles & Parameters model was particularly fruitful and led to a number of interesting studies in which commonalities and differences between languages were uncovered. From the point of view of language acquisition, the idea of a parameter which is set according to the input has often been criticized as it cannot be reconciled with observable facts: After setting a parameter, a learner would have to immediately have mastered certain aspects of that language. Chomsky (1986b: 146) uses the metaphor of switches which can be flipped one way or the other. As it is assumed that various areas of grammar are affected by parameters, setting one parameter should have a significant effect on the rest of the grammar of a given learner. However, the linguistic behaviour of children does not change in an abrupt fashion as would be expected (Bloom 1993: 731; Haider 1993: 6; Abney 1996: 3; Ackerman & Webelhuth 1998: Section 9.1; Tomasello 2000; 2003; Newmeyer 2005). Furthermore, it has not been possible to prove that there is a correlation between a certain parameter and various grammatical phenomena. For more on this, see Section 16.

The Principles and Parameters model nevertheless remains interesting for cross-linguistic research. Every theory has to explain why the verb precedes its objects in English and follows them in Japanese. One can name this difference a parameter and then classify languages accordingly. Whether this is actually relevant for language acquisition is being increasingly called in question.

3.6.2 Formalization

In his 1963 work on Transformational Grammar, Bierwisch writes the following:²³

It is very possible that the rules that we formulated generate sentences which are outside of the set of grammatical sentences in an unpredictable way, that is, they violate grammaticality due to properties that we did not deliberately exclude in our examination. This is meant by the statement that a grammar is a hypothesis about the structure of a language. A systematic check of the implications of a grammar that is appropriate for natural languages is surely a task that cannot be done by hand any more. This task could be solved by implementing the grammar

²³ Es ist also sehr wohl möglich, daß mit den formulierten Regeln Sätze erzeugt werden können, die auch in einer nicht vorausgesehenen Weise aus der Menge der grammatisch richtigen Sätze herausfallen, die also durch Eigenschaften gegen die Grammatikalität verstoßen, die wir nicht wissentlich aus der Untersuchung ausgeschlossen haben. Das ist der Sinn der Feststellung, daß eine Grammatik eine Hypothese über die Struktur einer Sprache ist. Eine systematische Überprüfung der Implikationen einer für natürliche Sprachen angemessenen Grammatik ist sicherlich eine mit Hand nicht mehr zu bewältigende Aufgabe. Sie könnte vorgenommen werden, indem die Grammatik als Rechenprogramm in einem Elektronenrechner realisiert wird, so daß überprüft werden kann, in welchem Maße das Resultat von der zu beschreibenden Sprache abweicht.

as a calculating task on a computer so that it becomes possible to verify to which degree the result deviates from the language to be described. (Bierwisch 1963: 163)

Bierwisch's claim is even more true in light of the empirical progress made in the last centuries. For example, Ross (1967) identified restrictions for movement and long-distance dependencies and Perlmutter (1978) discovered unaccusative verbs in the 70s. For German, see Grewendorf (1989) and Fanselow (1992a). Aside from analyses of these phenomena, restrictions on possible constituent positions have been developed (Lenerz 1977), as well as analyses of case assignment (Yip, Maling & Jackendoff 1987; Meurers 1999c; Przepiórkowski 1999b) and theories of verbal complexes and the fronting of parts of phrases (Evers 1975; Grewendorf 1988; Hinrichs & Nakazawa 1994; Kiss 1995; G. Müller 1998; Meurers 1999b; Müller 1999a; 2002a; De Kuthy 2002). All these phenomena interact!

Consider another quote:

A goal of earlier linguistic work, and one that is still a central goal of the linguistic work that goes on in computational linguistics, is to develop grammars that assign a reasonable syntactic structure to every sentence of English, or as nearly every sentence as possible. This is not a goal that is currently much in fashion in theoretical linguistics. Especially in Government-Binding theory (GB), the development of large fragments has long since been abandoned in favor of the pursuit of deep principles of grammar. The scope of the problem of identifying the correct parse cannot be appreciated by examining behavior on small fragments, however deeply analyzed. Large fragments are not just small fragments several times over—there is a qualitative change when one begins studying large fragments. As the range of constructions that the grammar accommodates increases, the number of undesired parses for sentences increases dramatically. (Abney 1996: 20)

Additionally, the aim is to formulate restrictions which ideally hold for all languages or at least for certain language classes. It follows from this, that one has to have an overview of the interaction of various phenomena in not just one but several languages. This task is so complex that individual researchers cannot manage it. This is the point at which computer implementations become helpful as they immediately flag inconsistencies in the theory. After removing these inconsistencies, computer implementations can be used to systematically analyze test data or corpora and thereby check the empirical adequacy of the theory (Müller, 1999a: Chapter 22; 2015a; 2014d; Oepen & Flickinger 1998; Bender 2008b, see Section 1.2).

More than 50 years after the first important published work by Chomsky, it is apparent that there has not been a large-scale implemented grammatical fragment on the basis of Transformational Grammar analyses. Chomsky has certainly contributed to the formalization of linguistics and developed important formal foundations, which are still relevant in the theory of formal languages in computer science and in theoretical computational linguistics (Chomsky 1959). However, in 1981, he had already turned his back on rigid formalization:

I think that we are, in fact, beginning to approach a grasp of certain basic principles of grammar at what may be the appropriate level of abstraction. At the same time, it is necessary to investigate them and determine their empirical adequacy by developing quite specific mechanisms. We should, then, try to distinguish as clearly as we can between discussion that bears on leading ideas and discussion that bears on the choice of specific realizations of them. (Chomsky 1981a: 2–3)

This is made explicit in a letter to *Natural Language and Linguistic Theory*:

Even in mathematics, the concept of formalization in our sense was not developed until a century ago, when it became important for advancing research and understanding. I know of no reason to suppose that linguistics is so much more advanced than 19th century mathematics or contemporary molecular biology that pursuit of Pullum's injunction would be helpful, but if that can be shown, fine. For the present, there is lively interchange and exciting progress without any sign, to my knowledge, of problems related to the level of formality of ongoing work. (Chomsky 1990: 146)

This departure from a single rigid formalism has led to there being a large number of publications inside Mainstream Generative Grammar with sometimes incompatible assumptions to the point where it is no longer clear how one can combine the insights of the various publications.

An example of this is the fact that the central notion of government has several different definitions (see Aoun & Sportiche (1983) for an overview²⁴).

This situation has been criticized repeatedly since the 80s and sometimes very harshly by proponents of GPSG (Gazdar, Klein, Pullum & Sag 1985: 6; Pullum 1985; 1989a; Pullum 1991: 48; Kornai & Pullum 1990).

The lack of precision and working out of the details²⁵ and the frequent modification of basic assumptions²⁶ has led to insights gained by Mainstream Generative Grammar rarely being translated into computer implementations. There are some implementations based on Transformational Grammar/GB/MP models or which borrow ideas from Mainstream Generative Grammar (Petrick 1965; Zwicky, Friedman, Hall & Walker 1965; Kay 1967; Friedman 1969; Friedman, Brecht, Doran, Pollack & Martner 1971; Morin 1973; Marcus 1980; Abney & Cole 1986; Kuhns 1986; Correa 1987; Stabler 1987; 1992; 2001; Kolb & Thiersch 1991; Fong 1991; Crocker & Lewin 1992; Lohnstein 1993; Fordham & Crocker 1994; Nordgård 1994; Veenstra 1998; Fong & Ginsburg 2012).²⁷ These implementations often do not use transformations or differ greatly from the theoretical assumptions of the publications. For example, Marcus (1980: 102–104) and Stabler (1987: 5) use rules

²⁴ A further definition can be found in Aoun & Lightfoot (1984). This is, however, equivalent to an earlier version as shown by Postal & Pullum (1986: 104–106).

²⁵ See e. g. Kuhns (1986: 550), Crocker & Lewin (1992: 508), Kolb & Thiersch (1991: 262), Kolb (1997: 3) and Freidin (1997: 580), Veenstra (1998: 25, 47), Lappin et al. (2000a: 888) and Stabler (2010: 397, 399, 400) for the latter.

²⁶ See e. g. Kolb (1997: 4), Fanselow (2009) and the quote from Stabler on page 175.

²⁷ See Fordham & Crocker (1994) for a combination of a GB approach with statistical methods.

which are only responsible for auxiliary movement.²⁸ These rules reverse the order of *John* and *has* for the analysis of sentences such as (51a) so that we get the order in (51b), which is then parsed with the rules for non-inverted structures.

- (51) a. Has John scheduled the meeting for Wednesday?
b. John has scheduled the meeting for Wednesday?

These rules for auxiliary movement are very specific and explicitly reference the category of the auxiliary. This does not correspond to the analyses proposed in GB in any way. As we have seen in Section 3.1.8, there are no special transformational rules for auxiliary movement. Auxiliary movement is carried out by the more general transformation move α and the associated restrictive principles. It is not unproblematic that the explicit formulation of the rule refers to the category *auxiliary* as is clear when one views Stabler's GB-inspired phrase structure grammar:

- (52) a. $s \rightarrow \text{switch}(\text{aux_verb}, \text{np}), \text{vp}$.
b. $s([\text{First}|\text{L0}], \text{L}, \text{X0}, \text{X}) :- \text{aux_verb}(\text{First}),$
 $\text{np}(\text{L0}, \text{L1}, \text{X0}, \text{X1}),$
 $\text{vp}([\text{First}|\text{L1}], \text{L}, \text{X1}, \text{X})$.

The rule in (52a) is translated into the Prolog predicate in (52b). The expression $[\text{First}|\text{L0}]$ after the s corresponds to the string, which is to be processed. The $'|'$ -operator divides the list into the beginning and the rest. *First* is the first word to be processed and L0 contains all other words. In the analysis of (51a), *First* is *has* and L0 is *John scheduled the meeting for Wednesday*. In the Prolog clause, it is then checked whether *First* is an auxiliary ($\text{aux_verb}(\text{First})$) and if this is the case, then it will be tried to prove that the list L0 begins with a noun phrase. Since *John* is an NP, this is successful. L1 is the sublist of L0 which remains after the analysis of L0 , that is *scheduled the meeting for Wednesday*. This list is then combined with the auxiliary (*First*) and now it will be checked whether the resulting list *has scheduled the meeting for Wednesday* begins with a VP. This is the case and the remaining list L is empty. As a result, the sentence has been successfully processed.

The problem with this analysis is that one particular word is checked in the lexicon. Sentences such as (53) can not be analyzed:²⁹

- (53) Could or should we pool our capital with that of other co-ops to address the needs of a regional “neighborhood”?³⁰

In this kind of sentence, two modal verbs have be coordinated. They then form an X^0 and – following GB analyses – can be moved together. If one wanted to treat these cases

²⁸ Nozohoor-Farshi (1986; 1987) has shown that Marcus' parser can only parse context-free languages. Since natural languages are of a greater complexity (see Chapter 18) and grammars of corresponding complexity are allowed by current versions of Transformational Grammar, Marcus' parser can be neither an adequate implementation of the Chomskyan theory in question nor a piece of software for analyzing any natural language.

²⁹ For arguments on the coordination of lexical elements, see Abeillé (2006).

³⁰ <http://www.cooperativegrocer.coop/articles/index.php?id=595>. 28.03.2010.

as Stabler does for the simplest case, then we would need to divide the list of words to be processed into two unlimited sub-lists and check whether the first list contains an auxiliary or several coordinated auxiliaries. We would require a recursive predicate `aux_verbs` which somehow checks whether the sequence *could or should* is a well-formed sequence of auxiliaries. This should not be done by a special predicate but rather by syntactic rules responsible for the coordination of auxiliaries. The alternative to a rule such as (52a) would be the one in (54), which is the one that is used in theories like GPSG (Gazdar et al. 1985: 62), LFG (Falk 1984: 491), some HPSG analyses (Ginzburg & Sag 2000: 36), and Construction Grammar (Fillmore 1999):

$$(54) \quad s \rightarrow v(\text{aux+}), \text{np}, \text{vp}.$$

This rule would have no problems with coordination data like (53) as coordination of multiple auxiliaries would produce an object with the category `v(aux+)` (for more on coordination see Section 23.6.2). If inversion makes it necessary to stipulate a special rule like (52a), then it is not clear why one could not simply use the transformation-less rule in (54).

In the MITRE system (Zwicky et al. 1965), there was a special grammar for the surface structure, from which the deep structure was derived via backward application of transformations, that is, instead of using one grammar to create deep structures, which are then transformed into other structures, one required two grammars. The deep structures that were determined by the parser were used as input to a transformational component since this was the only way to ensure that surface structure can actually be derived from the base structure (Kay 2011: 10).

There are other implementations discussed in this chapter that differ from transformation-based analyses. For example, Kolb & Thiersch (1991: 265, Section 4) arrive at the conclusion that a declarative, constraint-based approach to GB is more appropriate than a derivational one. Johnson (1989) suggests a *Parsing as Deduction* approach which reformulates sub-theories of GB (\bar{X} theory, Theta-Theory, Case Theory, ...) as logical expressions.³¹ These can be used independently of each other in a logical proof. In Johnson's analysis, GB theory is understood as a constraint-based system. More general restrictions are extracted from the restrictions on S- and D-Structure, which can then be used directly for parsing. This means that transformations are not directly carried out by the parser. As noted by Johnson, the language fragment he models is very small. It contains no description of *wh*-movement, for example (p. 114).

Probably the most detailed implementation of the GB/Barriers tradition is Stabler's Prolog implementation (1992). Stabler's achievement is certainly impressive, but his book confirms what has been claimed thus far: Stabler has to simply stipulate many things which are not explicitly mentioned in *Barriers* (e.g. using feature-value pairs when formalizing \bar{X} theory, which was borrowed from GPSG) and some assumptions cannot be properly formalized and are simply ignored (see Briscoe (1997) for details).

³¹ See Crocker & Lewin (1992: 511) and Fordham & Crocker (1994: 38) for another constraint-based Parsing-as-Deduction approach.

GB analyses which fulfill certain requirements can be reformulated so that they no longer make use of transformations. These transformation-less approaches are also called *representational*, whereas the transformation-based approaches are referred to as *derivational*. For representational analyses, there are only surface structures augmented by traces but none of these structures are connected to an underlying structure by means of transformations (see Koster: 1978; 1987: 235; Kolb & Thiersch 1991; Haider 1993: Section 1.4; Frey 1993: 14; Lohnstein 1993: 87–88, 177–178; Fordham & Crocker 1994: 38; Veenstra 1998: 58, for example).

These analyses can be implemented in the same way as corresponding HPSG analyses (see Chapter 9) as computational systems and this has in fact been carried out for the analysis of verb position in German.³² These analyses differ from GB analyses with regard to their basic architecture and in small, but important details such as how one deals with the interaction of long-distance dependencies and coordination (Gazdar 1981b). For a critical discussion and classification of movement analyses in Transformational Grammar, see Borsley (2012).

In conclusion of this somewhat critical overview, I offer the following comment: I do not wish to be understood to be claiming that all linguistic work should be completely formalized. There is simply not space for this in a, say, thirty page essay. Furthermore, I do not believe that all linguists should carry out formal work and implement their analyses as computational models. However, there has to be *somebody* who works out the formal details and these basic theoretical assumptions should be accepted and adopted for a sufficient amount of time by the research community in question.

Comprehension questions

1. Give some examples of functional and lexical categories.
2. How can one represent lexical categories with binary features and what advantages does this have?

Exercises

1. Draw syntactic trees for the following examples:

(55) a. dass die Frau den Mann liebt
 that the woman the.ACC man loves
 ‘that the woman loves the man’

³² This shows that ten Hacken’s comparison of HPSG and GB (ten Hacken 2007: Section 4.3) and the classification of these frameworks as belonging to different research paradigms is completely mistaken. In his classification, ten Hacken refers mainly to the model-theoretic approach that HPSG assumes. However, LFG also has a model-theoretic formalization (Kaplan 1995). Furthermore, there is also a model-theoretic variant of GB (Rogers 1998). See Chapter 14.

- b. dass der Mann geliebt wird
that the man loved is
'that the man is loved'
- c. Der Mann wird geliebt.
the man is loved
'The man is loved.'
- d. dass der Mann der Frau hilft
that the man the.DAT woman helps
'that the man helps the woman'
- e. Der Mann hilft der Frau.
the man helps the.DAT woman
'The man is helping the woman.'

For the passive sentences, use the analysis where the subject noun phrase is moved from the object position, that is, the analysis without an empty expletive as the subject.

Further reading

For Sections 3.1–3.5, I used material from Peter Gallmann from 2003. This has been modified, however, at various points. I am solely responsible for any mistakes or inadequacies. For current materials, see <http://www.syntax-theorie.de>.

In the book *Syntaktische Analyseperspektiven*, Lohnstein (2014) presents a variant of GB which more or less corresponds to what is discussed in this chapter (CP/IP, movement-based analysis of the passive). The chapters in said book have been written by proponents of various theories and all analyze the same newspaper article. This book is extremely interesting for all those who wish to compare the various theories out there.

Haegeman (1990) is a comprehensive introduction to GB. Those who do read German may consider the textbooks by Fanselow & Felix (1987), von Stechow & Sternefeld (1988) and Grewendorf (1988) since they are also addressing the phenomena that are covered in this book.

Borsley (1999) and Kim & Sells (2008) have parallel textbooks for GB and HPSG in English. For the comparison of Transformational Grammar and LFG, see Bresnan & Kaplan (1982). Kuhn (2007) offers a comparison of modern derivational analyses with constraint-based LFG and HPSG approaches. Borsley (2012) contrasts analyses of long-distance dependencies in HPSG with movement-based analyses as in GB/Minimalism. Borsley discusses four types of data which are problematic for movement-based approaches: Extraction without fillers, extraction with multiple gaps, extractions where fillers and gaps do not match and extraction without gaps.

4 Transformational Grammar – Minimalism

Chomsky (1993; 1995) argued that the problem of language evolution should be taken seriously and that the question of how linguistic knowledge could become part of our genetic endowment should be answered. To that end he suggested refocussing the theoretical developments towards models that have to make minimal assumptions regarding the machinery that is needed for linguistic analyses and hence towards models that assume less language specific innate knowledge.

While innovations like \bar{X} Theory and the analysis of clause structure in GB are highly influential and can be found in most of the other theories that are discussed in this book, this is less so for the technical work done in the Minimalist framework. It is nevertheless useful to familiarize with the technicalities since Minimalism is a framework in which a lot of work is done and understanding the basic machinery makes it possible to read empirically interesting work in that framework.

The degree of formalization of Minimalist theories is not different from what is known from GB times (see Sections 3.6.2 and 4.7.2) and as a result there are only few computer-processable implementations. Edward Stabler and colleagues developed so-called *Minimalist Grammars*, which are formalizations of some of Chomsky's and Kayne's ideas (Stabler 2001; Kobele 2006; Gärtner & Michaelis 2007). On Minimalist Grammars see also Section 4.6.4 of this book. This formal work was also implemented by Stabler and others. In addition there are implementations by Sandiway Fong (Fong & Ginsburg 2012; Fong 2014) and Niyogi & Berwick (2005). However, these implementations cover only some of the aspects that are suggested in theoretical papers. They will be discussed in more detail in Section 4.7.2.

While the GB literature of the 1980s and 1990s shared a lot of assumptions, there was an explosion of various approaches in the Minimalist framework that is difficult to keep track of. The presentation that follows is based on David Adger's textbook (Adger 2003).

4.1 General remarks on the representational format

The theories that are developed in the framework of the Minimalist Program build on the work done in the GB framework. So a lot of things that were explained in the previous chapter can be taken over to this chapter. However, there have been some changes in fundamental assumptions. The general parametrized principles were dropped from the theory and instead the relevant distinctions live in features. Languages differ in the values that certain features may have and in addition to this, features may be strong or

weak and feature strength is also a property that may vary from language to language. Strong features make syntactic objects move to higher positions. The reader is familiar with this feature-driven movement already since it was a component of the movement-based analysis of the passive in Section 3.4. In the GB analysis of passive, the object had to move to the specifier position of IP in order to receive case. Such movements that are due to missing feature values are a key component in Minimalist proposals.

4.1.1 Basic architecture

Chomsky assumes that there are just two operations (rules) for combining linguistic objects: External Merge and Internal Merge. External Merge simply combines two elements like *the* and *book* and results in a complex phrase. Internal Merge is used to account for movement of constituents. It applies to one linguistic object and takes some part of this linguistic objects and adjoins it to the left of the respective object. The application of External Merge and Internal Merge can apply in any order. For instance, two objects can be combined with External Merge and then one of the combined items is moved to the left by applying Internal Merge. The resulting object can be externally merged with another object and so on. As an example consider the NP in (1):

- (1) the man who we know

To derive this NP the verb *know* is externally merged with its object *who*. After several intermediate merges that will be discussed below, *know who* will be merged with *we* and finally the *who* is moved to the left by Internal Merge, resulting in *who we know*. This relative clause can be externally merged with *man* and so on.

So, Minimalist theories differ from GB in not assuming a Deep Structure that is generated by some \bar{X} grammar and a Surface Structure that is derived from the Deep Structure by move α . Instead it is assumed that there is a phase in which External and Internal Merge (combination and movement) apply in any order to derive a certain structure that is then said to be spelled out. It is said that the structure is sent to the interfaces: The articulatory-perceptual system (AP) on the one hand and the conceptual-intentional system (CI) on the other side. AP corresponds to the level of Phonological Form (PF) and CI to the level of Logical Form (LF) in GB. The new architecture is depicted in Figure 4.1 on the facing page. Overt syntax stands for syntactic operations that usually have a visible effect. After overt syntax the syntactic object is sent off to the interfaces and some transformations may take place after this Spell-Out point. Since such transformations do not affect pronunciation, this part of syntax is called *covert syntax*. Like in GB's LF, the covert syntax can be used to derived certain scope readings.

This architecture was later modified to allow Spell-Out at several point in the derivation. It is now assumed that there are Phases in a derivation and that a completed Phase is spelled out once it is used in a combination with a head (Chomsky 2008). For instance, a subordinated sentence like *that Peter comes* in (2) is one Phase and is sent to the interfaces before the whole sentence is completed.

- (2) He believes that Peter comes.

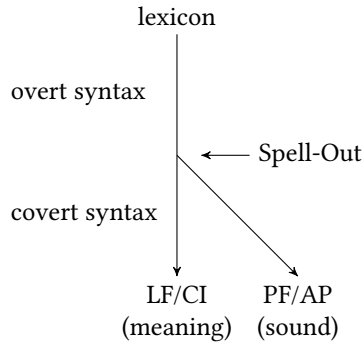


Figure 4.1: Architecture assumed in Minimalist theories before the Phase model

There are different proposals as to what categories form complete phases and I will ignore this concept in the following. See Section 15.1 on the psycholinguistic plausibility of Phases in particular and the Minimalist architecture in general. I will ignore Phases in the following.¹

4.1.2 Valence, feature checking, and agreement

The basic mechanism in Minimalist theories is feature checking. For instance, the noun *letters* may have a P feature, which means that it has to combine with a PP in order to form a complete phrase.

(3) letters to Peter

It is assumed that there are interpretable and uninterpretable features. An example of an interpretable feature is the number feature of nouns. The singular/plural distinction is semantically relevant. The category features for part of speech information are purely syntactic and hence cannot be interpreted semantically. Minimalism assumes that all uninterpretable features have to be used up during the derivation of a complex linguistic object. This process of eating up the feature is called *checking*. As an example, let us consider the noun *letters* again. The analysis is depicted in Figure 4.2 on the next page. The fact that the P feature of *letters* is uninterpretable is represented by the little *u* in front of the P. The uninterpretable P feature of *letters* can be checked against the P feature of *to Peter*. All checked features are said to delete automatically. The deletion is marked by striking the features out in the figures. Strings like (4) are ruled out as complete derivations since the N feature of P is not checked. This situation is shown in Figure 4.3 on the following page.

¹ Andreas Pankau (p. c. 2015) pointed out to me that there is a fundamental problem with such a conception of phases, since if it is the case that only elements that are in a relation to a head are sent off to the interface then the top-most phrase in a derivation would never be sent to the interfaces, since it does not depend on any head.

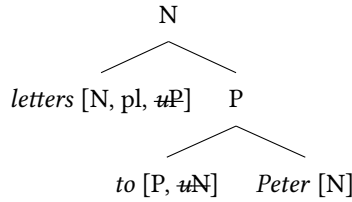


Figure 4.2: Valence representation via uninterpretable features

- (4) * letters to

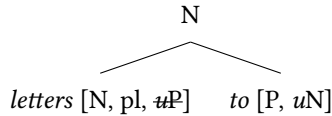


Figure 4.3: Illegitimate syntactic object due to an uninterpretable feature

If this structure would be used in a larger structure that is spelled out, the derivation would *crash* since the conceptual system could not make sense of the N feature that is still present at the P node.

Selectional features are atomic, that is, the preposition cannot select an NP[*acc*] as in GB and the other theories in this book unless NP[*acc*] is assumed to be atomic. Therefore an additional mechanism is assumed that can check other features in addition to selectional features. This mechanism is called *Agree*.

- (5) a. * letters to he
b. letters to him

The analysis of (6b) is shown in Figure 4.4. There is an interesting difference between the

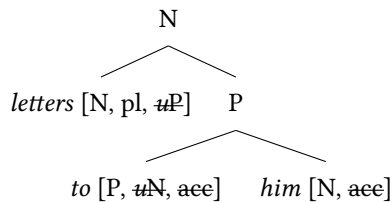


Figure 4.4: Feature checking via Agree

checking of selectional features and the checking of features via Agree. The features that

are checked via Agree do not have to be at the top node of the object that is combined with a head. This will play a role later in the analysis of the passive and local reordering.

4.1.3 Phrase structure and \bar{X} Theory

The projections of \bar{X} structures were given in Figure 2.9 on page 78. According to early versions of the \bar{X} Theory, there could be arbitrarily many complements that were combined with X^0 to form an \bar{X} . Arbitrarily many adjuncts could attach to \bar{X} and then at most one specifier could be combined with the \bar{X} yielding an XP. Minimalist theories assume binary branching and hence there is at most one complement, which is the first-merged item. Furthermore, it is not assumed that there is a unique specifier position. Rather Chomsky assumes that all items that are not complements are specifiers. That is he distinguishes between first-merged (complements) and later-merged items (specifiers). Figure 4.5 shows an example with two specifiers. It is also possible to have just

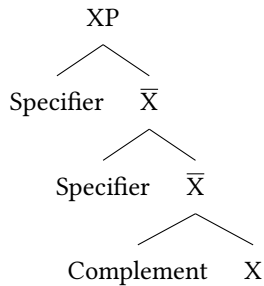


Figure 4.5: Complements and specifiers in Minimalist theories

a complement and no specifier or to have one or three specifiers. What structures are ultimately licensed depends on the features of the items that are involved in the Merge operations. Whether a phrasal projection counts as an \bar{X} or an XP depends on whether the phrase is used as a complement or specifier of another head or whether it is used as head in further Merge operations. If a phrase is used as specifier or complement its status is fixed to be a phrase (XP), otherwise the projectional status of resulting phrases is left underspecified. Lexical head daughters in Merge operations have the category X and complex head daughters in Merge operations have the category \bar{X} . This solves the problem that standard \bar{X} theoretic approaches had with pronouns and proper names: A lot of unary branching structure had to be assumed (See left picture in Figure 2.9). This is not necessary any longer in current Minimalist theories.²

² For problems with this approach see Brosziewski (2003: Chapter 2.1).

4.1.4 Little v

In Section 3.4 I used \bar{X} structures in which a ditransitive verb was combined with its accusative object to form a \bar{V} , which was then combined with the dative object to form a further \bar{V} . Such binary branching structures and also flat structures in which both objects are combined with the verb to form a \bar{V} are rejected by many practitioners of GB and Minimalism since the branching does not correspond to branchings that would be desired for phenomena like the binding of reflexives and negative polarity items. A binding in which *Benjamin* binds *himself* in (6a) is impossible:

- (6) a. *Emily showed himself Benjamin in the mirror.
 b. Peter showed himself Benjamin in the mirror.

What is required for the analysis of Binding and NPI phenomena in theories that analyze these phenomena in terms of tree configurations is that the reflexive pronoun is “higher” in the tree than the proper name *Benjamin*. More precisely, the reflexive pronoun *himself* has to c-command *Benjamin*. c-command is defined as follows (Adger 2003: 117):³

- (7) A node A c-commands B if, and only if A’s sister either:
 a. is B, or
 b. contains B

In the trees to the left and in the middle of Figure 4.6 the c-command relations are not as desired: In the left-most tree both NPs c-command each other and in the middle one *Benjamin* c-commands *himself* rather than the other way round. Hence it is assumed

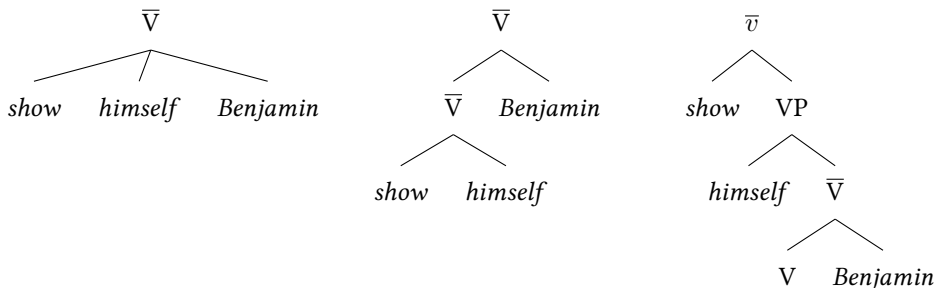


Figure 4.6: Three possible analyses of ditransitives

that the structures at the left and in the middle are inappropriate and that there is some additional structure involving the category v , which is called *little v* (Adger 2003: Section 4.4). The sister of *himself* is \bar{V} and \bar{V} contains *Benjamin*, hence *himself* c-commands

³ c-command also plays a prominent role in GB. In fact one part of Government & Binding is the Binding Theory, which was not discussed in the previous chapter since binding phenomena do not play a role in this book.

Benjamin. Since the sister of *Benjamin* is *V* and *V* neither is nor contains *himself*, *Benjamin* does not c-command *himself*.

The analysis of ditransitives involving little *v* goes back to Larson (1988). Larson assumes that the little *v* contributes a causative semantics. The structure in Figure 4.7 is derived by assuming that the verb *show* starts out in the *V* position and then moves to the *v* position. *show* is assumed to mean *see* and in the position of little *v* it picks up the causative meaning, which results in a *cause-see* meaning (Adger 2003: 133). The analysis is depicted in Figure 4.7.

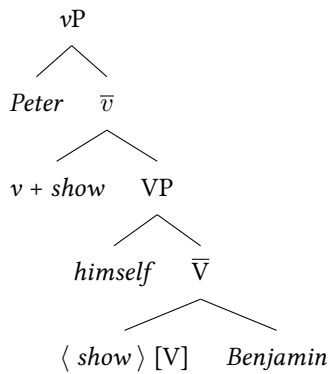


Figure 4.7: Analysis of ditransitives involving movement to little *v*

While this category was originally invented by Larson (1988) for the analysis of ditransitive verbs, it is now also used for the analysis of strictly transitive and even intransitive verbs.

Adger (2003: Section 4.5) argues that semantic roles are assigned uniformly in certain tree configurations:

- (8) a. NP daughter of *vP* → interpreted as agent
- b. NP daughter of *VP* → interpreted as theme
- c. PP daughter of *v̄* → interpreted as goal

Adger assumes that such uniformly assigned semantic roles help in the process of language acquisition and from this it follows that little *v* should also play a role in the analysis of examples with strictly transitive and intransitive verbs. The Figures 4.8 and 4.9 show the analysis of sentences containing the verbs *burn* and *laugh* respectively.⁴

⁴ If all intransitive verbs of this type are supposed to have agents as subjects, a very broad conception of agent has to be assumed that also subsumes the subject of verbs like *sleep*. Usually sleeping is not an activity that is performed intentionally.

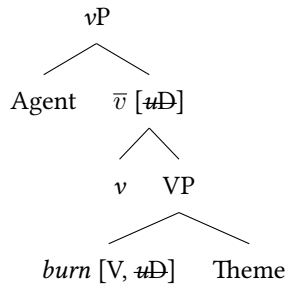


Figure 4.8: Analysis of strictly transitives involving little *v*

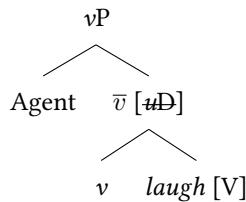


Figure 4.9: Analysis of intransitives involving little *v*

4.1.5 CP, TP, vP, VP

Section 3.1.8 dealt with the CP/IP system in GB. The Inflectional Phrase got renamed into Tense Phrase, but the core ideas of the CP/IP analysis have been transferred to the Minimalist analysis of English. This subsection will first discuss case assignment (Subsection 4.1.5.2) and then special features that are assumed to trigger movement (Subsection 4.1.5.1).

4.1.5.1 Features as triggers for movement: the EPP feature on T

The modals and auxiliaries were analyzed as members of the category I and the subjects as specifiers of IP. In the previous section, I showed how subjects are analyzed as specifiers of vP. Now, if one assumes that a modal verb combines with such a vP the subject follows the modal which does not correspond to the order that is observable in English. This problem is solved by assuming a strong uninterpretable D feature at T. Since the feature is strong, a suitable D has to move to the specifier of T and check the D locally. Figure 4.10 on the facing page shows the TP that plays a role in the analysis of (9):

- (9) Anna will read the book.

Full sentences are CPs. For the analysis of (9) an empty C head is assumed that is combined with the TP. The empty C contributes a clause type feature Decl. The full analysis

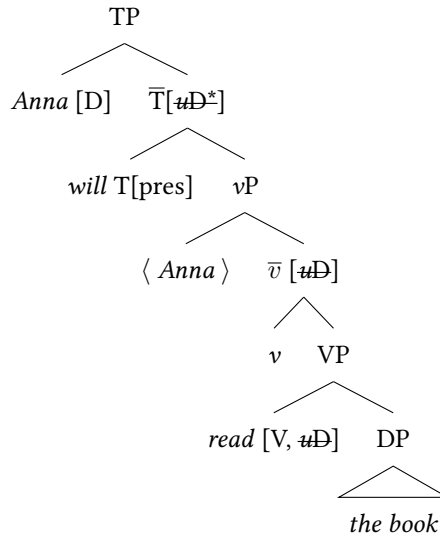


Figure 4.10: Analysis of *Anna will read the book*. involving a modal and movement of the subject from *v* to *T*

of (9) is shown in Figure 4.11.

The analysis of the question in (10) involves an unvalued clause-type feature on *T* for the sentence type question.

(10) What will Anna read?

The empty complementizer *C* has a *Q* feature that can value the clause-type feature on *T*. Since clause-type features on *T* that have the value *Q* are stipulated to be strong, the *T* element has to move to *C* to check the feature locally. In addition the *wh* element is moved. This movement is enforced by a strong *wh* feature on *C*. The analysis of (11) is given in Figure 4.12 on page 137.

4.1.5.2 Case assignment

In the GB analysis that was presented in Chapter 3, nominative was assigned by (finite) *I* and the other cases by the verb (see Section 3.4.2). The assignment of nominative is taken over to Minimalist analyses, so it is assumed that nominative is assigned by (finite) *T*. But in the Minimalist theory under consideration, there is not a single verb projection, but there are two verbal projections: *vP* and *VP*. Now, one could assume that *V* assigns accusative to its complement or that *v* assigns accusative to the complement of the verb it dominates. Adger (2003) argues for the latter approach. Figure 4.13 shows the *TP* for (11):

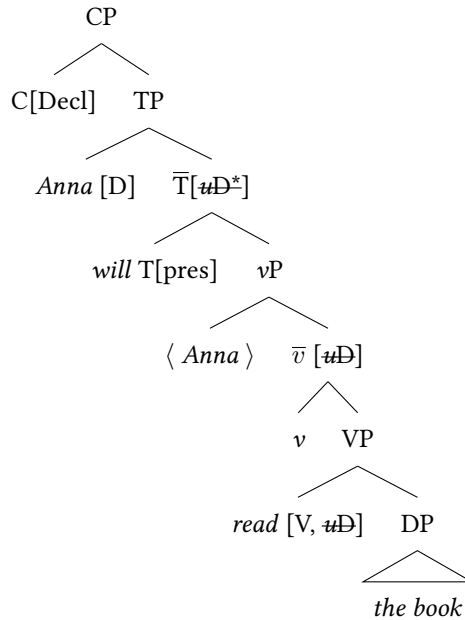


Figure 4.11: Analysis of *Anna will read the book.* as CP with an empty C with the clause-type feature Decl

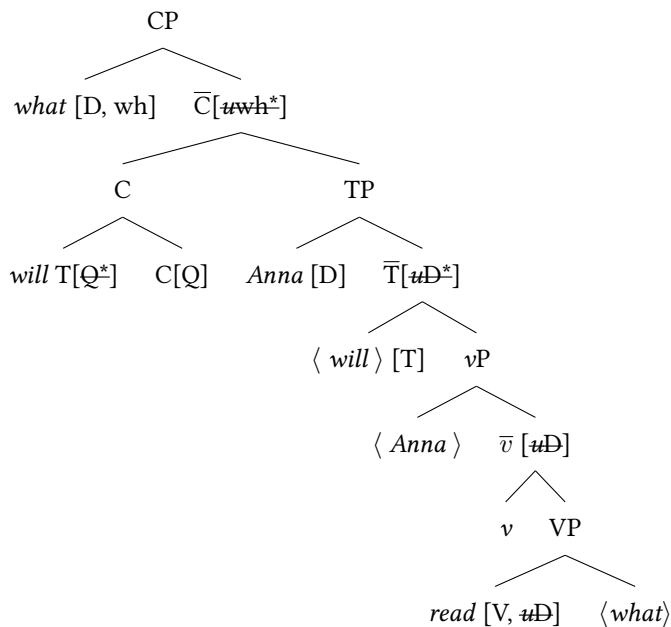
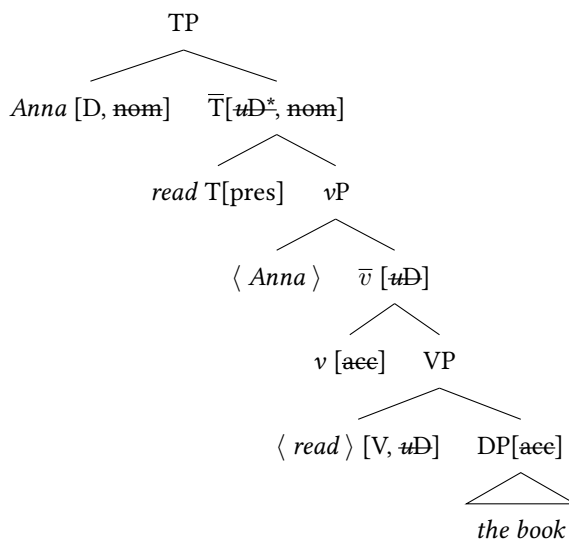
(11) Anna reads the book.

The two NPs *Anna* and *the book* start out with unvalued uninterpretable case features: [*ucase*:]. The features get valued by T and *v*. It is assumed that only one feature is checked by Merge, so this would be the D feature on T, leaving the case feature for the other available checking mechanism: Agree. Agree can be used to check features in sister nodes, but also features further away in the tree. The places that are possible candidates for Agree relations have to stand in a certain relation to each other. The first node has to c-command the node it Agrees with. c-command roughly means: one node up and then arbitrarily many nodes down. So *v* c-commands VP, V, the DP *the book*, and all the nodes within this DP. Since Agree can value features of c-commanded nodes, the accusative on *v* can value the case feature of the DP *the book*.

The non-locality that is built into Agree raises a problem: Why is it that (12) is ungrammatical?

(12) * Him likes she.

The accusative of *v* could be checked with its subject and the nominative of T with the object of *likes*. Both DP stand in the necessary c-command relations to T and *v*. This problem is solved by requiring that all Agree relations have to involve the closest possible element. Adger (2003: 218) formulates this constraint as follows:

Figure 4.12: Analysis of *What will Anna read?* with an empty C with a strong wh featureFigure 4.13: Case assignment by T and *v* in the TP for *Anna reads the book*.

- (13) Locality of matching: Agree holds between a feature F on X and a matching feature F on Y if and only if there is no intervening Z[F].

Intervention is defined as follows:

- (14) Intervention: In a structure [X ... Z ... Y], Z intervenes between X and Y iff X c-commands Y.

So, since T may Agree with *Anna* it must not Agree with *the book*. Hence nominative assignment to *she* in (14) is impossible and (14) is correctly ruled out.

4.1.6 Adjuncts

Adger (2003: Section 4.2.3) assumes that adjuncts attach to XP and form a new XP. He calls this operation *Adjoin*. Since this operation does not consume any features it is different from External Merge and hence the a new operation would be introduced into the theory contradicting Chomsky's claim that human languages use only Merge as a structure building operation. There are proposals to treat adjuncts as elements in special adverbial phrases with empty heads (see Section 4.6.1) that are also assumed to be part of a hierarchy of functional projections. Personally I prefer Adger's solution that corresponds to what is done in many other frameworks: There is a special rule or operation for the combination of adjuncts and heads (see for instance Section 9.1.7 on the HPSG schema for head adjunct combinations).

4.2 Verb position

The analysis of verb first sentences in German is straightforward, given the machinery that was introduced in the previous section. The basic idea is the same as in GB: The finite verb moves from V to *v* to T and then to C. The movement to T is forced by a strong tense feature on T and the movement of the T complex to C is enforced by a clause-type feature on T that is valued as a strong Decl by C. The analysis of (15) is shown in Figure 4.14 on the facing page.

- (15) Kennt jeder diesen Mann?
knows everybody this man
'Does everybody know this man?'

4.3 Long-distance dependencies

Having explained the placement of the verb in initial position, the analysis of V2 sentences does not come with a surprise: Adger (2003: 331) assumes a feature that triggers the movement of a constituent to a specifier position of C. Adger calls this feature *top*, but this is a misnomer since the initial position in German declarative sentences is not restricted to topics. Figure 4.15 on page 140 shows the analysis of (16):

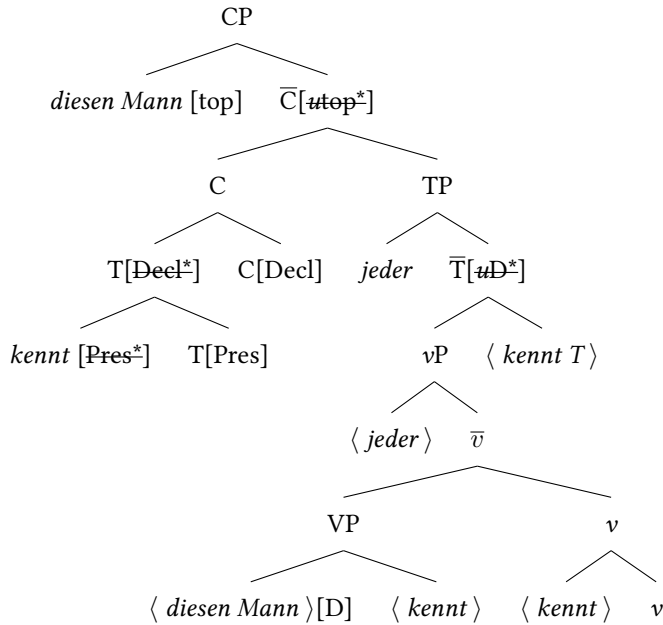


Figure 4.15: Analysis of *Diesen Mann kennt jeder*. ‘This man, everybody knows.’ following the analysis of Adger (2003: 331)

- b. der gestorbene Mann
the died man
‘the dead man’

The explanation of this difference is that adjectival participles predicate over what is the object in active sentences:

- (18) a. dass der Mann das Buch gelesen hat
that the man the book read has
‘that the man read the book’
b. das gelesene Buch
the read book

Now the assumption is that the argument of *gestorben* ‘died’ behaves like an object, while the argument of *getanzt* ‘danced’ behaves like a subject. If adjectival passives predicate over the object it is explained why (17b) is possible, while (17a) is not.

Adger (2003: 140) assumes the structure in Figure 4.16 on the facing page for vPs with unaccusative verbs. It is assumed that this unaccusative variant of little *v* plays a role in the analysis of the passive. Unaccusative verbs are similar to passivized verbs in that they do have a subject that somehow also has object properties. The special version of little

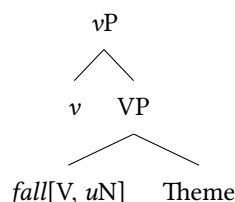


Figure 4.16: Structure of vP with unaccusative verbs like *fall*, *collapse*, *wilt* according to Adger (2003: 140)

v is selected by the Passive head *werden* ‘be’, which forms a Passive Phrase (abbreviated as PassP). See Figure 4.17 for the analysis of (19):

- (19) dass er geschlagen wurde
 that he beaten was
 ‘that he was beaten’

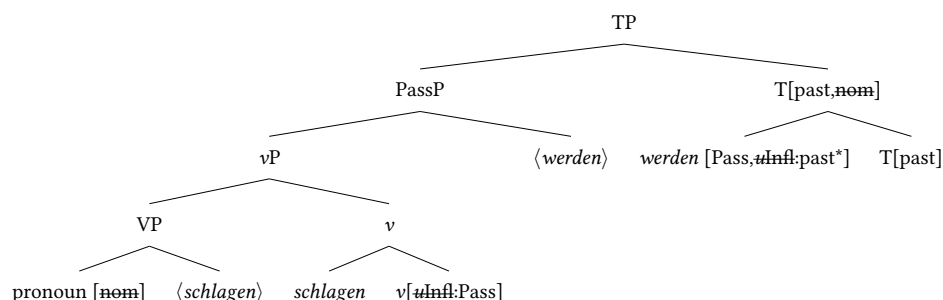


Figure 4.17: Minimalist analysis of the passive without movement but with non-local case assignment via Agree

The Pass head requires the Infl feature of little *v* to have the value Pass, which results in participle morphology at spellout. Hence the form that is used is *geschlagen* ‘beaten’. The auxiliary moves to T to check the strong Infl feature at T and since the Infl feature is past, the past form of *werden* ‘be’, namely *wurde* ‘was’, is used at spellout. T has a nom feature that has to be checked. Interestingly, the Minimalist approach does not require the object of *schlagen* to move to the specifier position of T in order to assign case, since case assignment is done via Agree. Hence the pronominal argument of *schlagen* in principle could stay in its object position and nevertheless get nominative from T. This would solve the problem of the GB analysis that was pointed out by Lenerz (1977: Section 4.4.3). See page 114 for Lenerz’ examples and discussion of the problem. However, Adger (2003: 332) assumes that German has a strong EPP feature on T. If this assumption

is upheld, all problems of the GB account will carry over to the Minimalist analysis: All objects have to move to T even when there is no reordering taking place. Furthermore impersonal passives of the kind in (20) would be problematic, since there is no noun phrase that could be moved to T in order to check the EPP feature:

- (20) weil getanzt wurde
 because danced was
 ‘because there was dancing there’

4.5 Local reordering

Adger (2003) does not treat local reordering. But there are several other suggestions in the literature. Since all reorderings in Minimalist theories are feature-driven, there must be an item that has a feature that triggers reorderings like those in (21b):

- (21) a. [weil] jeder diesen Mann kennt
 because everyone this man knows
 b. [weil] diesen Mann jeder kennt
 because this man everyone knows
 ‘because everyone knows this man’

There have been various suggestions involving functional projections like Topic Phrase (Laenzlinger 2004: 222) or AgrS and AgrO (Meinunger 2000: Chapter 4) that offer places to move to. G. Müller (2014a) offers a leaner solution though. In his approach the object simply moves to a second specifier position of little *v*. The analysis is depicted in Figure 4.18 on the facing page.

An option that was suggested by Laenzlinger (2004: 229–230) is to assume several Object Phrases for objects that may appear in any order. The objects move to the specifier positions of these projections and since the order of the Object Phrases is not restricted, both orders in (22) can be analyzed:

- (22) a. dass Hans diesen Brief meinem Onkel gibt
 that Hans this letter my uncle gives
 ‘that Hans gives this letter to my uncle’
 b. dass Hans meinem Onkel diesen Brief gibt
 that Hans my uncle this letter gives
 ‘that Hans gives to my uncle this letter’

4.6 New developments and theoretical variants

At the start of the 90s, Chomsky suggested a major rethink of the basic theoretical assumptions of GB and only keeping the parts of the theory which are absolutely necessary. In the *Minimalist Program*, Chomsky gives the central motivations for the far-reaching

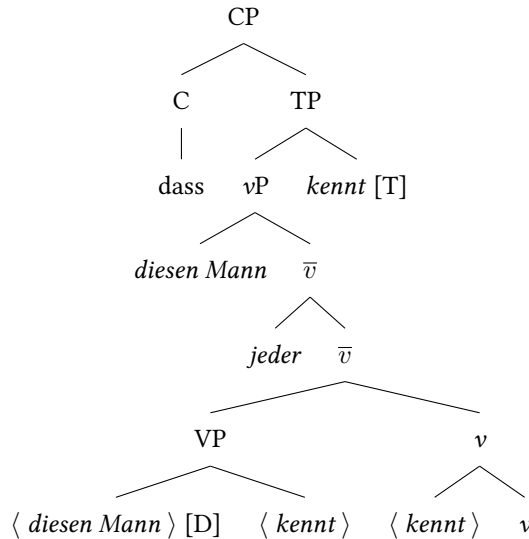


Figure 4.18: Analysis of *dass diesen Mann jeder kennt* ‘that everybody knows this man’ as movement of the object to a specifier position of *v*

revisions of GB theory (Chomsky 1993; 1995). Until the beginning of the 90s, it was assumed that Case Theory, the Theta Criterion, \bar{X} theory, Subjacency, Binding Theory, Control Theory etc. all belonged to the innate faculty for language (Richards 2015a: 804). This, of course, begs the question of how this very specific linguistic knowledge made its way into our genome. The Minimalist Program follows up on this point and attempts to explain properties of language through more general cognitive principles and reduce the number of innate language-specific knowledge postulated. The distinction between Deep Structure and Surface Structure, for example, was abandoned. Move still exists as an operation, but can be used directly to build sub-structures rather than after a complete D-Structure has been created. Languages differ with regard to whether this movement is visible or not.

Although Chomsky’s Minimalist Program should be viewed as a successor to GB, advocates of Minimalism often emphasize the fact that Minimalism is not a theory as such, but rather a research program (Chomsky: 2007: 4; 2013: 6). The suggestions made by Chomsky (1995) when introducing the research program have been reviewed by theoreticians and have sometimes come in for serious criticism (Kolb 1997; Johnson & Lappin 1997; 1999; Lappin, Levine & Johnson 2000a;b; 2001; Seuren 2004; Pinker & Jackendoff 2005), however, one should say that some criticisms overshoot the mark.

There are various strains of Minimalism. In the following sections, I will discuss some of its central ideas and criticisms.

4.6.1 Move, Merge, feature-driven movement and functional projections

Johnson, Lappin and Kolb have criticized the computational aspects of Chomsky's system. Chomsky suggested incorporating principles of economy into the theory. In certain cases, the grammatical system can create an arbitrary number of structures, but only the most economical, that is, the one which requires the least effort to produce, will be accepted as grammatical (transderivational economy). This assumption does not necessarily have to be taken too seriously and in reality, does not play a role in many works in the Minimalist framework (although see Richards (2015b) for recent approaches with derivations, which are compared in terms of economy). Nevertheless, there are other aspects of Chomsky's theory which can be found in many recent works. For example, Chomsky has proposed reducing the number of basic operations, which license structures. There are only the two operations Move and Merge (that is, External and Internal Merge). Move corresponds to the operation move α , which was already discussed above, and Merge is the combination of linguistics objects.

It is generally assumed that exactly two objects can be combined (Chomsky 1995: 226). For Move, it is the case that there must be a reason for a given movement operation. The reason for this movement is assumed to be that an element can check some feature in the position it is moved to. This idea was already presented in the analysis of the passive in Section 3.4: The accusative object does not bear case in passive sentences and therefore has to be moved to a position where it can receive case. This kind of argumentation is also used in newer analyses for a range of other phenomena. For example, it is assumed that there are phrases whose heads have the categories focus and topic. The corresponding functional heads are always empty in languages like German and English. Nevertheless, the assumption of these heads is motivated by the fact there are markers which signal the topic or focus of a sentence morphologically. This argumentation is only possible if one also assumes that the inventory of categories is the same for all languages. Then, the existence of a category in one language would suggest the existence of the same category in all other languages. This assumption of a shared universal component (Universal Grammar, UG) with detailed language-specific knowledge is, however, controversial and is shared by few linguists outside of the Chomskyan tradition. Even for those working in Chomskyan linguistics, there have been questions raised about whether it is permissible to argue in this way since – as Hauser, Chomsky & Fitch (2002) assume – if it is only the ability to create recursive structures which is specific to humans, then the individual syntactic categories are not part of UG and data from other languages cannot be used to motivate the assumption of invisible categories in another language.

The assumption that movement must be licensed by feature checking has led to an inflation of the number of (silent) functional heads.⁵ Rizzi (1997: 297) suggests the structure

⁵ The assumption of such heads is not necessary since features can be 'bundled' and then they can be checked together. For an approach in this vein, which is in essence similar to what theories such as HPSG assumes, see Sternefeld (2006: Section II.3.3.4, Section II.4.2).

In so-called cartographic approaches, it is assumed that every morphosyntactic feature corresponds to an independent syntactic head (Cinque & Rizzi 2010: 54, 61). For an explicitly formalized proposal, in which

in Figure 4.19 (see Grewendorf 2002: 85, 240; 2009, too). The functional categories Force,

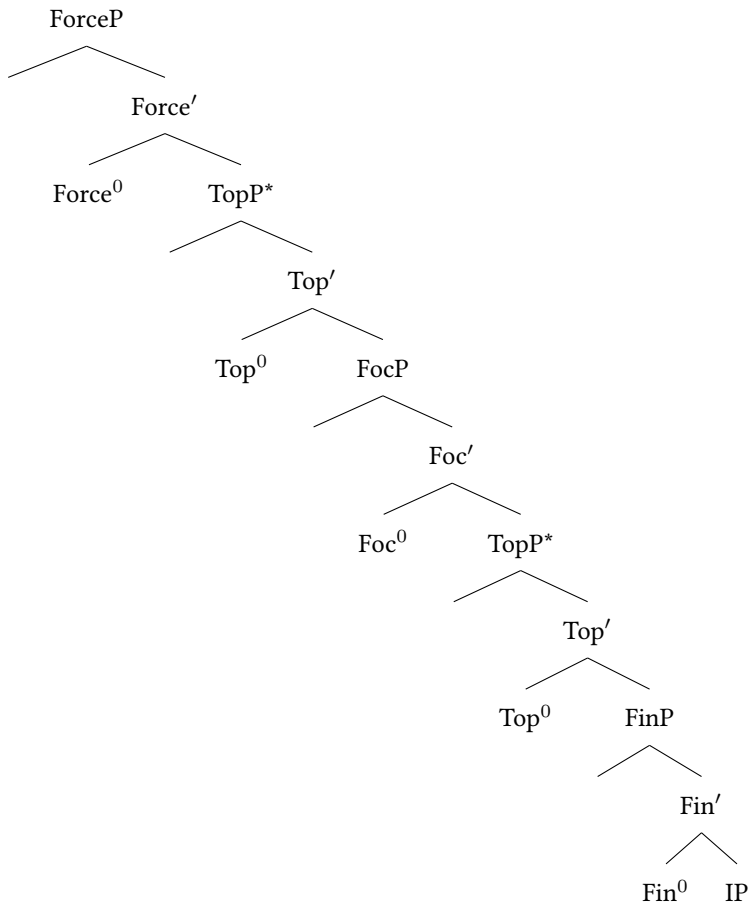


Figure 4.19: Syntactic structure of sentences following Rizzi (1997: 297)

Top, Foc and Fin correspond to clause type, topic, focus and finiteness. It is assumed that movement always targets a specifier position. Topics and focussed elements are always moved to the specifier position of the corresponding phrase. Topics can precede or follow focussed elements, which is why there are two topic projections: one above and one below FocP. Topic phrases are recursive, that is, an arbitrary number of TopPs can appear at the positions of TopP in the figure. Following Grewendorf (2002: 70), topic and focussed phrases are only realized if they are required for particular information

each feature is processed in its combination, see Stabler (2001: 335). Stabler's *Minimalist Grammars* are discussed in more detail in Section 4.6.4.

structural reasons such as movement.⁶ Chomsky (1995: 147) follows Pollock (1989) in assuming that all languages have functional projections for subject and object agreement as well as negation (AgrS, AgrO, Neg).⁷ Sternefeld (1995: 78), von Stechow (1996: 103) and Meinunger (2000: 100–101, 124) differentiate between two agreement positions for direct and indirect objects (AgrO, AgrIO). As well as AgrS, AgrO and Neg, Beghelli & Stowell (1997) assume the functional heads Share and Dist in order to explain scope phenomena in English as feature-driven movements at LF. For a treatment of scope phenomena without empty elements or movement, see Section 21.3. Błaszczak & Gärtner (2005: 13) assume the categories –PolP, +PolP and %PolP for their discussion of polarity.

Webelhuth (1995: 76) gives an overview of the functional projections that had been proposed up to 1995 and offers references for AgrA, AgrN, AgrV, Aux, Clitic Voices, Gender, Honorific, μ , Number, Person, Predicate, Tense, Z.

In addition to AdvP, NegP, AgrP, FinP, TopP and ForceP, Wiklund, Hrafnbjargarson, Bentzen & Hróarsdóttir (2007) postulate an OuterTopP.

Poletto (2000: 31) suggests both a HearerP and a SpeakerP for the position of clitics in Italian.

Cinque (1999: 106) adopts the 32 functional heads in Table 4.1 in his work. He assumes

Table 4.1: Functional heads following Cinque (1999: 106)

1. Mood _{Speech Act}	2. Mood _{Evaluative}	3. Mood _{Evidential}	4. Mood _{Epistemic}
5. T(Past)	6. T(Future)	7. Mood _{Irrealis}	8. Mod _{Necessity}
9. Mod _{Possibility}	10. Mod _{Volitional}	11. Mod _{Obligation}	12. Mod _{Ability/permission}
13. Asp _{Habitual}	14. Asp _{Repetitive(I)}	15. Asp _{Frequentative(I)}	16. Asp _{Celerative(I)}
17. T(Anterior)	18. Asp _{Terminative}	19. Asp _{Continuative}	20. Asp _{Perfect(?)}
21. Asp _{Retrospective}	22. Asp _{Proximative}	23. Asp _{Durative}	24. Asp _{Generic/progressive}
25. Asp _{Prospective}	26. Asp _{SgCompletive(I)}	27. Asp _{PlCompletive}	28. Asp _{Voice}
29. Asp _{Celerative(II)}	30. Asp _{SgCompletive(II)}	31. Asp _{Repetitive(II)}	32. Asp _{Frequentative(II)}

that all sentences contain a structure with all these functional heads. The specifier positions of these heads can be occupied by adverbs or remain empty. Cinque claims that these functional heads and the corresponding structures form part of Universal Grammar, that is, knowledge of these structures is innate (page 107).⁸ Laenzlinger (2004) follows Cinque in proposing this sequence of functional heads for German. He also follows

⁶ There are differing opinions as to whether functional projections are optional or not. Some authors assume that syntactic structures are always maximally present but functional heads can remain empty (e. g. Cinque (1999: 106) and Cinque & Rizzi (2010: 55)).

⁷ See Chomsky (1995: Section 4.10.1), however.

⁸ Table 4.1 shows only the functional heads in the clausal domain. Cinque (1994: 96, 99) also accounts for the order of adjectives with a cascade of projections: Quality, Size, Shape, Color, Nationality. These categories and their ordering are also assumed to belong to UG (p.100).

Cinque (1994: 96) claims that a maximum of seven attributive adjectives are possible and explains this with the fact that there are a limited number of functional projections in the nominal domain. As was shown on page 67, with a fitting context it is possible to use several adjectives of the same kind, which is why some of Cinque's functional projections would have to be subject to iteration.

Kayne (1994), who assumes that all syntactic structures have the sequence specifier head complement cross-linguistically, even if the surface order of the constituents seems to contradict this.

The realized structures are derived by leftward-movement⁹ Figure 4.20 on the next page shows the analysis of a verb-final clause where the functional adverbial heads have been omitted.¹⁰ Subjects and objects are generated as arguments inside of vP and VP, respectively. The subject is moved to the specifier of the subject phrase and the object is moved to the specifier of the object phrase. The verbal projection (VP_k) is moved in front of the auxiliary into the specifier position of the phrase containing the auxiliary. The only function of SubjP and ObjP is to provide a landing site for these movements. For a sentence, in which the object precedes the subject, Laenzlinger assumes that the object moves to the specifier of a topic phrase. Figure 4.20 contains only a ModP and an AspP, although Laenzlinger assumes that all the heads proposed by Cinque are present in the structure of all German clauses. For ditransitive verbs, Laenzlinger assumes multiple object phrases (page 230). A similar analysis with movement of object and subject from verb-initial VPs to Agr positions was suggested by Zwart (1994) for Dutch.

For general criticism of Kayne's model, see Haider (2000). Haider shows that a Kayne-like theory makes incorrect predictions and therefore fails to live up to its billing as a theory, which can explain all languages. Haider (1997a: Section 4) has shown that the assumption of an empty Neg head, as assumed by Pollock (1989), Haegeman (1995) and others, leads to problems. See Bobaljik (1999) for problems with the argumentation for Cinque's cascade of adverb-projections.

Furthermore, one should note that SubjP and ObjP, TraP (Transitive Phrase) and IntraP (Intransitive Phrase) (Karimi-Doostan 2005: 1745) and TopP (topic phrase), DistP (quantifier phrase), AspP (aspect phrase) (Kiss 2003: 22; Karimi 2005: 35), PathP and PlaceP (Svenonius 2004: 246) encode information about grammatical function, valency, information structure and semantics in the category symbols.¹¹ In a sense, this

⁹ This also counts for extraposition, that is, the movement of constituents into the postfield in German. Whereas this would normally be analyzed as rightward-movement, Kayne (1994: Chapter 9) analyzes it as movement of everything else to the left. Kayne assumes that (i.b) is derived from (i.a) by moving part of the NP:

- (i) a. just walked into the room [NP someone who we don't know].
- b. Someone_i just walked into the room [NP _i who we don't know].

(i.a) must have to be some kind of derived intermediate representation otherwise English would not be SV(O) underlyingly but rather V(O)S. (i.a) is therefore derived from (ii) by fronting the VP *just walked into the room*.

- (ii) Someone who we don't know just walked into the room

Such analyses have the downside that they cannot easily be combined with performance models (see Chapter 15).

¹⁰ These structures do not correspond to \bar{X} theory as it was presented in Section 2.5. In some cases, heads have been combined with the complements to form an XP rather than an X'. For more on \bar{X} theory in the Minimalist Program, see Section 4.6.3.

¹¹ For further examples and references, see Newmeyer 2004a: 194; 2005: 82. Newmeyer references also works which stipulate a projection for each semantic role.

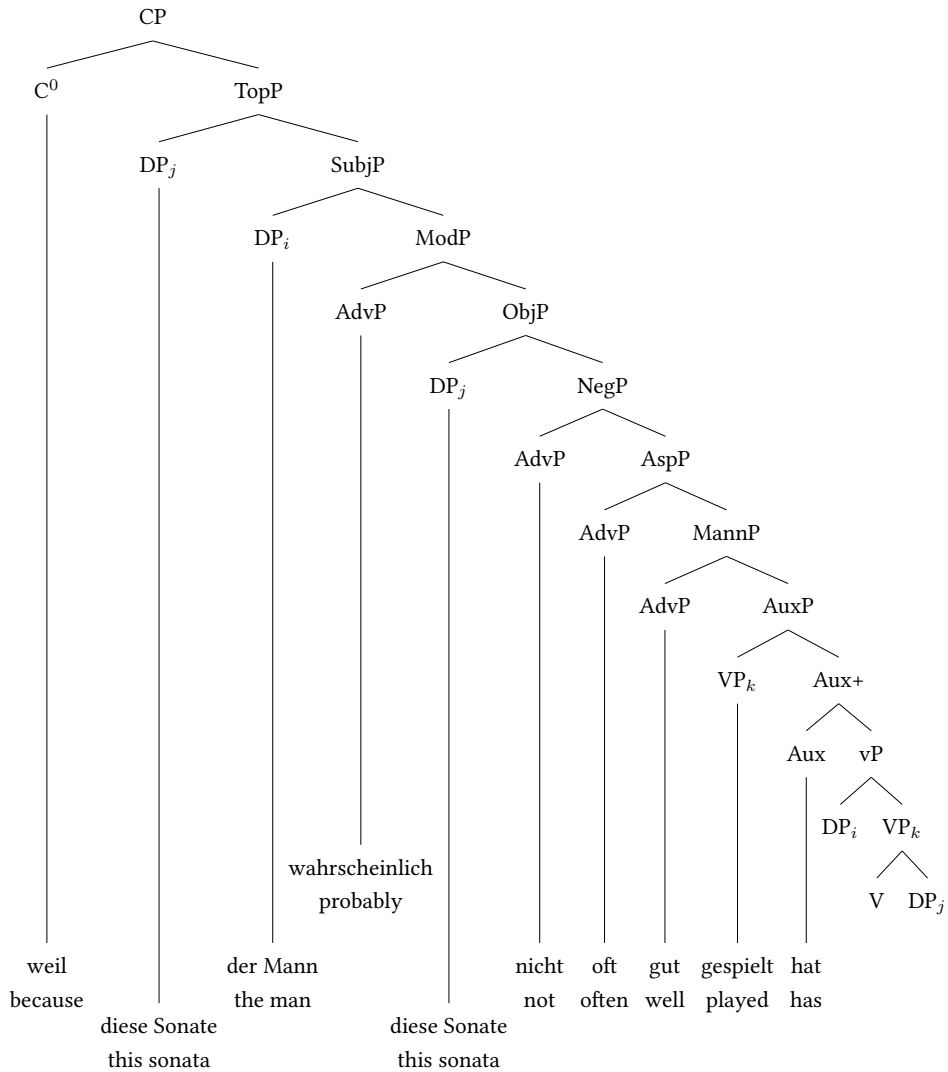


Figure 4.20: Analysis of sentence structure with leftward remnant movement and functional heads following Laenzlinger (2004: 224)

is a misuse of category symbols, but information structural and semantic categories are necessary since syntax, semantics, and information structure are tightly connected and since it is assumed that the semantics interprets the syntax directly and that semantics comes after syntax. By using semantically and pragmatically relevant categories, there is no longer a clean distinction between the levels of morphology, syntax, semantics and pragmatics: Everything has been ‘syntactified’. Felix Bildhauer has pointed out to me that approaches which assume a cascade of functional projections where the individual aspects of meaning are represented by nodes are actually very close to phrasal approaches in Construction Grammar (see Adger (2013: 470) also for a similar view). One simply lists configurations and these are assigned a meaning (or features which are interpreted post-syntactically. See Cinque & Rizzi (2010: 62) for the interpretation of TopP, for example).

If one takes the theory of feature checking in Specifier-Head relations to its logical conclusion, then one arrives at an analysis such as the one suggested by Radford (1997: 452). Radford assumes that prepositions occurring in the usual structure in (23) are also embedded in an Agreement Phrase and that the preposition adjoins to the head of the Agreement Phrase and the argument of the preposition is moved to the specifier.

(23) [PP P DP]

The problem here is that the object now precedes the preposition. In order to rectify this, Radford assumes a functional projection *p* (read *little p*) with an empty head to which the preposition then adjoins. This analysis is shown in Figure 4.21. This effort is

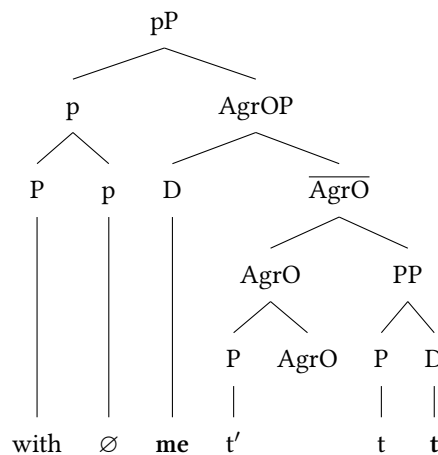


Figure 4.21: PP analysis following Radford with case assignment in specifier position in little *p*

only necessary in order to retain the assumption that feature checking takes place in specifier-head relations. If one were to allow the preposition to determine the case of its

object locally, then all this theoretical apparatus would not be necessary and it would be possible to retain the originally assumed structure.

Sternefeld (2006: 549–550) is critical of this analysis and compares it to Swiss cheese (being full of holes). The comparison to Swiss cheese is perhaps even too positive since, unlike Swiss cheese, the ratio of substance to holes in the analysis is extreme (2 words vs. 5 empty elements). We have already seen an analysis of noun phrases on page 72, where the structure of an NP, which only consisted of an adjective *klugen* ('clever'), contained more empty elements than overt ones. The difference to the PP analysis discussed here is that empty elements are only postulated in positions where overt determiners and nouns actually occur. The little *p* projection, on the other hand, is motivated entirely theory-internally. There is no theory-external motivation for any of the additional assumptions made for the analysis in Figure 4.21 (see Sternefeld (2006: 549–550)).

A variant of this analysis has been proposed by Hornstein, Nunes & Grohmann (2005: 124). The authors do without little *p*, which makes the structure less complex. They assume the structure in (24), which corresponds to the AgrOP-subtree in Figure 4.21.

$$(24) \quad [_{AgrP} DP_k [_{Agr'} P_i + Agr [_{PP} t_i t_k]]]$$

They assume that the movement of the DP to SpecAgrP happens invisibly, that is, covert. This solves Radford's problem and makes the assumption of *pP* redundant.

Furthermore, this analysis is argued for with reference to agreement phenomena in Hungarian. In Hungarian, there are postpositions, which agree with the preceding noun phrase in person and number. That is, the authors argue that English prepositional and Hungarian postpositional phrases have the same structure derived by movement, albeit the movement is covert in English.

In this way, it is possible to reduce the number and complexity of basic operations and, in this sense, the analysis is minimal. These structures are, however, still incredibly complex. No other kind of theory discussed in this book needs the amount of inflated structure to analyze the combination of a preposition with a noun phrase. The structure in (24) cannot be motivated by reference to data from English and it is therefore impossible to acquire it from the linguistic input. A theory, which assumes these kind of structures, would have to postulate a Universal Grammar with the information that features can only be checked in (certain) specifier positions (see Chapters 13 and 16 for more on Universal Grammar and language acquisition).

Another problem arises from the use of functional heads to encode linear order. In the classic CP/IP-system and all other theories discussed here, a category stands for a class of objects with the same distribution, that is, NP (or DP) stands for pronouns and complex noun phrases. Heads select phrases with a certain category. In the CP/IP-system, *I* selects a VP and an NP, whereas *C* selects an IP. In newer analyses, this kind of selectional mechanism does not work as easily. Since movement has taken place in (25b), we are dealing with a TopP or FocP in *das Buch dem Mann zu geben* 'the book the man to give'. Therefore, *um* cannot simply select a non-finite IP, but rather has to disjunctively be able to select a TopP, FocP or IP. It has to be ensured that TopPs and FocPs are marked with regard to the form on the verb contained inside them, since *um* can only be combined with *zu*-infinitives.

- (25) a. um dem Mann das Buch zu geben
 for the man the book to give
 ‘to give the man the book’
 b. um das Buch dem Mann zu geben
 for the book the man to give
 ‘to give the book to the man’

The category system, selectional mechanisms and projection of features would therefore have to be made considerably more complicated when compared to a system which simply base generates the orders or a system in which a constituent is moved out of the IP thereby creating a new IP.

Proposals that follow Cinque (1999) are problematic for similar reasons: Cinque assumes the category AdverbP for the combination of an adverb and a VP. There is an empty functional head, which takes the verbal projection as its complement and the adverb surfaces in the specifier of this projection. In these systems, the adverb phrase have to pass on inflectional properties of the verb since verbs with particular inflectional properties (finiteness, infinitives with *zu*, infinitives without *zu*, participles) have to be selected by higher heads (see page 181 and Section 9.1.4). For further, more serious problems with this analysis, see Haider (1997a: Section 5).

If one uses SpeakerP so that a clitic for first person singular can be moved to the correct specifier positions and a HearerP so that the clitic for second person can be moved to the correct position, then what one has are special projections which have all features that are relevant for clauses. In addition, the information that is needed to allow higher heads to select clauses containing clitics is included in the category labels. In other approaches and earlier variants of transformational grammar, selection was assumed to be strictly local so that higher heads only have access to those properties of embedded categories that are directly relevant for selection (Abraham (2005: 223); Sag (2007)). Locality will be discussed further in Chapter 19.

Finally, there is a conceptual problem with feature-driven movement, which has been pointed out by Gisbert Fanselow: Frey (2004a: 27) assumes a KontrP (contrastive phrase) and Frey (2004b) a TopP (topic phrase) (see Rizzi (1997) for TopP and FocP (focus phrase) in Italian and Haftka (1995), Grewendorf 2002: 85, 240; 2009, Abraham (2003: 19), Laenzlinger (2004: 224) and Hinterhölzel (2004: 18) for analyses of German with TopP and/or FocP). Constituents have to move to the specifier of these functional heads depending on their information structural status. Fanselow (2003a) has shown that such movement-based theories for the ordering of elements in the middle field are not compatible with current assumptions of the Minimalist Program. The reason for this is that some movement takes place in order to create space for other elements (altruistic movement). If the information structure of a sentence requires that the closest object to a verb is neither focussed nor not part of the focus, then the object closest to the verb should not receive the main stress in the clause. This can be achieved by deaccentuation, that is, by moving the accent to another constituent or even, as shown in (26b), by moving the object to a different position from the one in which it receives structural stress.

- (26) a. dass die Polizei gestern Linguisten verhaftete
 that the police yesterday linguists arrested
 ‘that the police arrested linguists yesterday’
 b. dass die Polizei Linguisten gestern verhaftete
 that the police linguists yesterday arrested
 ‘that the police arrested linguists YESTERDAY’

In Spanish, partial focus can be achieved not by special intonation, but rather only by altruistic movement in order to move the object out of the focus.

It is therefore not possible to assume that elements are moved to a particular position in the tree in order to check some feature motivated by information structural properties. Since feature checking is a prerequisite for movement in current minimalist theory, one would have to postulate a special feature, which only has the function of triggering altruistic movement. Fanselow (2003a: Section 4; 2006: 8) has also shown that the ordering constraints that one assumes for topic, focus and sentence adverbs can be adequately described by a theory which assumes firstly, that arguments are combined (in minimalist terminology: *merged*) with their head sequentially and secondly, that adjuncts can be adjoined to any projection level. The position of sentence adverbs directly before the focussed portion of the sentence receives a semantic explanation: Since sentence adverbs behave like focus-sensitive operators, they have to directly precede elements that they refer to. It follows from this that elements which do not belong to the focus of an utterance (topics) have to occur in front of the sentence adverb. It is therefore not necessary to assume a special topic position to explain local reorderings in the middle field. This analysis is also pursued in LFG and HPSG. These analyses are discussed in more detail in the corresponding chapters.

4.6.2 Labeling

In the Minimalist Program, Chomsky tries to keep combinatorial operations and mechanisms as simple as possible. He motivates this with the assumption that the existence of a UG with less language-specific knowledge is more plausible from an evolutionary point of view than a UG which contains a high degree of language-specific knowledge (Chomsky 2008: 135).

For this reason, he removes the projection levels of \bar{X} theory, traces, indices and similar elements of the grammar (page 138). All that remains is Merge and Move, that is, Internal and External Merge. Internal and External Merge combine two syntactic objects α and β into a larger syntactic object which is represented as a set $\{\alpha, \beta\}$. α and β can be either lexical items or internally complex syntactic objects. Internal Merge moves a part of an object to its periphery.¹² The result of internally merging an element is a set $\{\alpha, \beta\}$ where α was a part of β . External Merge also produces a set with two elements. However, two independent objects are merged. The objects that are created by Merge have a certain

¹² To be more specific, part of a syntactic object is copied and the copy is placed at the edge of the entire object. The original of this copy is no longer relevant for pronunciation (*Copy Theory of Movement*).

category (a set of features). For instance, if one combines the elements α and β , one gets $\{l, \{\alpha, \beta\}\}$, where l is the category of the resulting object. This category is also called a *label*. Since it is assumed that all constituents are headed, the category that is assigned to $\{\alpha, \beta\}$ has to be either the category of α or the category of β . Chomsky (2008: 145) discusses the following two rules for the determination of the label of a set.

- (27) a. In $\{H, \alpha\}$, H an LI, H is the label.
 b. If α is internally merged to β forming $\{\alpha, \beta\}$ then the label of β is the label of $\{\alpha, \beta\}$.

As Chomsky notes, these rules are not unproblematic since the label is not uniquely determined in all cases. An example is the combination of two lexical elements. If both H and α in (27a) are lexical items (LI), then both H and α can be the label of the resulting structure. Chomsky notices that this could result in deviant structures, but claims that this concern is unproblematic and ignores it. Chomsky offered a treatment of the combination of two lexical items in 2013. The solution to the problem is to assume that all combinations of lexical elements consist of a functional element and a root (Marantz 1997; Borer 2005). Roots are not considered as labels per definition¹³ and hence the category of the functional element determines the category of the combination (2013: 47). Such an analysis can only be rejected: The goal of the Minimalist Program is to simplify the theoretical proposals to such an extent that the models of language acquisition and language evolution become plausible, but in order to simplify basic concepts it is stipulated that a noun cannot simply be a noun but needs a functional element to tell the noun what category it has. Given that the whole point of Chomsky's Bare Phrase Structure was the elimination of the unary branching structures in \bar{X} Theory, it is unclear why they are reintroduced now through the backdoor, only more complex with an additional empty element.¹⁴ Theories like Categorical Grammar and HPSG can combine lexical items directly without assuming any auxiliary projections or empty elements. See Rauh (2013) also for a comparison of the treatment of syntactic categories in earlier versions of Transformational Grammar, HPSG, Construction Grammar, Role and Reference Grammar and root-based Neo-Constructivist proposals like the one assumed by Chomsky (2013). Rauh concludes that the direct connection of syntactic and semantic information is needed and that the Neo-Constructivism of Marantz and Borer has to be rejected. For further criticism of Neo-Constructivist approaches see Wechsler (2008a) and Müller & Wechsler (2014a: Sections 6.1 and 7).

The combination of a pronoun with a verbal projection poses a problem that is related to what has been said above. In the analysis of *He left* the pronoun *he* is a lexical element

¹³ Another category that is excluded as label per definition is *Conj*, which stands for conjunction (Chomsky 2013: 45–46). This is a stipulation that is needed to get coordination to work. See below.

¹⁴ The old \bar{X} rule in (i.a) corresponds to the binary combination in (i.b).

- (i) a. $N' \rightarrow N$
 b. $N \rightarrow N\text{-func root}$

In (i.a) a lexical noun is projected to an N' and in (i.b) a root is combined with a functional nominal head into a nominal category.

and hence would be responsible for the label of *He left*, since *left* is an internally complex verbal projection in Minimalist theories. The result would be a nominal label rather than a verbal one. To circumvent this problem, Chomsky (2013: 46) assumes that *he* has a complex internal structure: ‘perhaps D-pro’, that is, *he* is (perhaps) composed out of an invisible determiner and a pronoun.

The case in which two non-LIs are externally merged (for instance a nominal and a verbal phrase) is not discussed in Chomsky (2008). Chomsky (2013: 43–44) suggests that a phrase XP is irrelevant for the labelling of { XP, YP } if XP is moved (or rather copied in the Copy Theory of Movement) in a further step. Chomsky assumes that one of two phrases in a { XP, YP } combination has to move, since otherwise labelling would be impossible (p.12).¹⁵ The following coordination example will illustrate this: Chomsky assumes that the expression *Z and W* is analyzed as follows: First, Z and W are merged. This expression is combined with Conj (28a) and in the next step Z is raised (28b).

- (28) a. [α Conj [β Z W]]
 b. [γ Z [α Conj [β Z W]]

Since Z in β is only a copy, it does not count for labelling and β can get the label of W. It is stipulated for the combination of Z and α that Conj cannot be the label and hence the label of the complete structure is Z.¹⁶

A special case that is discussed by Chomsky is the Internal Merge of an LI α with a non LI β . According to rule (27a) the label would be α . According to (27b) the label would be β (see also Donati (2006)). Chomsky discusses the combination of the pronoun *what* with *you wrote* as an example.

- (29) what [C [you wrote *t*]]

If the label is determined according to (27b), one then has a syntactic object that would be called a CP in the GB framework; since this CP is, moreover, interrogative, it can function as the complement of *wonder* as in (30a). If the label is determined according to

¹⁵ His explanation is contradictory: on p.11 Chomsky assumes that a label of a combination of two entities with the same category is this category. But in his treatment of coordination, he assumes that one of the conjuncts has to be raised, since otherwise the complete structure could not be labeled.

¹⁶ As Bob Borsley (p.c. 2013) pointed out to me this makes wrong predictions for coordinations of two singular noun phrases with *and*, since the result of the coordination is a plural NP and not a singular one like the first conjunct. Theories like HPSG can capture this by grouping features in bundles that can be shared in coordinated structures (syntactic features and nonlocal features, see Pollard & Sag (1994: 202)). Furthermore the whole account cannot explain why (i.b) is ruled out.

- (i) a. both Kim and Lee
 b. * both Kim or Lee

The information about the conjunction has to be part of the representation for *or Lee* in order to be able to contrast it with *and Lee*.

A further problem is that the label of α should be the label of W since Conj does not count for label determination. This would lead to a situation in which we have to choose between Z and W to determine the label of γ . Following Chomsky’s logic either Z or W would have to move on to make it possible to label γ . Chomsky mentions this problem in footnote 40, but does not provide a solution.

(27a), one gets an object that can function as the accusative object of *read* in (30b), that is, something that corresponds to a DP in GB terminology.

- (30) a. I wonder what you wrote.
b. I read what you wrote.

what you wrote in (30b) is a so-called free relative clause.

Chomsky's approach to free relative clauses is interesting but is unable to describe the phenomenon in full breadth. The problem is that the phrase that contains the relative pronoun may be complex (contrary to Donati's claims, see also Citko (2008: 930–932)).¹⁷

(31) provides an English example from Bresnan & Grimshaw (1978: 333). German examples from Bausewein (1990: 155) and Müller (1999b: 78) are given in (32).

- (31) I'll read [whichever book] you give me.

- (32) a. Ihr könnt beginnen, [mit *wem*] ihr wollt.¹⁸
you can start with whom you want
'You can start with whoever you like.'
- b. [*Wessen Birne*] noch halbwegs in der Fassung steckt, pflegt solcherlei
whose nut yet halfway in the holder is uses such
Erloschene zu meiden;¹⁹
extinct to avoid
'Those who still have their wits half way about them tend to avoid such vacant characters;'
- c. [*Wessen Schuhe*] „danach“ besprenkelt sind, hat keinen Baum gefunden
whose shoes after.that speckled are has no tree found
und war nicht zu einem Bogen in der Lage.²⁰
and was not to a bow in the position
'If you end up with spattered shoes afterwards it is because you couldn't find a tree and you were incapable of peeing in an arc.'

Since *wessen Schuhe* 'whose shoes' is not a lexical item, rule (27b) has to be applied, provided no additional rules are assumed to deal with such cases. This means that the whole free relative clause *wessen Schuhe danach besprenkelt sind* is labeled as CP. For the free relatives in (31) and (32) the labeling as a CP is an unwanted result since they function as subjects or objects of the matrix predicates and hence should be labelled DP. However, since *wessen Schuhe* is a complex phrase and not a lexical item, (27a) does not apply and hence there is no analysis of the free relative clause as a DP. Therefore, it seems one must return to something like the GB analysis proposed by Groos & van

¹⁷ Chomsky (2013: 47) admits that there are many open questions as far as the labeling in free relative clauses is concerned and hence admits that there remain many open questions with labeling as such.

¹⁸ Bausewein (1990: 155).

¹⁹ Thomas Gsella, taz, 12.02.1997, p. 20.

²⁰ taz, taz mag, 08./09.08.1998, p. XII.

Riemsdijk (1981), at least for the German examples. Gross and van Riemsdijk assume that free relatives consist of an empty noun that is modified by the relative clause like a normal noun. On such an approach the complexity of the relative phrase is irrelevant. It is only the empty head that is relevant for labeling the whole phrase.²¹ However, once empty heads are countenanced in the analysis, the application of (27a) to (29) is undesirable since the application would result in two analyses for (30b): one with the empty nominal head and one in which (29) is labeled as NP directly. One might argue that in the case of several possible derivations, the most economical one wins, but the assumption of transderivational constraints leads to undesired consequences (Pullum 2013: Section 5).

Chomsky (2013) abandons the labeling condition in (27b) and replaces it with general labeling rules that hold for both internal and external Merge of two phrases. He distinguishes two cases. In the first case labeling becomes possible since one of the two phrases of the set { XP, YP } is moved away. This case was already discussed above. Chomsky writes about the other case: *X and Y are identical in a relevant respect, providing the same label, which can be taken as the label of the SO* (p. 11). He sketches an analysis of interrogative clauses on p. 13, in which the interrogative phrase has a Q feature and the remaining sentence from which the Q phrase was extracted has a Q feature as well. Since the two

²¹ Assuming an empty head is problematic since it may be used as an argument only in those cases in which an adjunct (the relative clause) is present (Müller 1999b: 97). See also Ott (2011: 187) for a later rediscovery of this problem. It can be solved in HPSG by assuming a unary projection that projects the appropriate category from a relative clause. I also use the unary projection to analyze so-called *non-matching* free relative clauses. In constructions with nonmatching free relative clauses the relative clause fills an argument slot that does not correspond to the properties of the relative phrase (Bausewein 1990). Bausewein discusses the following example, in which the relative phrase is a PP but the free relative fills the accusative slot of *kocht* ‘cooks’.

- (i) Sie kocht, worauf sie Appetit hat.
 she cooks where.on she appetite has
 ‘She cooks what she feels like eating.’

See Müller (1999b: 60–62) for corpus examples.

Minimalist theories do not employ unary projections. Ott (2011) develops an analysis in which the category of the relative phrase is projected, but he does not have a solution for nonmatching free relative clauses (p. 187). The same is true for Citko’s analysis, in which an internally merged XP can provide the label. Many other proposals for labeling or, rather, non-labeling exist. For instance some Minimalists want to eliminate labeling altogether and argue for a label-free syntax. As was pointed out by Osborne, Putnam & Groß (2011), such analyses bring Minimalism closer to Dependency Grammar. It is unclear how any of these models could deal with non-matching free relative clauses. Groß & Osborne (2009: Section 5.3.3) provide an analysis of free relatives in their version of Dependency Grammar, but deny the existence of nonmatching ones (p. 78). They suggest an analysis in which the relative phrase is the root/label of the free relative clause and hence they have the same problem as Minimalist proposals have with non-matching free relative clauses. As Groß & Osborne (2009: 73) and Osborne et al. (2011: 327) state: empty heads are usually not assumed in (their version of) Dependency Grammar. Neither are unary branching projections. This seems to make it impossible to state that free relative clauses with a relative phrase YP can function as XP provided XP is a category that is higher in the obliqueness hierarchy of Keenan & Comrie (1977), a generalization that was discovered by Bausewein (1990) (see also Müller (1999b: 60–62) and Vogel (2001: 4)). In order to be able to express the relevant facts, an element or a label has to exist that is different from the label of *worauf* in (i).

constituents share this property the label of the complete clause will be Q. This kind of labelling will ‘perhaps’ also be used for labeling normal sentences consisting of a subject and a verb phrase agreeing in person and number. These features would be responsible for the label of the sentence. The exact details are not worked out but almost certainly will be more complex than (27b).

A property that is inherent in both Chomsky (2005) and Chomsky (2013) is that the label is exclusively determined from one of the merged objects. As Bob Borsley pointed out to me, this is problematic for interrogative/relative phrases like (33).

(33) with whom

The phrase in (33) is both a prepositional phrase (because the first word is a preposition) and an interrogative/relative phrase (because the second word is an interrogative/relative word. So, what is needed for the correct labeling of PPs like the one in (33) is a well-defined way of percolating different properties from daughters to the mother node.²²

Summarizing, one can say that the labeling that was introduced to simplify the theory and reduce the amount of language specific innate knowledge that has to be assumed can only be made to work with a considerable amount of stipulations. For instance, the combination of lexical elements requires the assumption of empty functional heads, whose only purpose is determining the syntactic category of a certain lexical element. If this corresponded to linguistic reality, knowledge about labeling, the respective functional categories, and information about those categories that have to be ignored for the labeling would have to be part of innate language specific knowledge and nothing would be gained. One would be left with bizarre analyses with an enormous degree of complexity without having made progress in the Minimalist direction. Furthermore, there are empirical problems and a large number of unsolved cases.

The conclusion is that the label of a binary combination should not be determined in the ways suggested by Chomsky (2008; 2013). An alternative option for computing the label is to use the functor of a functor argument structure as the label (Berwick & Epstein 1995: 145). This is the approach taken by Categorical Grammar (Ajdukiewicz 1935; Steedman 2000) and in Stabler’s Minimalist Grammars (2011).²³ Stabler’s formalization

²² HPSG solves this problem by distinguishing head features including part of speech information and non-local features containing information about extraction and interrogative/relative elements. Head features are projected from the head, the nonlocal features of a mother node are the union of the nonlocal features of the daughters minus those that are bound off by certain heads or in certain configurations.

Citko (2008: 926) suggests an analysis in which both daughters can contribute to the mother node. The result is a complex label like $\{P, \{D, N\}\}$. This is a highly complex data structure and Citko does not provide any information on how the relevant information that it contains is accessed. Is an object with the label $\{P, \{D, N\}\}$ a P, a D or an N? One could say that P has priority since it is in the least embedded set, but D and N are in one set. What about conflicting features? How does a preposition that selects for a DP decide whether $\{D, N\}$ is a D or an N? In any case it is clear that a formalization will involve recursive relations that dig out elements of subsets in order to access their features. This adds to the overall complexity of the proposal and is clearly dispreferred over the HPSG solution, which provides one feature structure per linguistic object.

²³ For the Categorical Grammar approach to work, it is necessary to assign the category x/x to an adjunct, where x stands for the category of the head to which the adjunct attaches. For instance, an adjective

of Merge will be discussed in Section 4.6.4.

4.6.3 Specifiers, complements, and the remains of \bar{X} Theory

Chomsky (2008: 146) assumes that every head has exactly one complement but an arbitrary number of specifiers. In standard \bar{X} Theory, the restriction that there can be at most one complement followed from the general \bar{X} schema and the assumption that structures are at most binary branching: In standard \bar{X} Theory a lexical head was combined with all its complements to form an X' . If there are at most two daughters in a phrase, it follows that there can be only one complement (Sentences with ditransitive verbs have been analyzed with an empty head licensing an additional argument; see Larson (1988) and Müller & Wechsler (2014a: Sections 6.1 and 7) for a critical assessment of approaches involving little *v*). In standard \bar{X} Theory there was just one specifier. This restriction has now been abandoned. Chomsky writes that the distinction between specifier and complement can now be derived from the order in which elements are merged with their head: Elements that are *first-merged* are complements and all others – those which are *later-merged* – are specifiers.

Such an approach is problematic for sentences with mono-valent verbs: according to Chomsky's proposal subjects of mono-valent verbs would not be specifiers but complements.²⁴ This problem will be discussed in more detail in Section 4.6.4.

Apart from this, theories assuming that syntactic objects merged with word groups are specifiers do not allow for analyses in which two lexical verbs are directly coordinated as in (34):²⁵

combines with a nominal object to form a nominal object. Therefore its category is *n/n* rather than *adj*. Similarly, Stabler's approach does not extend to adjuncts unless he is willing to assign the category *noun* to attributive adjectives. One way out of this problem is to assume a special combination operation for adjuncts and their heads (see Frey & Gärtner (2002: Section 3.2)). Such a combination operation is equivalent to the Head Adjunct Schema of HPSG.

²⁴ Pauline Jacobson (p.c. 2013) pointed out that the problem with intransitive verbs could be solved by assuming that the last-merged element is the specifier and all non-last-merged elements are complements. This would solve the problems with intransitive verbs and with the coordination of verbs in (34) but it would not solve the problem of coordination in head-final languages as in (37). Furthermore, current Minimalist approaches make use of multiple specifiers and this would be incompatible with the Jacobsonian proposal unless one would be willing to state more complicated restrictions on the status of non-first-merged elements.

²⁵ Chomsky (2013: 46) suggests the coordination analysis in (28): according to this analysis, the verbs would be merged directly and one of the verbs would be moved around the conjunction in a later step of the derivation. As was mentioned in the previous section such analyses do not contribute to the goal of making minimal assumptions about innate language specific knowledge since it is absolutely unclear how such an analysis of coordination would be acquired by language learners. Hence, I will not consider this coordination analysis here.

Another innovation of Chomsky's 2013 paper is that he eliminates the concept of specifier. He writes in footnote 27 on page 43: *There is a large and instructive literature on problems with Specifiers, but if the reasoning here is correct, they do not exist and the problems are unformulable.* This is correct, but this also means that everything that was explained with reference to the notion of specifier in the Minimalist framework until now does not have an explanation any longer. If one follows Chomsky's suggestion, a large part of the linguistic research of the past years becomes worthless and has to be redone.

Chomsky did not commit himself to a particular view on linearization in his earlier work, but somehow

- (34) He [knows and loves] this record.

For example, in an analysis suggested by Steedman (1991: 264), *and* (being the head) is first merged with *loves* and then *knows*. The result of this combination is a complex object that has the same syntactic properties as the combined parts: the result is a complex verb that needs a subject and an object. After the combination of the conjunction with the two verbs, the result has to be combined with *this record* and *he*. *this record* behaves in all relevant respects like a complement. Following Chomsky's definition, however, it should be a specifier, since it is combined with the third application of Merge. The consequences are unclear. Chomsky assumes that Merge does not specify constituent order. According to him the linearization happens at the level of Phonological Form (PF). The restrictions that hold there are not described in his recent papers. However, if the categorization as complement or specifier plays a role for linearization as in Kayne's work (2011: 2, 12) and in Stabler's proposal (see Section 4.6.4), *this record* would have to be serialized before *knows and loves*, contrary to the facts. This means that a Categorical Grammar-like analysis of coordination is not viable and the only remaining option would seem to be assuming that *knows* is combined with an object and then two VPs are coordinated. Kayne (1994: 61, 67) follows Wexler & Culicover (1980: 303) in suggesting such an analysis and assumes that the object in the first VP is deleted. However, Borsley (2005: 471) showed that such an analysis makes wrong predictions, since (35a) would be derived from (35b) although these sentences differ in meaning.²⁶

- (35) a. Hobbs whistled and hummed the same tune.
b. Hobbs whistled the same tune and hummed the same tune.

Since semantic interpretation cannot see processes such as deletion that happen at the level of Phonological Form (Chomsky 1995: Chapter 3), the differences in meaning cannot be explained by an analysis that deletes material.

In a further variant of the VP coordination analysis, there is a trace that is related to *this record*. This would be a *Right-Node-Raising* analysis. Borsley (2005) has shown that such analyses are problematic. Among the problematic examples that he discusses is the following pair (see also Bresnan (1974: 615)).

- (36) a. He tried to persuade and convince him.
b. * He tried to persuade, but couldn't convince, him.

The second example is ungrammatical if *him* is not stressed. In contrast, (36a) is well-formed even with unstressed *him*. So, if (36a) were an instance of Right-Node-Raising, the contrast would be unexpected. Borsley therefore excludes a Right-Node-Raising analysis.

one has to ensure that the entities that were called specifier are realized in a position in which constituents are realized that used to be called specifier. This means that the following remarks will be relevant even under current Chomskian assumptions.

²⁶ See also Bartsch & Vennemann (1972: 102), Dowty (1979: 143), den Besten (1983: 104–105), Klein (1985: 8–9) and Eisenberg (1994b) for similar observations and criticism of similar proposals in earlier versions of Transformational Grammar.

The third possibility to analyze sentences like (34) assumes discontinuous constituents and uses material twice: the two VPs *knows this record* and *loves this record* are coordinated with the first VP being discontinuous. (See Crysmann (2001); Beavers & Sag (2004) for such proposals in the framework of HPSG.) However, discontinuous constituents are not usually assumed in the Minimalist framework (see for instance Kayne (1994: 67)). Furthermore, Abeillé (2006) showed that there is evidence for structures in which lexical elements are coordinated directly. This means that one needs analyses like the CG analysis discussed above, which would result in the problems with the specifier/complement status just discussed.

Furthermore, Abeillé has pointed out that NP coordinations in head-final languages like Korean and Japanese present difficulties for Merge-based analyses. (37) shows a Japanese example.

- (37) Robin-to Kim
Robin-and Kim
‘Kim and Robin’

In the first step *Robin* is merged with *to*. In a second step *Kim* is merged. Since *Kim* is a specifier, one would expect that *Kim* is serialized before the head as it is the case for other specifiers in head-final languages.

Chomsky tries to get rid of the unary branching structures of standard \bar{X} Theory, which were needed to project lexical items like pronouns and determiners into full phrases, referring to work by Muysken (1982b). Muysken used the binary features MIN and MAX to classify syntactic objects as minimal (words or word-like complex objects) or maximal (syntactic objects that stand for complete phrases). Such a feature system can be used to describe pronouns and determiners as [+MIN, +MAX]. Verbs like *give*, however, are classified as [+MIN, –MAX]. They have to project in order to reach the [+MAX]-level. If specifiers and complements are required to be [+MAX], then determiners and pronouns fulfill this requirement without having to project from X^0 via X' to the XP-level.

In Chomsky’s system the MIN/MAX distinction is captured with respect to the completeness of heads (complete = phrase) and to the property of being a lexical item. However, there is a small but important difference between Muysken’s and Chomsky’s proposal: the predictions with regard to the coordination data that was discussed above. Within the category system of \bar{X} Theory, it is possible to combine two X^0 s to get a new, complex X^0 . This new object has basically the same syntactic properties that simple X^0 s have (see Jackendoff (1977: 51) and Gazdar, Klein, Pullum & Sag (1985)). In Muysken’s system the coordination rule (or the lexical item for the conjunction) can be formulated such that the coordination of two +MIN items is a +MIN item. In Chomsky’s system an analogous rule cannot be defined, since the coordination of two lexical items is not a lexical item any longer.

Like Chomsky in his recent Minimalist work, Categorical Grammar (Ajdukiewicz 1935) and HPSG (Pollard and Sag: 1987; 1994: 39–40) do not (strictly) adhere to \bar{X} Theory. Both theories assign the symbol NP to pronouns (for CG see Steedman & Baldridge (2006: p. 615), see Steedman (2000: Section 4.4) for the incorporation of lexical type raising in

order to accommodate quantification). The phrase *likes Mary* and the word *sleeps* have the same category in Categorial Grammar ($s \backslash np$). In both theories it is not necessary to project a noun like *tree* from N^0 to \bar{N} in order to be able to combine it with a determiner or an adjunct. Determiners and mono-valent verbs in controlled infinitives are not projected from an X^0 level to the XP level in many HPSG analyses, since the valence properties of the respective linguistic objects (an empty SUBCAT or COMPS list) are sufficient to determine their combinatoric potential and hence their distribution (Müller 1996d; Müller 1999a). If the property of being minimal is needed for the description of a phenomenon, the binary feature LEX is used in HPSG (Pollard and Sag: 1987: 172; 1994: 22). However, this feature is not needed for the distinction between specifiers and complements. This distinction is governed by principles that map elements of an argument structure list (ARG-ST) onto valence lists that are the value of the SPECIFIER and the COMPLEMENTS feature (abbreviated as SPR and COMPS respectively).²⁷ Roughly speaking, the specifier in a verbal projection is the least oblique argument of the verb for configurational languages like English. Since the argument structure list is ordered according to the obliqueness hierarchy of Keenan & Comrie (1977), the first element of this list is the least oblique argument of a verb and this argument is mapped to the SPR list. The element in the SPR list is realized to the left of the verb in SVO languages like English. The elements in the COMPS list are realized to the right of their head. Approaches like the one by Ginzburg & Sag (2000: 34, 364) that assume that head-complement-phrases combine a word with its arguments have the same problem with coordinations like (34) since the head of the VP is not a word.²⁸ However, this restriction for the head can be replaced by one that refers to the LEX feature rather than to the property of being a word or lexical item.

Pollard and Sag as well as Sag and Ginzburg assume flat structures for English. Since one of the daughters is marked as lexical, it follows that the rule does not combine a head with a subset of its complements and then apply a second time to combine the result with further complements. Therefore a structure like (38a) is excluded, since *gave John* is not a word and hence cannot be used as the head daughter in the rule.

- (38) a. [[gave John] a book]
 b. [gave John a book]

Instead of (38a) only analyses like (38b) are admitted; that is, the head is combined with all its arguments all in one go. The alternative is to assume binary branching structures (Müller 2015c; Müller & Ørnsnes 2015: Section 1.2.2). In such an approach, the head complement schema does not restrict the word/phrase status of the head daughter. The binary branching structures in HPSG correspond to External Merge in the MP.

²⁷ Some authors assume a three-way distinction between subjects, specifiers, and complements.

²⁸ As mentioned above, a multidomination approach with discontinuous constituents is a possible solution for the analysis of (34) (see Crysmann (2001); Beavers & Sag (2004)). However, the coordination of lexical items has to be possible in principle as Abeillé (2006) has argued. Note also that the HPSG approach to coordination cannot be taken over to the MP. The reason is that the HPSG proposals involve special grammar rules for coordination and MP comes with the claim that there is only Merge. Hence the additional introduction of combinatorial rules is not an option within the MP.

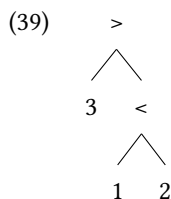
In the previous two sections certain shortcomings of Chomsky’s labeling definition and problems with the coordination of lexical items were discussed. In the following section I discuss Stabler’s definition of Merge in Minimalist Grammar, which is explicit about labeling and in one version does not have the problems discussed above. I will show that his formalization corresponds rather directly to HPSG representations.

4.6.4 Minimalism, Categorical Grammar, and HPSG

In this section, I will relate Minimalism, Categorical Grammar and HPSG to one another. Readers, who are not yet familiar with Categorical Grammar and HPSG, should skip this section and jump to the discussions in Chapters 6, 8 and 9 and return here after.

In Section 4.6.2 it was shown that Chomsky’s papers leave many crucial details about labeling unspecified. Stabler’s work is relatively close to recent Minimalist approaches, but is worked out with considerable precision (see also Stabler (2010: 397, 399, 400) on formalization of post GB approaches). Stabler (2001) shows how Kayne’s theory of remnant movement can be formalized and implemented. Stabler refers to his particular way of formalizing Minimalist theories as *Minimalist Grammars* (MG). There are a number of interesting results which follow from MG and variants thereof. In particular, with regard to the weak capacity of grammars (Michaelis 2001). It has been shown, for instance, that the number of possible languages one could create with MGs includes the set of those which can be created by Tree Adjoining Grammars (see Chapter 12). This means that it is possible to assign a greater number of word strings to structures with MGs, however, the structures derived by MGs are not necessarily always the same as the structures created by TAGs. For more on the generative capacity of grammars, see Chapter 18.

Although Stabler’s work can be regarded as a formalization of Chomsky’s Minimalist ideas, Stabler’s approach differs from Chomsky’s in certain matters of detail. Stabler assumes that the results of the two Merge operations are not sets but pairs. The head in a pair is marked by a pointer (‘<’ or ‘>’). Bracketed expressions like $\{ \alpha, \{ \alpha, \beta \} \}$ (discussed in Section 4.6.2) are replaced by trees like the one in (39).



1 is the head in (39), 2 is the complement and 3 the specifier. The pointer points to the part of the structure that contains the head. The daughters in a tree are ordered, that is, 3 is serialized before 1 and 1 before 2.

Stabler (2010: 402) defines External Merge as follows:

$$(40) \quad \text{em}(t_1[-f], t_2[f]) = \begin{cases} < \\ \begin{array}{cc} \diagup & \diagdown \\ t_1 & t_2 \end{array} & \text{if } t_1 \text{ has exactly 1 node} \\ > \\ \begin{array}{cc} \diagup & \diagdown \\ t_2 & t_1 \end{array} & \text{otherwise} \end{cases}$$

$=f$ is a selection feature and f the corresponding category. When $t_1[-f]$ and $t_2[f]$ are combined, the result is a tree in which the selection feature of t_1 and the respective category feature of t_2 are deleted. The upper tree in (40) represents the combination of a (lexical) head with its complement. t_1 is positioned before t_2 . The condition that t_1 has to have exactly one node corresponds to Chomsky's assumption that the first Merge is a Merge with a complement and that all further applications of Merge are Merges with specifiers (Chomsky 2008: 146).

A common assumption in current Minimalist theories is that movement is feature driven, that is, an item moves to a special position in a tree structure to check a feature. An example would be a *wh* element that is specified as $-wh$ and has to check this feature in a specifier position of a CP. In Stabler's system movement and feature checking is done in the definition of Internal Merge, which is given in (41).²⁹

$$(41) \quad \text{im}(t_1[+f]) = \begin{array}{c} > \\ \diagup \quad \diagdown \\ t_2^> \quad t_1\{t_2[-f]^> \mapsto \epsilon\} \end{array}$$

t_1 is a tree with a subtree t_2 which has the feature f with the value ' $-$ '. This subtree is deleted ($t_2[-f]^> \mapsto \epsilon$) and a copy of the deleted subtree without the $-f$ feature ($t_2^>$) is positioned in specifier position. The element in specifier position has to be a maximal projection. This requirement is visualized by the raised '>'.³⁰

Stabler provides an example derivation for the sentence in (42).

(42) *who Marie praises*

praises is a two-place verb with two $=D$ features. This encodes the selection of two determiner phrases. *who* and *Marie* are two Ds and they fill the object and subject position of the verb. The resulting verbal projection *Marie praises who* is embedded under an empty complementizer which is specified as $+wh$ and hence provides the position for the movement of *who*, which is placed in the specifier position of CP by the application of Internal Merge. The $-wh$ feature of *who* is deleted and the result of the application of Internal Merge is *who Marie praises*.

²⁹ In addition to what is shown in (41), Stabler's definition contains a variant of the *Shortest Move Constraint* (SMC), which is irrelevant for the discussion at hand and hence will be omitted.

This analysis has a problem that was pointed out by Stabler himself in unpublished work cited by Veenstra (1998: 124): It makes incorrect predictions in the case of mono-valent verbs. If a verb is combined with an NP, the definition of External Merge in (40) treats this NP as a complement³⁰ and serializes it to the right of the head. Instead of analyses of sentences like (43a) one gets analyses of strings like (43b).³¹

- (43) a. Max sleeps.
b. *Sleeps Max.

To solve this problem, Stabler assumes that mono-valent verbs are combined with a nonovert object (see Veenstra (1998: 61, 124) who, quoting Stabler’s unpublished work, also adopts this solution). With such an empty object, the resulting structure contains the empty object as a complement. The empty object is serialized to the right of the verb and *Max* is the specifier and hence serialized to the left of the verb as in (44)).

- (44) Max sleeps _.

Of course any analysis of this kind is both stipulative and entirely ad hoc, being motivated only by the wish to have uniform structures. Moreover, it exemplifies precisely one of the methodological deficiencies of Transformational Generative Grammar (the excessive appeal to uniformity) discussed at length by Culicover & Jackendoff (2005: Section 2.1.2).

An alternative is to assume an empty verbal head that takes *sleeps* as complement and *Max* as subject. Such an analysis is often assumed for ditransitive verbs in Minimalist theories which assume Larsonian verb shells (Larson 1988). Larsonian analyses usually assume that there is an empty verbal head that is called little *v* and that contributes a causative meaning. Adger (2003) adopts a little *v*-based analysis for intransitive verbs. Omitting the TP projection, his analysis is provided in Figure 4.22. Adger argues that

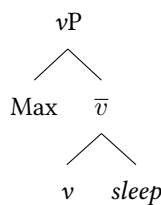


Figure 4.22: Little *v*-based analysis of *Max sleeps*

the analysis of sentences with unergative verbs involves a little *v* that selects an agent, while the analysis of unaccusative verbs involves a little *v* that does not select an N head.

³⁰ Compare also Chomsky’s definition of specifier and complement in Section 4.6.3.

³¹ More elaborated analyses assume that the subject of a verb has to move to the specifier position of a Tense Phrase (TP). These analyses would not license (43b), but they would fail to derive (43a) since *Max* would be treated as a complement rather than a specifier.

For unaccusatives he assumes that the verb selects a theme. He states that little *v* does not have necessarily a causative meaning but introduces the agent. But note that in the example at hand the subject of *sleep* is neither causing an event nor is it necessarily deliberately doing something. So it is rather an undergoer than an agent. This means that the assumption of the empty *v* head is made for purely theory-internal reasons without any semantic motivation in the case of intransitives. If the causative contribution of little *v* in ditransitive constructions is assumed, this would mean that one needs to little *vs*, one with and one without a causative meaning. In addition to the lack of theory-external motivation for little *v*, there are also empirical problems for such analyses. The reader is referred to Müller & Wechsler (2014a: Sections 6.1 and 7) for further details.

Apart from the two operations that were defined in (40) and (41), there are no other operations in MG.³² Apart from the problems with mono-valent verbs this results in the problem that was discussed in Section 4.6.3: There is no analysis with a direct combination of verbs for (34) – repeated here as (45).

(45) He [knows and loves] this record.

The reason is that the combination of *knows*, *and*, and *loves* consists of three nodes and the Merge of *knows and loves* with *this record* would make *this record* the specifier of the structure. Therefore *this record* would be serialized before *knows and loves*, contrary to the facts. Since the set of languages that can be generated with MGs contains the languages that can be generated with certain TAGs and with Combinatorial Categorical Grammar (Michaelis 2001), the existence of a Categorical Grammar analysis implies that the coordination examples can be derived in MGs somehow. But for linguists, the fact that it is possible to generate a certain string at all (the weak capacity of a grammar) is of less significance. It is the actual structures that are licensed by the grammar that are important (the strong capacity).

4.6.4.1 Directional Minimalist Grammars and Categorical Grammar

Apart from reintroducing X^0 categories, this problem can be solved by changing the definition of Merge in a way that allows heads to specify the direction of combination with their arguments: Stabler (2011: p. 635) suggests marking the position of an argument relative to its head together with the selection feature and gives the following redefinition of External Merge.

³² For extensions see Frey & Gärtner (2002: Section 3.2).

$$(46) \quad \text{em}(t_1[\alpha], t_2[x]) = \begin{cases} \begin{array}{c} < \\ \swarrow \quad \searrow \\ t_1 \quad t_2 \end{array} & \text{if } \alpha \text{ is } =x \\ \\ \begin{array}{c} > \\ \swarrow \quad \searrow \\ t_2 \quad t_1 \end{array} & \text{if } \alpha \text{ is } x= \end{cases}$$

The position of the equal sign specifies on which side of the head an argument has to be realized. This corresponds to Forward and Backward Application in Categorical Grammar. Stabler calls this form of grammar Directional MG (DMG). This variant of MG avoids the problem with mono-valent verbs and the coordination data is unproblematic as well if one assumes that the conjunction is a head with a variable category that selects for elements of the same category to the left and to the right of itself. *know* and *love* would both select an object to the right and a subject to the left and this requirement would be transferred to the result of coordinating *knows and loves*.³³ See Steedman (1991: 264) for the details of the CG analysis and Bouma & van Noord (1998: 52) for an earlier HPSG proposal involving directionality features along the lines suggested by Stabler for his DMGs.

4.6.4.2 Minimalist Grammars and Head-Driven Phrase Structure Grammar

The notation for marking the head of a structure with ‘>’ and ‘<’ corresponds directly to the HPSG representation of heads. Since HPSG is a sign-based theory, information about all relevant linguistic levels is represented in descriptions (phonology, morphology, syntax, semantics, information structure). (47) gives an example: the lexical entry for the word *grammar*.

$$(47) \quad \left[\begin{array}{c} \text{PHON} \\ \\ \text{SYNSEM|LOC} \\ \\ \text{word} \end{array} \left[\begin{array}{c} \langle 'gramər \rangle \\ \\ \left[\begin{array}{c} \text{CAT} \left[\begin{array}{c} \text{HEAD } \textit{noun} \\ \text{SPR} \quad \langle \text{DET} \rangle \\ \text{cat} \end{array} \right] \\ \\ \text{CONT} \dots \left[\begin{array}{c} \text{INST } X \\ \textit{grammar} \end{array} \right] \\ \\ \textit{loc} \end{array} \right] \right] \right]$$

The part of speech of *grammar* is *noun*. In order to form a complete phrase, it requires a determiner. This is represented by giving the SPR feature the value ⟨ DET ⟩. Semantic

³³ Note however, that this transfer makes it necessary to select complex categories, a fact that I overlooked in Müller (2013c). The selection of simplex features vs. complex categories will be discussed in Section 4.6.5.

information is listed under CONT. For details see Chapter 9.

Since we are dealing with syntactic aspects exclusively, only a subset of the used features is relevant: valency information and information about part of speech and certain morphosyntactic properties that are relevant for the external distribution of a phrase is represented in a feature description under the path SYNSEM|LOC|CAT. The features that are particularly interesting here are the so-called head features. Head features are shared between a lexical head and its maximal projection. The head features are located inside CAT and are grouped together under the path HEAD. Complex hierarchical structure is also modelled with feature value pairs. The constituents of a complex linguistic object are usually represented as parts of the representation of the complete object. For instance, there is a feature HEAD-DAUGHTER the value of which is a feature structure that models a linguistic object that contains the head of a phrase. The Head Feature Principle (48) refers to this daughter and ensures that the head features of the head daughter are identical with the head features of the mother node, that is, they are identical to the head features of the complete object.

$$(48) \text{ headed-phrase} \Rightarrow \left[\begin{array}{l} \text{SYNSEM|LOC|CAT|HEAD } \boxed{1} \\ \text{HEAD-DTR|SYNSEM|LOC|CAT|HEAD } \boxed{1} \end{array} \right]$$

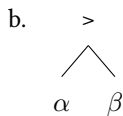
Identity is represented by boxes with the same number.

Ginzburg & Sag (2000: 30) represent all daughters of a linguistic object in a list that is given as the value of the DAUGHTERS attribute. The value of the feature HEAD-DAUGHTER is identified with one of the elements of the DAUGHTERS list:

$$(49) \quad \begin{array}{ll} \text{a.} & \left[\begin{array}{ll} \text{HEAD-DTR } \boxed{1} & \\ \text{DTRS} & \langle \boxed{1} \alpha, \beta \rangle \end{array} \right] \\ \text{b.} & \left[\begin{array}{ll} \text{HEAD-DTR } \boxed{1} & \\ \text{DTRS} & \langle \alpha, \boxed{1} \beta \rangle \end{array} \right] \end{array}$$

α and β are shorthands for descriptions of linguistic objects. The important point about the two descriptions in (49) is that the head daughter is identical to one of the two daughters, which is indicated by the $\boxed{1}$ in front of α and β , respectively. In the first feature description, the first daughter is the head and in the second description, the second daughter is the head. Because of the Head Feature Principle the syntactic properties of the whole phrase are determined by the head daughter. That is, the syntactic properties of the head daughter correspond to the label in Chomsky's definition. This notation corresponds exactly to the one that is used by Stabler: (49a) is equivalent to (50a) and (49b) is equivalent to (50b).

$$(50) \quad \text{a.} \quad \begin{array}{c} < \\ \swarrow \quad \searrow \\ \alpha \quad \beta \end{array}$$



An alternative structuring of this basic information, discussed by Pollard & Sag (1994: Chapter 9), eliminates the DAUGHTERS feature, using instead the two features HEAD-DAUGHTER and NON-HEAD-DAUGHTERS. This gives rise to feature descriptions like (51a), which corresponds directly to Chomsky’s set-based representations, discussed in Section 4.6.2 and repeated here as (51b).

- (51) a. $\left[\begin{array}{ll} \text{HEAD-DTR} & \alpha \\ \text{NON-HEAD-DTRS} & \langle \beta \rangle \end{array} \right]$
- b. $\{ \alpha, \{ \alpha, \beta \} \}$

The representation in (51a) does not contain information about linear precedence of α and β . Linear precedence of constituents is constrained by linear precedence rules, which are represented independently from constraints regarding (immediate) dominance.

The definition of Internal Merge in (41) corresponds to the Head-Filler-Schema in HPSG (Pollard & Sag 1994: 164). Stabler’s derivational rule deletes the subtree $t_2[-f]^>$. HPSG is monotonic, that is, nothing is deleted in structures that are licensed by a grammar. Instead of deleting t_2 inside of a larger structure, structures containing an empty element (NB – not a tree) are licensed directly.³⁴ Both in Stabler’s definition and in the HPSG schema, t_2 is realized as a filler in the structure. In Stabler’s definition of Internal Merge the category of the head daughter is not mentioned, but Pollard & Sag (1994: 164) restrict the head daughter to be a finite verbal projection. Chomsky (2007: 17) assumes that all operations but External Merge operate on Phase level. Chomsky assumes that CP and v^*P are Phases. If this constraint is incorporated into the definition in (41), the restrictions on the label of t_1 would have to be extended accordingly. In HPSG, sentences like (52) have been treated as VPs, not as CPs and hence Pollard and Sag’s requirement that the head daughter in the Head Filler Schema be verbal corresponds to Chomsky’s restriction.

- (52) Bagels, I like.

Hence, despite minor presentational differences, we may conclude that the formalization of Internal Merge and that of the Head-Filler Schema are very similar.

An important difference between HPSG and Stabler’s definition is that ‘movement’ is not feature driven in HPSG. This is an important advantage since feature-driven movement cannot deal with instances of so-called altruistic movement (Fanselow 2003a), that is, movement of a constituent that happens in order to make room for another constituent in a certain position. See also Bildhauer & Cook (2010: p. 72) for a discussion of such ‘altruistic’ multiple frontings in German.

³⁴ See Bouma, Malouf & Sag (2001a) for a traceless analysis of extraction in HPSG and Müller (2013b: Section 11.9) for a general discussion of empty elements.

A further difference between general \bar{X} Theory and Stabler's formalization of Internal Merge on the one hand and HPSG on the other is that in the latter case there is no restriction regarding the completeness (or valence 'saturation') of the filler daughter. Whether the filler daughter has to be a maximal projection (English) or not (German) follows from restrictions that are enforced locally when the trace is combined with its head. This makes it possible to analyze sentences like (53) without remnant movement.³⁵

- (53) Gelesen_i hat_j das Buch keiner _{-i} _{-j}.
 read has the book nobody

In contrast, Stabler is forced to assume an analysis like the one in (54b) (see also G. Müller (1998) for a remnant movement analysis). In a first step, *das Buch* is moved out of the VP (54a) and in a second step the emptied VP is fronted, as in (54b).

- (54) a. Hat [das Buch]_j [keiner [VP _{-j} gelesen]].
 b. [VP _{-j} Gelesen]_i hat [das Buch]_j [keiner _{-i}].

Haider (1993: 281), De Kuthy & Meurers (2001: Section 2) and Fanselow (2002) showed that this kind of remnant movement analysis is problematic for German. The only phenomenon that Fanselow identified as requiring a remnant movement analysis is the problem of multiple fronting (see Müller (2003a) for an extensive discussion of relevant data). Müller (2005b;c; 2015b) develops an alternative analysis of these multiple frontings which uses an empty verbal head in the *Vorfeld* (the position before the finite verb in root clauses), but does not assume that adjuncts or arguments like *das Buch* in (54b) are extracted from the *Vorfeld* constituent. Instead of the remnant movement analysis the mechanism of argument composition from Categorical Grammar (Geach 1970; Hinrichs & Nakazawa 1994) is used to ensure the proper realization of arguments in the sentence. Chomsky (2007: 20) already uses argument composition as part of his analysis of TPs and CPs. Hence both remnant movement and argument composition are already assumed in recent Minimalist proposals. The HPSG alternative, however, would appear to need less theoretical apparatus.

Finally, it should be mentioned that all transformational accounts have problems with Across the Board extraction like (55a) and (55) in which one element corresponds to several gaps.

- (55) a. Bagels, I like and Ellison hates.³⁶
 b. The man who_i [Mary loves _{-i}] and [Sally hates _{-i}] computed my tax.

This problem was solved for GPSG by Gazdar (1981b) and the solution carries over to HPSG. The Minimalist community tried to address these problems by introducing operations like sideward movement (Nunes 2004) where constituents can be inserted into sister trees. So in the example in (55a) *Bagels* is copied from the object position of *hates*

³⁵ See also Müller & Ørnsnes (2013b) for an analysis of object shift in Danish that can account for verb fronting without remnant movement. The analysis does not have any of the problems that remnant movement analyses have.

³⁶ Pollard & Sag (1994: 205).

into the object position of *like* and then these two copies are related to the fronted element. Koebele criticized such solutions since they overgenerate massively and need complicated filters. What he suggests instead is the introduction of a GPSG-style SLASH mechanism into Minimalist theories (Koebele 2008).

Furthermore, movement paradoxes (Bresnan 2001: Chapter 2) can be avoided by not sharing all information between filler and gap, a solution that is not available for transformational accounts, which usually assume identity of filler and gap or – as under the Copy Theory of Movement – assume that a derivation contains multiple copies of one object only one of which is spelled out. See also Borsley (2012) for further puzzles for, and problems of, movement-based approaches.

A further difference between MG and HPSG is that the Head-Filler-Schema is not the only schema for analysing long-distance dependencies. As was noted in footnote 9 on page 147, there is dislocation to the right (extraposition) as well as fronting. Although these should certainly be analyzed as long-distance dependencies, they differ from other long-distance dependencies in various respects (see Section 13.1.5). For analyses of extraposition in the HPSG framework, see Keller (1995); Bouma (1996), Müller (1999a: Chapter 13).

Apart from the schema for long-distance dependencies, there are of course other schemata in HPSG which are not present in MG or Minimalism. These are schemata which describe constructions without heads or are necessary to capture the distributional properties of parts of constructions, which cannot be easily captured in lexical analyses (e.g. the distribution of *wh*- and relative pronouns). See Section 23.10.

Chomsky (2010) has compared a Merge-based analysis of auxiliary movement to a HPSG analysis and critiqued that the HPSG analysis uses ten schemata rather than one (Merge). Ginzburg & Sag (2000) distinguish three types of construction with moved auxiliaries: inverted sentences such as those with fronted adverbial and with *wh*-questions (56a,b), inverted exclamatives (56c) and polar interrogatives (56d):

- (56) a. Under no circumstances *did she think they would do that*.
- b. Whose book *are you reading*?
- c. Am I tired!
- d. Did Kim leave?

Fillmore (1999) captures various different usage contexts in his Construction Grammar analysis of auxiliary movement and shows that there are semantic and pragmatic differences between the various contexts. Every theory must be able to account for these. Furthermore, one does not necessarily require ten schemata. It is possible to determine this – as Categorical Grammar does – in the lexical entry for the auxiliary or on an empty head (see Chapter 23 for a more general discussion of lexical and phrasal analyses). Regardless of this, every theory has to somehow account for these ten differences. If one wishes to argue that this has nothing to do with syntax, then this has to somehow be modelled in the semantic component. This means that there is no reason to prefer one theory over another at this point.

4.6.5 Selection of atomic features vs. selection of complex categories

Berwick & Epstein (1995) pointed out that Minimalist theories are very similar to Categorical Grammar and I have discussed the similarities between Minimalist theories and HPSG in Müller (2013c) and in the previous subsections. However, I overlooked one crucial difference between the usual assumptions about selection in Minimalist proposals on the one hand and Categorical Grammar, LFG, HPSG, TAG, and Construction Grammar on the other hand: What is selected in the former type of theory is a single feature, while the latter theories select for feature bundles. This seems to be a small difference but the consequences are rather severe. Stabler's definition of External Merge that was given on page 162 removes the selection feature (=f) and the corresponding feature of the selected element (f). In some publications the selection features are called uninterpretable features and are marked with an *u*. The uninterpretable features have to be checked and then they are removed from the linguistic object as in Stabler's definition. The fact that they have been checked is represented by crossed them out. It is said that all uninterpretable features have to be checked before a syntactic object is sent to the interfaces (semantics and pronunciation). If uninterpretable features are not checked that derivation crashes. Adger (2003: Section 3.6) explicitly discusses the consequences of these assumptions: A selecting head checks a feature of the selected object. It is not possible to check features of elements that are contained in the object that a head combines with. Only features at the top-most node, the so-called root node, can be checked with external merge. The only way features inside complex objects can be checked is by means of movement. This means that a head may not combine with a partially saturated linguistic object, that is, with a linguistic object that has an unchecked selection feature. I will to discuss this design decision with reference to an example provided by (Adger 2003: 95). The noun *letters* selects for a P and Ps select for an N. The analysis of (57a) corresponds to the left tree in Figure 4.23.

- (57) a. letters to Peter
b. * letters to

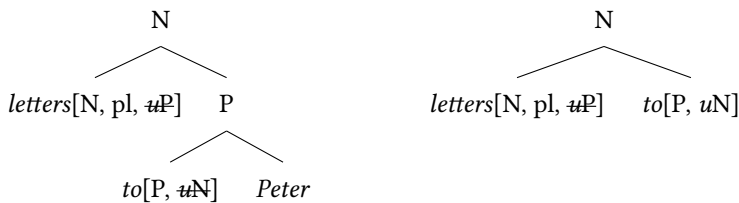


Figure 4.23: The analysis of *letters to Peter* according to Adger (2003: 95)

The string in (57b) is ruled out since the uninterpretable N feature of the preposition *to* is not checked. So this integrates the constraint that all dependent elements have to be maximal into the core mechanism. This makes it impossible to analyze examples like

(58) in the most straightforward way namely as involving a complex preposition and a noun that is lacking a determiner:

- (58) vom Bus
from.the bus

In theories in which complex descriptions can be used to describe dependants, the dependent may be partly saturated. So for instance in HPSG fused prepositions like *vom* ‘from.the’ can select an \bar{N} , which is a nominal projection lacking a specifier:

- (59) $N[SPR \langle DET \rangle]$

The description in (59) is an abbreviation for an internally structured set of feature-value pairs (see Section 9.6). The example here is given for the illustration of the differences only, since there may be ways of accounting for such cases in a single-feature-Merge system. For instance, one could assume a DP analysis and have the complex preposition select a complete NP (something of category N with no uninterpretable features). Alternatively, one can assume that there is indeed a full PP with all the structure that is usually assumed and the fusion of preposition and determiner happens during pronunciation. The first suggestion eliminates the option of assuming an NP analysis as it was suggested by Bruening (2009) in the Minimalist framework.

Apart from this illustrative example with a fused preposition, there are other cases in which one may want to combine unsaturated linguistic objects. I already discussed coordination examples above. Another example is verbal complexes in German, Dutch, Japanese, and Korean. Of course there are analyses of these languages that do not assume a verbal complex (G. Müller 1998; Wurmbrand 2003a), but these are not without problems. Some of the problems were discussed in the previous section as well.

Summing up this brief subsection, it has to be said that the feature checking mechanism that is built into the conception of Merge is more restrictive than the selection that is used in Categorical Grammar, Lexical Functional Grammar, HPSG, Construction Grammar, and TAG. In my opinion, it is too restrictive.

4.6.6 Summary

In sum, one can say that the computational mechanisms of the Minimalist Program (e. g. transderivational constraints and labeling) as well as the theory of feature-driven movement are problematic and the assumption of empty functional categories is sometimes ad hoc. If one does not wish to assume that these categories are shared by all languages, then proposing two mechanisms (Merge and Move) does not represent a simplification of grammar since every single functional category which must be stipulated constitutes a complication of the entire system.

The labeling mechanism is not yet fully developed and should be replaced by the head/functor-based labeling that is used in Categorical Grammar and HPSG.

4.7 Summary and classification

This section is similar to Section 3.6. I first comment on language acquisition and then on formalization.

4.7.1 Explaining language acquisition

Chomsky (2008: 135) counts theories in the MP as Principle and Parameter analyses and identifies MP parameters as being in the lexicon. Also, see Hornstein (2013: 396). UG is defined as possibly containing non-language-specific components, which are genetically determined (Chomsky 2007: 7). UG consists of unbounded Merge and the condition that expressions derived by a grammar must fulfill the restrictions imposed by the phonological and conceptual-intentional interfaces. In addition, a specific repertoire of features is assumed to be part of UG (Chomsky 2007: 6–7). The exact nature of these features has not been explained in detail and, as a result, the power of UG is somewhat vague. However, there is a fortunate convergence between various linguistic camps as Chomsky does not assume that the swathes of functional projections, which we encountered in Section 4.6.1, also form part of UG (However, authors like Cinque & Rizzi (2010) do assume that a hierarchy of functional projections is part of UG). Since there are still parameters, the same arguments used against GB approaches to language acquisition that were mentioned in Section 3.6.1 are still relevant for theories of language acquisition in the Minimalist Program. See Section 16 for an in-depth discussion approaches to language acquisition and the Principles and Parameters model as well as input-based approaches.

Chomsky's main goal in the Minimalist Program is to simplify the theoretical assumptions regarding formal properties of language and the computational mechanisms that are used so much as to make it plausible that they or relevant parts of them are part of our genetic endowment. But if we recapitulate what was assumed in this chapter, it is difficult to believe that Minimalist theories achieve this goal. To derive a simple sentence with an intransitive verb, one needs several empty heads and movements. Features can be strong or weak, Agree operates non-locally in trees across several phrase boundaries. And in order to make correct predictions, it has to be made sure that Agree can only see the closest possible element (13)–(14). This is a huge machinery in comparison to a Categorical Grammar that just combines adjacent things. Categorical Grammars can be acquired from input (see Section 13.8.3), while it is really hard to imagine how the fact that there are features that trigger movement when they are strong but do not trigger it when they are weak should be acquired from data alone.

4.7.2 Formalization

Section 3.6.2 commented on the lack of formalization in transformational grammar up until the 1990s. The general attitude towards formalization did not change in the minimalist era and hence there are very few formalizations and implementations of Minimalist theories.

Stabler (2001) shows how it is possible to formalize and implement Kayne's theory of remnant movement. In Stabler's implementation³⁷, there are no transderivational constraints, no numerations³⁸, he does not assume Agree (see Fong 2014: 132) etc. The following is also true of Stabler's implementation of Minimalist Grammars and GB systems: There are no large grammars. Stabler's grammars are small, meant as a proof of concept and purely syntactic. There is no morphology³⁹, no treatment of multiple agreement (Stabler 2011: Section 27.4.3) and above all no semantics. PF and LF processes are not modelled.⁴⁰ The grammars and computational system developed by Sandiway Fong are of similar size and faithfulness to the theory (Fong & Ginsburg 2012; Fong 2014): The grammar fragments are small, encode syntactic aspects such as labeling directly in the phrase structure (Fong & Ginsburg 2012: Section 4) and therefore fall behind \bar{X} theory. Furthermore, they do not contain any morphology. Spell-Out is not implemented so it is in the end not possible to process or create any utterances.⁴¹ The benchmark here has

³⁷ His system can be downloaded from his website: <http://www.linguistics.ucla.edu/people/stabler/coding.html>. 31.03.2010.

³⁸ There is a numeration lexicon in Veenstra (1998: Chapter 9). This lexicon consists of a set of numerations, which contain functional heads, which can be used in sentences of a certain kind. For example, Veenstra assumes numerations for sentences with bivalent verbs and subjects in initial position, for embedded sentences with mono-valent verbs, for *wh*-questions with mono-valent verbs and for polar interrogatives with mono-valent verbs. An element from this set of numerations corresponds to a particular configuration and a phrasal constructions in the spirit of Construction Grammar. Veenstra's analysis is not a formalization of the concept of the numeration that one finds in Minimalist works. Normally, it is assumed that a numeration contains all the lexical entries which are needed for the derivation of a sentence. As (i) shows, complex sentences can consist of various sentence types:

- (i) Der Mann, der behauptet hat, dass Maria gelacht hat, steht neben der Palme, die im letzten Jahr gepflanzt wurde.
 The man who claimed has that Maria laughed has stands next to the palm tree which in last year planted was
 'The man who claimed that Maria laughed is standing next to the palm tree that was planted last year.'

In (i), there are two relative clauses with verbs of differing valency, an embedded sentence with a mono-valent verb and the matrix clause. Under a traditional understanding of numerations, Veenstra would have to assume a infinitely large numeration lexicon containing all possible combinations of sentence types.

³⁹ The test sentences have the form as in (i).

- (i) a. the king will -s eat
 b. the king have -s eat -en
 c. the king be -s eat -ing
 d. the king -s will -s have been eat -ing the pie

⁴⁰ See Sauerland & Elbourne (2002) for suggestions of PF and LF-movement and the deletion of parts of copies (p. 285). The implementation of this would be far from trivial.

⁴¹ The claim by Berwick, Pietroski, Yankama & Chomsky (2011: 1221) in reference to Fong's work is just plain wrong: *But since we have sometimes adverted to computational considerations, as with the ability to "check" features of a head/label, this raises a legitimate concern about whether our framework is computationally realizable. So it is worth noting that the copy conception of movement, along with the locally oriented "search and labeling" procedure described above, can be implemented computationally as an efficient parser; see Fong, 2011, for details.* If one has a piece of software which cannot parse a single sentence, then one cannot claim that it is efficient since one does not know whether the missing parts of the program could make it

been set by implementations of grammars in constraint-based theories. For example, the HPSG grammars of German, English and Japanese that were developed in the 90s as part of *Verbmobil* (Wahlster 2000) for the analysis of spoken language or the LFG or CCG systems with large coverage. These grammars can analyze up to 83 % of utterances in spoken language (for *Verbmobil* from the domains of appointment scheduling and trip planning) or written language. Linguistic knowledge is used to generate and analyze linguistic structures. In one direction, one arrives at a semantic representation of a string of words and in the other one can create a string of words from a given semantic representation. A morphological analysis is indispensable for analyzing naturally occurring data from languages with elaborated morphological marking systems. In the remainder of this book, the grammars and computational systems developed in other theories will be discussed at the beginning of the respective chapters.

The reason for the lack of larger fragments inside of GB/MP could have to do with the fact that the basic assumptions of Minimalist community change relatively quickly:

In Minimalism, the triggering head is often called a *probe*, the moving element is called a *goal*, and there are various proposals about the relations among the features that trigger syntactic effects. Chomsky (1995: p. 229) begins with the assumption that features represent requirements which are checked and deleted when the requirement is met. The first assumption is modified almost immediately so that only a proper subset of the features, namely the ‘formal’, ‘uninterpretable’ features are deleted by checking operations in a successful derivation (Collins, 1997; Chomsky 1995: §4.5). Another idea is that certain features, in particular the features of certain functional categories, may be initially unvalued, becoming valued by entering into appropriate structural configurations with other elements (Chomsky 2008; Hiraiwa, 2005). And some recent work adopts the view that features are never deleted (Chomsky 2007: p. 11). These issues remain unsolved. (Stabler 2010: 397)

In order to fully develop a grammar fragment, one needs a least three years (compare the time span between the publication of *Barriers* (1986) and Stalber’s implementation (1992)). Particularly large grammars require the knowledge of several researchers working in international cooperation over the space of years or even centuries. This process is disrupted if fundamental assumptions are repeatedly changed at short intervals.

Further reading

This chapter heavily draws from Adger (2003). Other textbooks on Minimalism are Radford (1997), Grewendorf (2002), and Hornstein et al. (2005).

Kuhn (2007) offers a comparison of modern derivational analyses with constraint-based LFG and HPSG approaches. Borsley (2012) contrasts analyses of long-distance dependencies in HPSG with movement-based analyses as in GB/Minimalism. Borsley

extremely inefficient. Furthermore, one cannot compare the software to other programs. As has already been discussed, labeling is not carried out by Fong as was described in Chomsky’s work but instead he uses a phrase structure grammar of the kind described in Chapter 2.

discusses four types of data which are problematic for movement-based approaches: extraction without fillers, extraction with multiple gaps, extractions where fillers and gaps do not match and extraction without gaps.

The discussion of labeling, abandonment of \bar{X} theory and a comparison between Stabler's Minimalist Grammars and HPSG from Sections 4.6.2–4.6.4 can be found in Müller (2013c).

Intonational Phrasing, Discontinuity, and the Scope of Negation by Błaszczak & Gärtner (2005) is recommended for the more advanced reader. The authors compare analyses of negated quantifiers with wide scope in the framework of Minimalism (following Kayne) as well as Categorical Grammar (following Steedman).

In many of his publications, Chomsky discusses alternative, transformation-less approaches as “notational variants”. This is not appropriate as analyses without transformations can make different predictions to transformation-based approaches (e.g. w.r.t. coordination and extraction. See Section 5.5 for a discussion of GPSG in this respect). In Gazdar (1981a), one can find a comparison of GB and GPSG as well as a discussion of the classification of GPSG as a notational variant of Transformational Grammar with contributions from Noam Chomsky, Gerald Gazdar and Henry Thompson.

Sternefeld (2006) is a good, detailed introduction to syntax (839 pages) which develops a Transformational Grammar analysis of German which (modulo transformations) almost matches what is assumed in HPSG (featural descriptions for arguments ordered in a valence list according to a hierarchy). Sternefeld's structures are minimal since he does not assume any functional projections if they cannot be motivated for the language under discussion. Sternefeld is critical regarding certain aspects which some other analyses take for granted. Sternefeld views the book explicitly as a textbook from which one can learn how to argue coherently when creating theories. For this reason, this book is therefore not just recommended for students and PhD students.

Sternefeld & Richter (2012) discuss the situation in theoretical linguistics with particular focus on the theories described in this chapter. I can certainly understand the frustration of the authors with regard to the vagueness of analyses, argumentation style, empirical base of research, rhetorical clichés, immunization attempts and general respect for scientific standards: A current example of this is the article *Problems of Projection* by Chomsky (2013).⁴² I, however, do not share the general, pessimistic tone of this article.

⁴² Vagueness: In this article, *perhaps* occurs 19 times, *may* 17 as well as various *ifs*. Consistency: The assumptions made are inconsistent. See footnote 15 on page 154. Argumentation style: The term specifier is abolished and it is claimed that the problems associated with this term can no longer be formulated. They are therefore now not of this world. See footnote 25. Immunization: Chomsky writes the following regarding the Empty Category Principle: *apparent exceptions do not call for abandoning the generalization as far as it reaches, but for seeking deeper reasons to explain where and why it holds* p. 9. This claim is most certainly correct, but one wonders how much evidence one needs in a specific case in order to disregard a given analysis. In particular regarding the essay *Problems of Projection*, one has to wonder why this essay was even published only five years after *On Phases*. The evidence against the original approach is overwhelming and several points are taken up by Chomsky (2013) himself. If Chomsky were to apply his own standards (for a quote of his from 1957, see page 6) as well as general scientific methods (Ockham's Razor), the consequence would surely be a return to head-based analyses. For detailed comments on this essay, see Sections 4.6.2 and 4.6.3.

In my opinion, the patient's condition is critical, but he is not dead yet. As a reviewer of the Sternefeld and Richter paper pointed out, the situation in linguistics has changed so much that now having a dissertation from MIT does not necessary guarantee you a position (footnote 16) later on. One could view a reorientation of certain scientists with regard to certain empirical questions, adequate handling of data (Fanselow: 2004b; 2009: 137) and improved communication between theoretical camps as a way out of this crisis.

Since the 90s, it is possible to identify an increased empirical focus (especially in Germany), which manifests itself, for example, in the work of linguistic Collaborative Research Centers (SFBs) or the yearly *Linguistic Evidence* conference. As noted by the reviewer cited above, in the future it will not be enough to focus on Chomsky's problems in determining the syntactic category of sentences such as *He left* (see Section 4.6.2). Linguistic dissertations will have to have an empirical section, which shows that the author actually understands something about language. Furthermore, dissertations, and of course other publications, should give an indication that the author has not just considered theories from a particular framework but is also aware of the broad range of relevant descriptive and theoretical literature.

As I have shown in Section 4.6.4 and Müller (2013c) and will also show in the following chapters and concluding discussion chapter in particular, there are most certainly similarities between the various analyses on the market and they do converge in certain respects. The way of getting out the current crisis lies with the education and training of following generations, which should be empirically-grounded and theoretically broad.

Finally: Both teachers and students should read the prognosis by Sternefeld and Richter. I implore the students not abandon their studies straight after reading it, but rather to postpone this decision at least until after they have read the remaining chapters of this book.

5 Generalized Phrase Structure Grammar

Generalized Phrase Structure Grammar (GPSG) was developed as an answer to Transformational Grammar at the end of the 1970s. The book by Gazdar, Klein, Pullum & Sag (1985) is the main publication in this framework. Hans Uszkoreit has developed a large GPSG fragment for German (1987). Analyses in GPSG were so precise that it was possible to use them as the basis for computational implementations. The following is a possibly incomplete list of languages with implemented GPSG fragments:

- German (Weisweber 1987; Weisweber & Preuss 1992; Naumann 1987; 1988; Volk 1988)
- English (Evans 1985; Phillips & Thompson 1985; Phillips 1992; Grover, Carroll & Briscoe 1993)
- French (Emirkanian, Da Sylva & Bouchard 1996)
- Persian (Bahrani, Sameti & Manshadi 2011)

As was discussed in Section 3.1.1, Chomsky (1957) showed that simple phrase structure grammars are not well-suited to describing relations between linguistic structures and claimed that one needs transformations to explain them. These assumptions remained unchallenged for two decades (with the exception of publications by Harman (1963) and Freidin (1975)) until alternative theories such as LFG and GPSG emerged, which addressed Chomsky's criticisms and developed non-transformational explanations of phenomena for which there were previously only transformational analyses or simply none at all. The analysis of local movement of arguments, passives and long-distance dependencies are some of the most important phenomena that have been discussed in this framework. Following some introductory remarks on the representational format of GPSG in Section 5.1, I will present the GPSG analyses of these phenomena in some more detail.

5.1 General remarks on the representational format

This section has three parts. The general assumptions regarding features and the representation of complex categories is explained in Section 5.1.1, the assumptions regarding the linearization of daughters in a phrase structure rule is explained in Section 5.1.2. Section 5.1.3 introduces Mata rules, Section 5.1.4 deals with semantics, and Section 5.1.5 with adjuncts.

5.1.1 Complex categories, the Head Feature Convention, and \bar{X} rules

In Section 2.2, we augmented our phrase structure grammars with features. GPSG goes one step further and describes categories as sets of feature-value pairs. The category in (1a) can be represented as in (1b):

- (1) a. NP(3,sg,nom)
 b. { CAT n, BAR 2, PER 3, NUM sg, CASE nom }

It is clear that (1b) corresponds to (1a). (1a) differs from (1b) with regard to the fact that the information about part of speech and the \bar{X} projection (in the symbol NP) are prominent, whereas in (1b) these are treated just like the information about case, number or person.

Lexical entries have a feature SUBCAT. The value is a number which says something about the kind of grammatical rules in which the word can be used. (2) shows examples for grammatical rules and lists some verbs which can occur in these rules.¹

- (2) V2 → H[5] (kommen ‘come’, schlafen ‘sleep’)
 V2 → H[6], N2[CASE acc] (kennen ‘know’, suchen ‘search’)
 V2 → H[7], N2[CASE dat] (helfen ‘help’, vertrauen ‘trust’)
 V2 → H[8], N2[CASE dat], N2[CASE acc] (geben ‘give’, zeigen ‘show’)
 V2 → H[9], V3[+dass] (wissen ‘know’, glauben ‘believe’)

These rules license VPs, that is, the combination of a verb with its complements, but not with its subject. The numbers following the category symbols (V or N) indicate the \bar{X} projection level. For Uszkoreit, the maximum number of projections of a verbal projection is three rather than two as is often assumed.

The H on the right side of the rule stands for *Head*. The *Head Feature Convention* (HFC) ensures that certain features of the mother node are also present on the node marked with H (for details see Gazdar, Klein, Pullum & Sag 1985: Section 5.4 and Uszkoreit 1987: 67):

Principle 3 (Head Feature Convention)

The mother node and the head daughter must bear the same head features except if the features have an explicit value.

In (2), examples for verbs which can be used in the rules are given in brackets. As with ordinary phrase structure grammars, one also requires corresponding lexical entries for verbs in GPSG. Two examples are provided in (3):

- (3) V[5, vFORM inf] → einzuschlafen
 V[6, vFORM inf] → aufzuessen

The first rule states that *einzuschlafen* ‘to fall asleep’ has a SUBCAT value of 5 and the second indicates that *aufzuessen* ‘to finish eating’ has a SUBCAT value of 6. It follows, then, that *einzuschlafen* can only be used in the first rule (2) and *aufzuessen* can only be

¹ The analyses discussed in the following are taken from Uszkoreit (1987).

used in the second. Furthermore, (3) contains information about the form of the verb (*inf* stands for infinitives with *zu* ‘to’).

If we analyze the sentence in (4) with the second rule in (2) and the second rule in (3), then we arrive at the structure in Figure 5.1.

- (4) Karl hat versucht, [den Kuchen aufzuessen].
 Karl has tried the cake to.eat.up
 ‘Karl tried to finish eating the cake.’

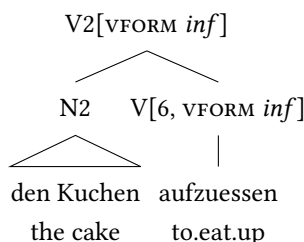


Figure 5.1: Projection of head features in GPSG

The rules in (2) say nothing about the order of the daughters which is why the verb ($H[6]$) can also be in final position. This aspect will be discussed in more detail in Section 5.1.2. With regard to the HFC, it is important to bear in mind that information about the infinitive verb form is also present on the mother node. Unlike simple phrase structure rules such as those discussed in Chapter 2, this follows automatically from the Head Feature Convention in GPSG. In (3), the value of *VFORM* is given and the HFC ensures that the corresponding information is represented on the mother node when the rules in (2) are applied. For the phrase in (4), we arrive at the category $V2[VFORM\ inf]$ and this ensures that this group of words only occurs in the contexts it is supposed to:

- (5) a. [Den Kuchen aufzuessen] hat er nicht gewagt.
 the cake to.eat has he not dared
 ‘He didn’t dare to eat the cake.’
 b. * [Den Kuchen aufzuessen] hat er nicht.
 the cake to.eat has he not
 c. * [Den Kuchen aufgeessen] hat er nicht gewagt.
 the cake eaten has he not dared
 d. [Den Kuchen aufgeessen] hat er nicht.
 the cake eaten has he not
 ‘He hasn’t eaten the cake.’

This works in an analogous way for noun phrases: There are rules for nouns which do not take an argument as well as for nouns with certain arguments. Examples of rules

for nouns, which either require no argument or two PPs are given in (6) (Gazdar, Klein, Pullum & Sag 1985: 127):

- (6) $N1 \rightarrow H[30]$ (*Haus* ‘house’, *Blume* ‘flower’)
 $N1 \rightarrow H[31], PP[mit], PP[über]$ (*Gespräch* ‘talk’, *Streit* ‘argument’)

The rule for the combination of \bar{N} and a determiner is as follows:

- (7) $N2 \rightarrow Det, H1$

$N2$ stands for another NP, that is, for a projection of a noun phrase on bar level one, whereas $H1$ stands for a projection of the head daughter on the bar level one. The Head Feature Convention ensures that the head daughter is also a nominal projection, since all features on the head daughter apart from the \bar{X} level are identified with those of the whole NP. When analysing (8), the second rule in (6) licenses the \bar{N} *Gesprächs mit Maria über Klaus*. The fact that *Gesprächs* (‘conversation’) is in the genitive is represented in the lexical entry of *Gesprächs* and since *Gesprächs* is the head, it is also present at \bar{N} following the Head Feature Convention.

- (8) des Gespräch-s mit Maria über Klaus
the.GEN conversation-GEN with Maria about Klaus
‘the conversation with Maria about Klaus’

For the combination of \bar{N} with the determiner, we apply the rule in (7). The category of the head determines the word class of the element on the left of the rule, which is why the rule in (7) corresponds to the classical \bar{X} rules that we encountered in (64c) on page 76. Since *Gesprächs mit Maria über Klaus* is again the head daughter, the information about the genitive of \bar{N} is also present at the NP node.

5.1.2 Local reordering

The first phenomenon we will discuss is local movement of arguments. As was already discussed in Section 3.5, arguments in the middle-field can occur in an almost arbitrary order. (9) gives some examples:

- (9) a. [weil] der Mann der Frau das Buch gibt
because the man.NOM the woman.DAT the book.ACC gives
b. [weil] der Mann das Buch der Frau gibt
because the man the book the woman gives
c. [weil] das Buch der Mann der Frau gibt
because the book the man the woman gives
d. [weil] das Buch der Frau der Mann gibt
because the book the woman the man gives
e. [weil] der Frau der Mann das Buch gibt
because the woman the man the book gives

- f. [weil] der Frau das Buch der Mann gibt
 because the woman the book the man gives
 ‘because the man gives the book to the woman’

In the phrase structure grammars in Chapter 2, we used features to ensure that verbs occur with the correct number of arguments. The following rule in (10) was used for the sentence in (9a):

- (10) $S \rightarrow NP[nom] NP[dat] NP[acc] V_nom_dat_akk$

If one wishes to analyze the other orders in (9), then one requires an additional five rules, that is, six in total:

- (11) $S \rightarrow NP[nom] NP[dat] NP[acc] V_nom_dat_akk$
 $S \rightarrow NP[nom] NP[acc] NP[dat] V_nom_dat_akk$
 $S \rightarrow NP[acc] NP[nom] NP[dat] V_nom_dat_akk$
 $S \rightarrow NP[acc] NP[dat] NP[nom] V_nom_dat_akk$
 $S \rightarrow NP[dat] NP[nom] NP[acc] V_nom_dat_akk$
 $S \rightarrow NP[dat] NP[acc] NP[nom] V_nom_dat_akk$

In addition, it is necessary to postulate another six rules for the orders with verb initial order:

- (12) $S \rightarrow V_nom_dat_akk NP[nom] NP[dat] NP[acc]$
 $S \rightarrow V_nom_dat_akk NP[nom] NP[acc] NP[dat]$
 $S \rightarrow V_nom_dat_akk NP[acc] NP[nom] NP[dat]$
 $S \rightarrow V_nom_dat_akk NP[acc] NP[dat] NP[nom]$
 $S \rightarrow V_nom_dat_akk NP[dat] NP[nom] NP[acc]$
 $S \rightarrow V_nom_dat_akk NP[dat] NP[acc] NP[nom]$

Furthermore, one would also need parallel rules for transitive and intransitive verbs with all possible valencies. Obviously, the commonalities of these rules and the generalizations regarding them are not captured. The point is that we have the same number of arguments, they can be realized in any order and the verb can be placed in initial or final position. As linguists, we find it desirable to capture this property of the German language and represent it beyond phrase structure rules. In Transformational Grammar, the relationship between the orders is captured by means of movement: The Deep Structure corresponds to verb-final order with a certain order of arguments and the surface order is derived by means of move α . Since GPSG is a non-transformational theory, this kind of explanation is not possible. Instead, GPSG imposes restrictions on *immediate dominance* (ID), which differ from those which refer to *linear precedence* (LP): Rules such as (13) are to be understood as dominance rules, which do not have anything to say about the order of the daughters (Pullum 1982).

- (13) $S \rightarrow V, NP[nom], NP[acc], NP[dat]$

The rule in (13) simply states that S dominates all other nodes. Due to the abandonment of ordering restrictions for the right-hand side of the rule, we only need one rule rather than twelve.

Nevertheless, without any kind of restrictions on the right-hand side of the rule, there would be far too much freedom. For example, the following order would be permissible:

- (14) * Der Frau der Mann gibt ein Buch.
 the woman.DAT the man.NOM gives the book.ACC

Such orders are ruled out by so-called *Linear Precedence Rules* or LP-rules). LP-constraints are restrictions on local trees, that is, trees with a depth of one. It is, for example, possible to state something about the order of V, NP[nom], NP[acc] and NP[dat] in Figure 5.2 using linearization rules.

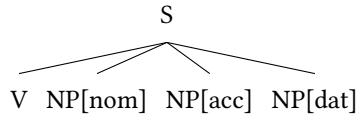


Figure 5.2: Example of a local tree

The following linearization rules serve to exclude orders such as those in (14):

- (15) $V[+MC] < X$
 $X < V[-MC]$

MC stands for *main clause*. The LP-rules ensure that in main clauses (+MC), the verb precedes all other constituents and follows them in subordinate clauses (−MC). There is a restriction that says that all verbs with the MC-value ‘+’ also have to be (+FIN). This will rule out infinitive forms in initial position.

These LP rules do not permit orders with an occupied prefield or postfield in a local tree. This is intended. We will see how fronting can be accounted for in Section 5.4.

5.1.3 metarules

We have previously encountered linearization rules for sentences with subjects, however our rules have the form in (16), that is, they do not include subjects:

- (16) $V2 \rightarrow H[7], N2[CASE\ dat]$
 $V2 \rightarrow H[8], N2[CASE\ dat], N2[CASE\ acc]$

It is not possible to analyze the verb phrases in (17) such as *dem Mann das Buch zu geben* and *das Buch dem Mann zu geben* as we did with the sentences in (9) since the subject does not occur on the right-hand side of the rule in (16).

- (17) a. Er verspricht, [dem Mann das Buch zu geben].
 he promises the.DAT man the.ACC book to give

- b. Er verspricht, [das Buch dem Mann zu geben].
 he promises the.ACC book the.DAT man to give
 ‘He promises to give the man the book.’

A rule with the format of (18) does not make much sense for a GPSG analysis of German since it cannot derive all the orders in (9) as the subject can occur between the elements of the VP as in (9c).

(18) $S \rightarrow N2\ V2$

With the rule in (18), it is possible to analyze (9a) as in Figure 5.3 and (9b) would also be possible with a different ordering of the NPs inside the VP. The remaining examples in (9) cannot be captured by the rule in (18), however. This has to do with the fact that only elements in the same local tree, that is, elements which occur on the right-hand side of the rule, can be moved. While we can move the parts of the VP and thereby derive (9b), it is not possible to move the subject to a lower position between the objects. Instead, a metarule can be used to analyze sentences where the subject occurs between other arguments of the verb. This rule relates phrase structure rules to other phrase structure rules. A metarule can be understood as a kind of instruction that creates another rule for each rule with a certain form and these newly created rules will in turn license local trees.

For the example at hand, we can formulate a metarule which says the following: If there is a rule with the form “V2 consists of something” in the grammar, then there also has to be another rule “V3 consists of whatever V2 consists + an NP in the nominative”. In formal terms, this looks as follows:

- (19) $V2 \rightarrow W \mapsto$
 $V3 \rightarrow W, N2[\text{CASE nom}]$

W is a variable which stands for an arbitrary number of categories ($W = \text{whatever}$). The metarule creates the following rule in (20) from the rules in (16):

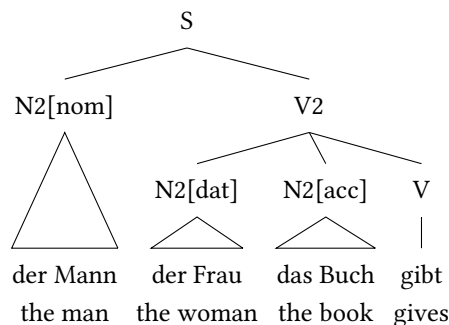


Figure 5.3: VP analysis for German (not appropriate in the GPSG framework)

- (20) $V3 \rightarrow H[7], N2[\text{CASE dat}], N2[\text{CASE nom}]$
 $V3 \rightarrow H[8], N2[\text{CASE dat}], N2[\text{CASE acc}], N2[\text{CASE nom}]$

Now, the subject and other arguments both occur in the right-hand side of the rule and can therefore be freely ordered as long as no LP rules are violated.

5.1.4 Semantics

The semantics adopted by Gazdar, Klein, Pullum & Sag (1985: Chapter 9–10) goes back to Richard Montague (1974). Unlike a semantic theory which stipulates the combinatorial possibilities for each rule (see Section 2.3), GPSG uses more general rules. This is possible due to the fact that the expressions to be combined each have a semantic type. It is customary to distinguish between entities (e) and truth values (t). Entities refer to an object in the world (or in a possible world), whereas entire sentences are either true or false, that is, they have a truth value. It is possible to create more complex types from the types e and t . Generally, the following holds: If a and b are types, then $\langle a, b \rangle$ is also a type. Examples of complex types are $\langle e, t \rangle$ and $\langle e, \langle e, t \rangle \rangle$. We can define the following combinatorial rule for these kind of typed expressions:

- (21) If α is of type $\langle b, a \rangle$ and β of type b , then $\alpha(\beta)$ is of type a .

This type of combination is also called *Functional Application*. With the rule in (21), it is possible that the type $\langle e, \langle e, t \rangle \rangle$ corresponds to an expression which still has to be combined with two expressions of type e in order to result in an expression of type t . The first combination step with e will yield $\langle e, t \rangle$ and the second step of combination with a further e will give us t . This is similar to what we saw with λ -expressions on page 64: $\lambda y \lambda x \text{ like}'(x, y)$ has to combine with a y and an x . The result in this example was $\text{mögen}'(\text{max}', \text{lotte}')$, that is, an expression that is either true or false in the relevant world.

In Gazdar et al. (1985), an additional type is assumed for worlds in which an expression is true or false. For reasons of simplicity, I will omit this here. The types that we need for sentences, NPs and N's, determiners and VPs are given in (22):

- (22) a. $\text{TYP}(S) = t$
 b. $\text{TYP}(\text{NP}) = \langle \langle e, t \rangle, t \rangle$
 c. $\text{TYP}(\text{N}') = \langle e, t \rangle$
 d. $\text{TYP}(\text{Det}) = \langle \text{TYP}(\text{N}'), \text{TYP}(\text{NP}) \rangle$
 e. $\text{TYP}(\text{VP}) = \langle e, t \rangle$

A sentence is of type t since it is either true or false. A VP needs an expression of type e to yield a sentence of type t . The type of the NP may seem strange at first glance, however, it is possible to understand it if one considers the meaning of NPs with quantifiers. For sentences such as (23a), a representation such as (23b) is normally assumed:

- (23) a. Alle Kinder lachen.
 all children laugh
 'All the children are laughing.'

$$\text{b. } \forall x \text{ child}'(x) \rightarrow \text{laugh}'(x)$$

The symbol \forall stands for the universal quantifier. The formula can be read as follows. For every object, for which it is the case that it has the property of being a child, it is also the case that it is laughing. If we consider the contribution made by the NP, then we see that the universal quantifier, the restriction to children and the logical implication come from the NP:

$$(24) \quad \forall x \text{ child}'(x) \rightarrow P(x)$$

This means that an NP is something that must be combined with an expression which has exactly one open slot corresponding to the x in (24). This is formulated in (22b): An NP corresponds to a semantic expression which needs something of type $\langle e, t \rangle$ to form an expression which is either true or false (that is, of type t).

An N' stands for a nominal expression for the kind $\lambda x \text{ child}(x)$. This means if there is a specific individual which one can insert in place of the x , then we arrive at an expression that is either true or false. For a given situation, it is the case that either John has the property of being of child or he does not. An N' has the same type as a VP.

N' and NP in (22d) stand for the types given in (22c) and (22b), that is, a determiner is semantically something which has to be combined with the meaning of N' to give the meaning of an NP.

Gazdar, Klein, Pullum & Sag (1985: 209) point out a redundancy in the semantic specification of grammars which follow the rule-by-rule hypothesis (see Section 2.3) since, instead of giving rule by rule instructions with regard to combinations, it suffices in many cases simply to say that the functor is applied to the argument. If we use types such as those in (22), it is also clear which constituent is the functor and which is the argument. In this way, a noun cannot be applied to a determiner, but rather only the reverse is possible. The combination in (25a) yields a well-formed result, whereas (25b) is ruled out.

- (25) a. $\text{Det}'(N')$
 b. $N'(\text{Det}')$

The general combinatorial principle is then as follows:

- (26) Use Functional Application for the combination of the semantic contribution of the daughters to yield a well-formed expression corresponding to the type of the mother node.

The authors of the GPSG book assume that this principle can be applied to the vast majority of GPSG rules so that only a few special cases have to be dealt with by explicit rules.

5.1.5 Adjuncts

For nominal structures in English, Gazdar et al. (1985: 126) assume the \bar{X} analysis and, as we have seen in Section 2.4.1, this analysis is applicable to nominal structures in German. Nevertheless, there is a problem regarding the treatment of adjuncts in the verbal

domain if one assumes flat branching structures, since adjuncts can freely occur between arguments:

- (27) a. weil der Mann der Frau das Buch *gestern* gab
 because the man the woman the book yesterday gave
 b. weil der Mann der Frau *gestern* das Buch gab
 because the man the woman yesterday the book gave
 c. weil der Mann *gestern* der Frau das Buch gab
 because the man yesterday the woman the book gave
 d. weil *gestern* der Mann der Frau das Buch gab
 because yesterday the man the woman the book gave
 ‘because the man give the book to the woman yesterday’

For (27), one requires the following rule:

- (28) $V3 \rightarrow H[8], N2[\text{CASE dat}], N2[\text{CASE acc}], N2[\text{CASE nom}], \text{AdvP}$

Of course, adjuncts can also occur between the arguments of verbs from other valency classes:

- (29) weil (oft) die Frau (oft) dem Mann (oft) hilft
 because often the woman often the man often helps
 ‘because the woman often helps the man’

Furthermore, adjuncts can occur between the arguments of a VP:

- (30) Der Mann hat versucht, der Frau heimlich das Buch zu geben.
 the man has tried the woman secretly the book to give
 ‘The man tried to secretly give the book to the woman.’

In order to analyze these sentences, we can use a metarule which adds an adjunct to the right-hand side of a V2 (Uszkoreit 1987: 146).

- (31) $V2 \rightarrow W \mapsto$
 $V2 \rightarrow W, \text{AdvP}$

By means of the subject introducing metarule in (19), the V3-rule in (28) is derived from a V2-rule. Since there can be several adjuncts in one sentence, a metarule such as (31) must be allowed to apply multiple times. The recursive application of metarules is often ruled out in the literature due to reasons of generative capacity (see Chapter 18) (Thompson 1982; Uszkoreit 1987: 146). If one uses the Kleene star, then it is possible to formulate the adjunct metarule in such a way that it does not have to apply recursively (Uszkoreit 1987: 146):

- (32) $V2 \rightarrow W \mapsto$
 $V2 \rightarrow W, \text{AdvP}^*$

If one adopts the rule in (32), then it is not immediately clear how the semantic contribution of the adjuncts can be determined.² For the rule in (31), one can combine the semantic contribution of the AdvP with the semantic contribution of the V2 with the insertion rule. This is of course also possible if the rule is applied multiple times. If this rule is applied to (33a), for example, the V-node in (33a) will receive the semantic contribution of the first adverb.

- (33) a. $V2 \rightarrow V, NP, AdvP$
 b. $V2 \rightarrow V, NP, AdvP, AdvP$

The V2-node in (33b) contains the semantic representation of the adverb when applied to the V2-node in (33a).

Weisweber & Preuss (1992) have shown that it is also possible to use rules such as (31) if one does not use metarules to calculate a set of phrase structure rules, but rather directly applies the metarules during the analysis of a sentence. Since sentences are always of finite length and the metarule introduces more AdvPs to the right-hand side of the rule, the metarule can only be applied a finite number of times.

5.2 Passive as a metarule

The German passive can be described in an entirely theory-neutral way as follows:³

- The subject is suppressed.
- If there is an accusative object, this becomes the subject.

This is true for all verb classes which can form the passive. It does not make a difference whether the verbs takes one, two or three arguments:

- (34) a. weil er noch gearbeitet hat
 because he still worked has
 'because he has still worked'
 b. weil noch gearbeitet wurde
 because still worked was
 'because people were still working'

² In LFG, an adjunct is entered into a set in the functional structure (see Section 7.1.6). This also works with the use of the Kleene Star notation. From the f-structure, it is possible to calculate the semantic denotation with corresponding scope by making reference to the c-structure. In HPSG, Kasper (1994) has made a proposal which corresponds to the GPSG proposal with regard to flat branching structures and an arbitrary number of adjuncts. In HPSG, however, one can make use of so-called relational constraints. This are similar to small programs which can create relations between values inside complex structures. Using such relational constraints, it is then possible to calculate the meaning of an unrestricted number of adjuncts in a flat branching structure.

³ This characterization does not hold for other languages. For instance, Icelandic allows for dative subjects. See Zaenen, Maling & Thráinsson (1985).

- (35) a. weil er an Maria gedacht hat
because he on Mary thought has
'because he thought of Mary'
b. weil an Maria gedacht wurde
because on Mary thought was
'because Mary was thought of'
- (36) a. weil sie ihn geschlagen hat
because she it him has
'because she has hit him'
b. weil er geschlagen wurde
because he hit was
'because he was hit'
- (37) a. weil er ihm den Aufsatz gegeben hat
because he him the essay given has
'because he has given him the essay'
b. weil ihm der Aufsatz gegeben wurde
because him the essay given was
'because he was given the essay'

In a simple phrase structure grammar, we would have to list two separate rules for each pair of sentences making reference to the valency class of the verb in question. The characteristics of the passive discussed above would therefore not be contained in these rules. In GPSG, it is possible to explain this relation using a metarule: For an active rule with a subject and an accusative object, a corresponding passive rule with suppressed subject is also licensed. The link between active and passive clauses can therefore be captured in this way.

An important difference to Transformational Grammar/GB is that we are not creating a relation between two trees, but rather between active and passive rules. The two rules result in the derivation of two unrelated structures, that is, the structure of (38b) is not related to the structure of (38a).

- (38) a. weil sie ihn geschlagen hat
because she him beaten has
'because she has beaten him'
b. weil er geschlagen wurde
because he beaten was
'because he was beaten'

The generalization with regard to active/passive is captured nevertheless.

In what follows, I will discuss the analysis of the passive given in Gazdar, Klein, Pullum & Sag (1985) in some more detail: The authors suggest the following metarule for English:⁴

- (39) $VP \rightarrow W, NP \mapsto$
 $VP[PAS] \rightarrow W, (PP[by])$

This rule states that verbs which take an object can occur in a passive VP without this object. Furthermore, a *by*-PP can be added. If we apply this metarule to the rules in (40), then this will yield the rules listed in (41):

- (40) $VP \rightarrow H[2], NP$
 $VP \rightarrow H[3], NP, PP[to]$
- (41) $VP[PAS] \rightarrow H[2], (PP[by])$
 $VP[PAS] \rightarrow H[3], PP[to], (PP[by])$

It is possible to use the rules in (40) to analyze verb phrases in active sentences:

- (42) a. [_S The man [_{VP} devoured the carcass]]
 b. [_S The man [_{VP} handed the sword to Tracy]]

The combination of a VP with the subject is licensed by an additional rule ($S \rightarrow NP, VP$).

With the rules in (41), one can analyze the VPs in the corresponding passive sentences in (43):

- (43) a. [_S The carcass was [_{VP[PAS]} devoured (by the man)]]
 b. [_S The sword was [_{VP[PAS]} handed to Tracy (by the man)]]

At first glance, this analysis may seem odd as an object is replaced inside the VP by a PP which would be the subject in an active clause. Although this analysis makes correct predictions with regard to the syntactic well-formedness of structures, it seems unclear how one can account for the semantic relations. It is possible, however, to manipulate the information in the passive metarule in such a way that the *by*-PP will be correctly incorporated semantically (Gazdar et al. 1985: 219).

We arrive at a problem, however, if we try to apply this analysis to German since the impersonal passive cannot be derived by simply suppressing an object. The V2-rules for verbs such as *arbeiten* ‘work’ and *denken* ‘think’ as used for the analysis of (34a) and (35a) have the following form:

- (44) $V2 \rightarrow H[5]$
 $V2 \rightarrow H[13], PP[an]$

⁴ See Weisweiler & Preuss (1992: 1114) for a parallel rule for German which refers to accusative case on the left-hand side of the metarule.

There is no NP on the right-hand side of these rules which could be turned into a *von*-PP. If the passive is to be analysed as a rule suppressing the NP argument, then it should follow from the existence of the impersonal passive that the passive metarule has to be applied to rules which license finite clauses, since information about whether there is a subject or not is only present in rules for finite clauses.⁵ In this kind of system, the rules for finite sentences (V3) are the basic rules and the rules for V2 would be derived from these.

It would only make sense to have a metarule which applies to V3 for German since English does not have V3 rules which contain both the subject and its object on the right-hand side of the rule.⁶ For English, it is assumed that a sentence consists of a subject and a VP (see Gazdar et al. 1985: 139). The result is that we arrive at two very different analyses for the passive in English and German, which still do not capture the descriptive insight that the passive is the suppression of the subject and the subsequent promotion of the object. The central difference between German and English seems to be that English obligatorily requires a subject,⁷ which is why English does not have an impersonal passive. This is a property independent of passives, which affects the possibility of having a passive structure, however.

The problem with the GPSG analysis is the fact that valency is encoded in phrase structure rules and that subjects are not present in the rules for verb phrases. In the following chapters, we will encounter approaches from LFG, Categorical Grammar, HPSG and Construction Grammar which encode valency separately from phrase structure rules and therefore do not have a principle problem with impersonal passive.

See Jacobson (1987b: 394–396) for more problematic aspects of the passive analysis in GPSG and for the insight that a lexical representation of valency – as assumed in Categorical Grammar, GB, LFG and HPSG – allows for a lexical analysis of the phenomenon, which is however unformulable in GPSG for principled reasons having to do with the fundamental assumptions regarding valence representations.

5.3 Verb position

Uszkoreit (1987) analysed verb-initial and verb-final order as linearization variants of a flat tree. The details of this analysis have already been discussed in Section 5.1.2.

An alternative suggestion in a version of GPSG comes from (Jacobs 1986: 110): Jacobs' analysis is a rendering of the verb movement analysis in GB. He assumes that there is an empty verb in final position and links this to the verb in initial position using technical means which we will see in more detail in the following section.

⁵ GPSG differs from GB in that infinitive verbal projections do not contain empty subjects. This is also true for all other theories discussed in this book with the exception of Tree-Adjoining Grammar.

⁶ Gazdar et al. (1985: 62) suggest a metarule similar to our rule for introducing subjects on page 185. This rule is used to analyze the position of auxiliaries in English and only licenses sequences of the form AUX NP VP. In such structures, subjects and objects are not in the same local tree.

⁷ Under certain conditions, the subject can also be omitted in English. For more on imperatives and other subject-less examples, see page 506.

5.4 Long-distance dependencies as the result of local dependencies

One of the main innovations of GPSG is its treatment of long-distance dependencies as a sequence of local dependencies (Gazdar 1981b). This approach will be explained taking constituent fronting to the pre-field in German as an example. Until now, we have only seen the GPSG analysis for verb-initial and verb-final position: The sequences in (45) are simply linearization variants.

- (45) a. [dass] der Mann der Frau das Buch gibt
 that the man the woman the book gives
 ‘that the man gives the book to the woman’
 b. Gibt der Mann der Frau das Buch?
 gives the man the woman the book
 ‘Does the man give the book to the woman?’

What we want is to derive the verb-second order in the examples in (46) from V1 order in (45b).

- (46) a. Der Mann gibt der Frau das Buch.
 the man gives the woman the book
 ‘The man gives the woman the book.’
 b. Der Frau gibt der Mann das Buch.
 the woman gives the man the book
 ‘The man gives the woman the book.’

For this, the metarule in (47) has to be used. This metarule removes an arbitrary category X from the set of categories on the right-hand side of the rule and represents it on the left-hand side with a slash (‘/’):⁸

- (47) $V3 \rightarrow W, X \mapsto$
 $V3/X \rightarrow W$

This rule creates the rules in (49) from (48):

- (48) $V3 \rightarrow H[8], N2[CASE\ dat], N2[CASE\ acc], N2[CASE\ nom]$
 (49) $V3/N2[CASE\ nom] \rightarrow H[8], N2[CASE\ dat], N2[CASE\ acc]$
 $V3/N2[CASE\ dat] \rightarrow H[8], N2[CASE\ acc], N2[CASE\ nom]$
 $V3/N2[CASE\ acc] \rightarrow H[8], N2[CASE\ dat], N2[CASE\ nom]$

The rule in (50) connects a sentence with verb-initial order with a constituent which is missing in the sentence:

⁸ An alternative to Uszkoreit’s trace-less analysis (1987: 77), which is explained here, consists of using a trace for the extracted element as in GB.

(50) $V3[+FIN] \rightarrow X[+TOP], V3[+MC]/X$

In (50), X stands for an arbitrary category which is marked as missing in V3 by the ‘/’. X is referred to as a *filler*.

The interesting cases of values for X with regard to our examples are given in (51).

(51) $V3[+FIN] \rightarrow N2[+TOP, CASE\ nom], V3[+MC]/N2[CASE\ nom]$
 $V3[+FIN] \rightarrow N2[+TOP, CASE\ dat], V3[+MC]/N2[CASE\ dat]$
 $V3[+FIN] \rightarrow N2[+TOP, CASE\ acc], V3[+MC]/N2[CASE\ acc]$

(51) does not show actual rules. Instead, (51) shows examples for insertions of specific categories into the X-position, that is, different instantiations of the rule.

The following linearization rule ensures that a constituent marked by [+TOP] in (50) precedes the rest of the sentence:

(52) $[+TOP] < X$

TOP stands for *topicalized*. As was mentioned on page 109, focussed elements and expletives as well as topics can also occur in the prefix, which is why this name is not ideal. However, it is possible to replace this with another term (e. g. *prefield*). This would not affect the analysis. X in (52) stands for an arbitrary category. This is a new X and is independent from the one in (50).

Figure 5.4 shows the interaction of the rules for the analysis of (53).⁹

(53) Dem Mann gibt er das Buch.
 the.DAT man gives he the book
 ‘He gives the man the book.’

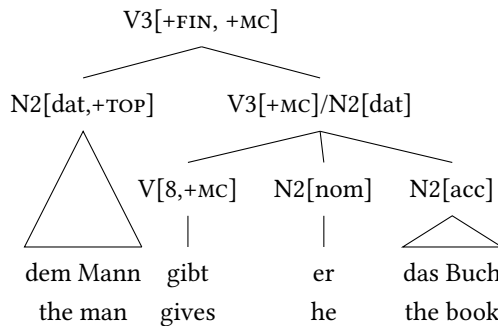


Figure 5.4: Analysis of fronting in GPSG

The metarule in (47) licenses a rule which adds a dative object into slash. This rule now licenses the subtree for *gibt er das Buch*. The linearization rule $V[+MC] < X$ orders the

⁹ The FIN feature has been omitted on some of the nodes since it is redundant: +MC-verbs always require the FIN value ‘+’.

verb to the very left inside of the local tree. In the next step, the constituent following the slash is bound off. Following the LP-rule [+TOP] < X, the bound constituent must be ordered to the left of the V3 node.

The analysis given in Figure 5.4 may seem too complex since the noun phrases in (53) all depend on the same verb. It is possible to invent a system of linearization rules which would allow one to analyze (53) with an entirely flat structure. One would nevertheless still need an analysis for sentences such as those in (37) on page 109 – repeated here as (54):

- (54) a. [Um zwei Millionen Mark]_i soll er versucht haben, [eine
around two million Deutsche.Marks should he tried have an
Versicherung _i zu betrügen].¹⁰
insurance.company to deceive
'He apparently tried to cheat an insurance company out of two million
Deutsche Marks.'
- b. „Wer_i, glaubt er, daß er _i ist?“ erregte sich ein Politiker vom Nil.¹¹
who believes he that he is retort REFL a politician from.the Nile
'Who does he think he is?', a politician from the Nile exclaimed.'
- c. Wen_i glaubst du, daß ich _i gesehen habe.¹²
who believe you that I seen have
'Who do you think I saw?'
- d. [Gegen ihn]_i falle es den Republikanern hingegen schwerer,
against him fall it the Republicans however more.difficult
[[Angriffe _i] zu lancieren].¹³
attacks to launch
'It is, however, more difficult for the Republicans to launch attacks against
him.'

The sentences in (54) cannot be explained by local reordering as the elements in the prefield are not dependent on the highest verb, but instead originate in the lower clause. Since only elements from the same local tree can be reordered, the sentences in (54) cannot be analyzed without postulating some kind of additional mechanism for long-distance dependencies.

In the following, I will discuss another example of fronting, namely one of the more complex examples in (54). The analysis of (54c) consists of several steps: the introduction, percolation and finally binding off of information about the long-distance dependency. This is shown in Figure 5.5 on the next page. Simplifying somewhat, I assume that *gesehen habe* ('have seen') behaves like a normal transitive verb.¹⁴ A phrase structure rule

¹⁰ taz, 04.05.2001, p. 20.

¹¹ Spiegel, 8/1999, p. 18.

¹² Scherpenisse (1986: 84).

¹³ taz, 08.02.2008, p. 9.

¹⁴ See Nerbonne (1986a) and Johnson (1986), for analyses of verbal complexes in GPSG.

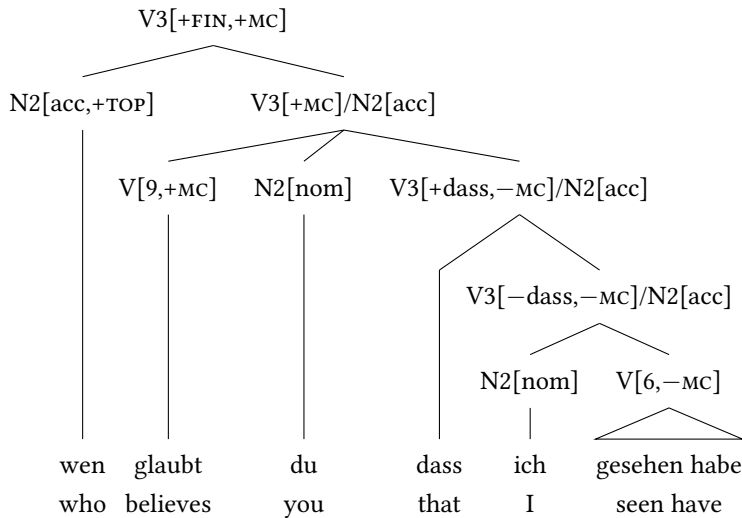


Figure 5.5: Analysis of long-distance dependencies in GPSG

licensed by the metarule in (47) licenses the combination of *ich* 'I' and *gesehen habe* 'has seen' and represents the missing accusative object on the V3 node. The complementizer *dass* 'that' is combined with *ich gesehen habe* 'I have seen' and the information about the fact that an accusative NP is missing is percolated up the tree. This percolation is controlled by the so-called *Foot Feature Principle* which states that all foot features of all the daughters are also present on the mother node. Since the SLASH-feature is a foot feature, the categories following the '/' percolate up the tree if they are not bound off in the local tree. In the final step, the V3/N2[acc] is combined with the missing N2[acc]. The result is a complete finite declarative clause of the highest projection type.

5.5 Summary and classification

Some twenty years after Chomsky's criticism of phrase structure grammars, the first large grammar fragment in the GPSG framework appeared and offered analyses of phenomena which could not be described by simple phrase structure rules. Although works in GPSG essentially build on Harman's 1963 idea of a transformation-less grammar, they also go far beyond this. A special achievement of GPSG is, in particular, the treatment of long-distance dependencies as worked out by Gazdar (1981b). By using the SLASH-mechanism, it was possible to explain the simultaneous extraction of elements from conjuncts (Across the Board Extraction, Ross 1967). The following examples from Gazdar (1981b: 173) show that gaps in conjuncts must be identical, that is, a filler of a certain category must correspond to a gap in every conjunct:

- (55) a. The kennel which Mary made and Fido sleeps in has been stolen.
(= S/NP & S/NP)
- b. The kennel in which Mary keeps drugs and Fido sleeps has been stolen.
(= S/PP & S/PP)
- c. * The kennel (in) which Mary made and Fido sleeps has been stolen.
(= S/NP & S/PP)

GPSG can plausibly handle this with mechanisms for the transmission of information about gaps. In symmetric coordination, the slash elements in each conjunct have to be identical. On the one hand, a transformational approach is not straightforwardly possible since one normally assumes in such analyses that there is a tree and something is moved to another position in the tree thereby leaving a trace. However, in coordinate structures, the filler would correspond to two or more traces and it cannot be explained how the filler could originate in more than one place.

The encoding of valency in GPSG is problematic for several reasons. For example, morphological processes take into account the valency properties of words. Adjectival derivation with the suffix *-bar* ‘-able’ is only productive with transitive verbs and verbs with an accusative object which can undergo passivization:

- (56) a. lös-bar (nominative, accusative)
solv-able
- b. vergleich-bar (nominative, accusative, PP[mit])
compar-able
- c. *schlaf-bar (nominative)
sleep-able
- d. *helf-bar (nominative, dative)
help-able

A rule for derivations with *-bar* ‘-able’ must therefore make reference to valency information. This is not possible in GPSG grammars since every lexical entry is only assigned one number which says something about the rules in which this entry can be used. For *-bar*-derivations, one would have to list in the derivational rule all the numbers which correspond to rules with accusative objects, which of course does not adequately describe the phenomenon. Furthermore, the valency of the resulting adjective also depends on the valency of the verb. For example, a verbs such as *vergleichen* ‘compare’ requires a *mit* (with)-PP and *vergleichbar* ‘comparable’ does too (Riehemann: 1993: 7, 54; 1998: 68). In the following chapters, we will encounter models which assume that lexical entries contain information as to whether a verb selects for an accusative object or not. In this kind of model, morphological rules which need to access the valency properties of linguistic objects can be adequately formulated.

Nerbonne (1986a) and Johnson (1986) investigate fronting of partial VPs in the GPSG framework.

- (57) a. Erzählen wird er seiner Tochter ein Märchen können.
tell will he his daughter a fairy.tale can

- b. Ein Märchen erzählen wird er seiner Tochter können.
a fairy.tale tell will he his daughter can
- c. Seiner Tochter ein Märchen erzählen wird er können.
his daughter a fairy.tale tell will he can
'He will be able to tell his daughter a fairy tale.'

The problem with sentences such as those in (57) is that the valency requirements of the verb *erzählen* 'to tell' are realized in various positions in the sentence. For fronted constituents, one requires a rule which allows a ditransitive to be realized without its arguments or with one or two objects. Furthermore, it has to be ensured that the arguments that are missing in the prefield are realized in the remainder of the clause. It is not legitimate to omit obligatory arguments or realize arguments with other properties like a different case, as the examples in (58) show:

- (58) a. Verschlungen hat er es nicht.
devoured has he it.ACC not
'He did not devour it.'
- b. *Verschlungen hat er nicht.
devoured has he not
- c. *Verschlungen hat er ihm nicht.
devoured has he him.DAT not

The obvious generalization is that the fronted and unfronted arguments must add up to the total set belonging to the verb. This is scarcely possible with the rule-based valency representation in GPSG. In theories such as Categorical Grammar (see Chapter 8), it is possible to formulate elegant analyses of (58) (Geach 1970). Nerbonne and Johnson both suggest analyses for sentences such as (58) which ultimately amount to changing the representation of valency information in the direction of Categorical Grammar.

In GPSG, the system of linearization, dominance and metarules is normally restricted by conditions we will not discuss here in such a way that one could create a phrase structure grammar of the kind we saw in Chapter 2 from the specification of a GPSG grammar. Such grammars are also called context-free grammars. In the mid-80s, it was shown that context-free grammars are not able to describe natural language (Shieber 1985; Culy 1985; see Pullum (1986) for a historical overview). Following the emergence of constraint-based models such as HPSG (see Chapter 9) and unification-based variants of Categorical Grammar (see Chapter 8 and Uszkoreit (1986a)), most authors previously working in GPSG turned to other frameworks. The GPSG analysis of long-distance dependencies and the distinction between immediate and linear order are still used in HPSG and variants of Construction Grammar to this day.

Comprehension questions

1. What does it mean for a grammar to be in an ID/LP-format?

2. How are linear variants of constituents in the middle-field handled by GPSG?
3. Think of some phenomena which have been described by transformations and consider how GPSG has analysed these data using other means.

Exercises

1. Write a small GPSG grammar which can analyze the following sentences:

- (59)
- a. [dass] der Mann ihn liest
that the man it reads
'that the man reads it'
 - b. [dass] ihn der Mann liest
that it the man reads
'that the man reads it'
 - c. Der Mann liest ihn.
the man reads it
'The man reads it.'

Include all arguments in a single rule without using the metarule for introducing subjects.

Further reading

The main publication in GPSG is Gazdar, Klein, Pullum & Sag (1985). This book has been critically discussed by Jacobson (1987b). Some problematic analyses are contrasted with alternatives from Categorical Grammar and reference is made to the heavily Categorical Grammar influenced work of Pollard (1984), which counts as one of the predecessors of HPSG. Some of Jacobson's suggestions can be found in later works in HPSG.

Grammars of German can be found in Uszkoreit (1987) and Busemann (1992). Gazdar (1981b) developed an analysis of long-distance dependencies, which is still used today in theories such as HPSG.

A history of the genesis of GPSG can be found in Pullum (1989b).

6 Feature Descriptions

In the previous chapter, we talked about sets of feature-value pairs, which can be used to describe linguistic objects. In this chapter, we will introduce feature descriptions which play a role in theories such as LFG, HPSG, Construction Grammar, versions of Categorical Grammar and TAG (and even some formalizations of Minimalist theories (Veenstra 1998)). This chapter will therefore lay some of the groundwork for the chapters to follow.

Feature structures are complex entities which can model properties of a linguistic object. Linguists mostly work with feature descriptions which describe only parts of a given feature structure. The difference between models and descriptions will be explained in more detail in Section 6.7.

Alternatively, feature structures are called:

- feature-value structure
- attribute-value structure

Other terms for feature description are the following:

- *attribute-value matrix* (AVM)
- *feature matrix*

In what follows, I will restrict the discussion to the absolutely necessary details in order to keep the formal part of the book as short as possible. I refer the interested reader to Shieber (1986), Pollard & Sag (1987: Chapter 2), Johnson (1988), Carpenter (1992), King (1994) and Richter (2004). Shieber's book is an accessible introduction to Unification Grammars. The works by King and Richter, which introduce important foundations for HPSG, would most probably not be accessible for those without a good grounding in mathematics. However, it is important to know that these works exist and that the corresponding linguistic theory is built on a solid foundation.

6.1 Feature Descriptions

When describing linguistic signs, we have to say something about their properties. For a noun, we can say that it has case, gender, number and person features. For a word such as *Mannes* 'man.GEN', we can say that these features have the values *genitive*, *masculine*, *singular* and 3. If we were to write these as a list of feature-value pairs, we would arrive at the following feature description:

6 Feature Descriptions

(1) Merkmal-Wert-Paare für *Mannes*:

CASE	<i>genitive</i>
GENDER	<i>masculine</i>
NUMBER	<i>singular</i>
PERSON	3

It is possible to describe a variety of different things using feature descriptions. For example, we can describe a person as in (2):

(2)	FIRSTNAME	<i>max</i>
	LASTNAME	<i>meier</i>
	DATE-OF-BIRTH	<i>10.10.1985</i>

People are related to other people – a fact that can also be expressed in feature-value pairs. For example, the fact that Max Meier has a father called Peter Meier can be captured by expanding (2) as follows:

(3)	FIRSTNAME	<i>max</i>		
	LASTNAME	<i>meier</i>		
	DATE-OF-BIRTH	<i>10.10.1985</i>		
	FATHER	FIRSTNAME	<i>peter</i>]
		LASTNAME	<i>meier</i>	
		DATE-OF-BIRTH	<i>10.05.1960</i>	
		FATHER	...	
MOTHER	...			
MOTHER	...			

The value of the FATHER-feature is another feature description containing the same features as (2).

In feature descriptions, a *path* is a sequence of features, which immediately follow each other. The *value of a path* is the feature description at the end of the path. Therefore, the value of FATHER|DATE-OF-BIRTH is *10.05.1960*.

One can think of many different features, which could be included in representations such as (3). One may wonder how to integrate information about offspring into (3).

An obvious solution would be to add features for DAUGHTER und SON:

(4)	FIRSTNAME	<i>max</i>
	LASTNAME	<i>meier</i>
	DATE-OF-BIRTH	<i>10.10.1985</i>
	FATHER	...
	MOTHER	...
	DAUGHTER	...

This solution is not satisfactory as it not immediately clear how one could describe a person with several daughters. Should one really introduce features such as DAUGHTER-1 or DAUGHTER-3?

$$(5) \left[\begin{array}{ll} \text{FIRSTNAME} & \textit{max} \\ \text{LASTNAME} & \textit{meier} \\ \text{DATE-OF-BIRTH} & \textit{10.10.1985} \\ \text{FATHER} & \dots \\ \text{MOTHER} & \dots \\ \text{DAUGHTER-1} & \dots \\ \text{DAUGHTER-2} & \dots \\ \text{DAUGHTER-3} & \dots \end{array} \right]$$

How many features do we want to assume? Where is the limit? What would the value of DAUGHTER-32 be?

For this case, it makes much more sense to use a list. Lists are indicated with angle brackets. Any number of elements can occur between these brackets. A special case is when no element occurs between the brackets. In this case, we call it an *empty list*. In the following example, Max Meier has a daughter called Clara, who herself has no daughter.

$$(6) \left[\begin{array}{ll} \text{FIRSTNAME} & \textit{max} \\ \text{LASTNAME} & \textit{meier} \\ \text{DATE-OF-BIRTH} & \textit{10.10.1985} \\ \text{FATHER} & \dots \\ \text{MOTHER} & \dots \\ \text{DAUGHTER} & \left\langle \begin{array}{ll} \text{FIRSTNAME} & \textit{clara} \\ \text{LASTNAME} & \textit{meier} \\ \text{DATE-OF-BIRTH} & \textit{10.10.2004} \\ \text{FATHER} & \dots \\ \text{MOTHER} & \dots \\ \text{DAUGHTER} & \langle \rangle \end{array} \right\rangle \end{array} \right]$$

Now, we are left with the question of sons. Should we add another list for sons? Do we want to differentiate between sons and daughters? It is certainly the case that the gender of the children forms an important characteristic, but these are characteristics of the objects themselves, since every person has a gender. The description in (7) therefore offers a more adequate representation.

At this point, one could ask why the parents are not included in a list as well. In fact, we find similar questions also in linguistic works: how is information best organized for the job at hand? For a representation of parental descriptions under separate features, one could argue that one can make certain claims about a mother or father without having to necessarily find a matching item in a list.

(7)	FIRSTNAME		max
	LASTNAME		meier
	DATE-OF-BIRTH		10.10.1985
	GENDER		male
	FATHER		...
	MOTHER		...
	CHILDREN	<	FIRSTNAME
			clara
			LASTNAME
			meier
			DATE-OF-BIRTH
			10.10.2004
			GENDER
			female
		>	FATHER
			...
			MOTHER
			...
			CHILDREN

If the order of the elements is irrelevant, then we could use sets rather than lists. Sets are written inside curly brackets.¹

6.2 Types

In the previous section, we introduced feature descriptions consisting of feature-value pairs and showed that it makes sense to allow for complex values for features. In this section, feature descriptions will be further expanded to include types. Feature descriptions which are assigned a type are also called *typed feature descriptions*. Types say something about which features can or must belong to a particular structure. The description previously discussed describes an object of the type *person*.

(8)	FIRSTNAME		max
	LASTNAME		meier
	DATE-OF-BIRTH		10.10.1985
	GENDER		male
	FATHER		...
	MOTHER		...
	CHILDREN		<..., ...>
	<i>person</i>		

Types are written in *italics*.

The specification of a type determines which properties a modelled object has. It is then only possible for a theorist to say something about these properties. Properties such

¹ The definition of a set requires many technicalities. In this book, sets will only be used for collections of semantic information. This can be done equally well using lists, which is why we avoid introducing sets here and instead use lists.

as OPERATING VOLTAGE are not relevant for objects of the type *person*. If we know the type of a given object, then we also know that this object must have certain properties even if we do not yet know their exact values. In this way, (9) is still a description of Max Meier even though it does not contain any information about Max' date of birth:

$$(9) \left[\begin{array}{ll} \text{FIRSTNAME} & \textit{max} \\ \text{LASTNAME} & \textit{meier} \\ \text{GENDER} & \textit{male} \\ \textit{person} & \end{array} \right]$$

We know, however, that Max Meier must have been born on some day since this is a description of the type *person*. The question *What is Max' date of birth?* makes sense for a structure such as (9) in a way that the question *Which operating voltage does Max have?* does not. If we know that an object is of the type *person*, then we have the following basic structure:

$$(10) \left[\begin{array}{ll} \text{FIRSTNAME} & \textit{firstname} \\ \text{LASTNAME} & \textit{lastname} \\ \text{DATE-OF-BIRTH} & \textit{date} \\ \text{GENDER} & \textit{gender} \\ \text{FATHER} & \textit{person} \\ \text{MOTHER} & \textit{person} \\ \text{CHILDREN} & \textit{list of person} \\ \textit{person} & \end{array} \right]$$

In (10) and (9), the values of features such as FIRSTNAME are in italics. These values are also types. They are different from types such as *person*, however, as no features belong to them. These kinds of features are called *atomic*.

Types are organized into hierarchies.

It is possible to define the subtypes *woman* and *man* for *person*. These would determine the gender of a given object. (11) shows the feature structure for the type *woman*, which is analogous to that of *man*

$$(11) \left[\begin{array}{ll} \text{FIRSTNAME} & \textit{firstname} \\ \text{LASTNAME} & \textit{lastname} \\ \text{DATE-OF-BIRTH} & \textit{date} \\ \text{GENDER} & \textit{female} \\ \text{FATHER} & \textit{person} \\ \text{MOTHER} & \textit{person} \\ \text{CHILDREN} & \textit{list of person} \\ \textit{female person} & \end{array} \right]$$

At this point, we could ask ourselves if we really need the feature GENDER. The necessary information is already represented in the type *woman*. The question if specific

information is represented by special features or whether it is stored in a type without a corresponding individual feature will surface again in the discussion of linguistic analyses. Both alternatives differ mostly in the fact that the information which is modelled by types is not immediately accessible for structure sharing, which is discussed in Section 6.4.

Type hierarchies play an important role in capturing linguistic generalizations, which is why type hierarchies and the inheritance of constraints and information will be explained with reference to a further example in what follows. One can think of type hierarchies as an effective way of organizing information. In an encyclopedia, the individual entries are linked in such a way that the entries for monkey and mouse will each contain a pointer to mammal. The description found under mammal does therefore not have to be repeated for the subordinate concepts. In the same way, if one wishes to describe various electrical appliances, one can use the hierarchy in Figure 6.1. The

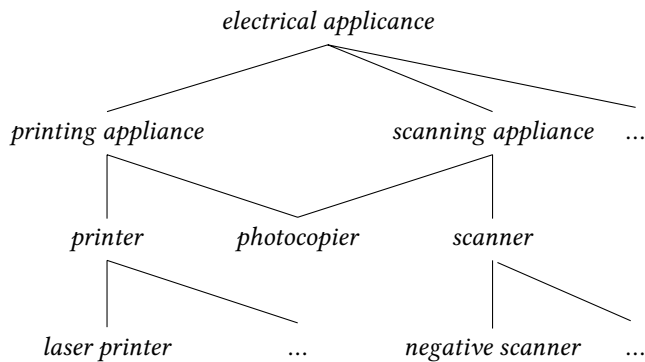


Figure 6.1: Non-linguistic example of multiple inheritance

most general type *electrical appliance* is the highest in Figure 6.1. Electrical appliances can have certain properties, e.g. a power supply with a certain power consumption. All subtypes of *electrical appliance* “inherit” this property. In this way, *printing appliance* and *scanning appliance* also have a power supply with a specific power consumption. A *printing appliance* can produce information and a *scanning appliance* can read in information. A *photocopier* can both produce information and read it. Photocopiers have both the properties of scanning and printing appliances. This is expressed by the connection between the two superordinate types and *photocopier* in Figure 6.1. If a type is at the same time the subtype of several superordinate types, then we speak of *multiple inheritance*. If appliances can print, but not scan, they are of type *printer*. This type can have further more specific subtypes, which in turn may have particular properties, e.g. *laser printer*. New features can be added to subtypes, but it is also possible to make values of inherited features more specific. For example, *negative scanner* is far more restricted when compared to its subordinate type *scanner*, since the material that can be scanned is restricted to negatives.

The objects that we model have a maximally specific type. In the example above, this means that we can have objects of the type *laser printer* and *negative scanner* but not of the type *printing appliance*. This is due to the fact that *printing appliance* is not maximally specific since this type has two subtypes.

Type hierarchies with multiple inheritance are an important means for expressing linguistic generalizations (Flickinger, Pollard & Wasow 1985; Flickinger 1987; Sag 1997). Types of words or phrases which occur at the very top of these hierarchies correspond to constraints on linguistic objects, which are valid for linguistic objects in all languages. Subtypes of such general types can be specific to certain languages or language types.

6.3 Disjunction

Disjunctions can be used if one wishes to express the fact that a particular object can have various different properties. If one were to organize a class reunion twenty years after leaving school and could not recall the exact names of some former classmates, it would be possible to search the web for “Julia (Warbanow or Barbanow)”. In feature descriptions, this “or” is expressed by a ‘ \vee ’.

$$(12) \begin{bmatrix} \text{FIRSTNAME} & \text{ julia} \\ \text{LASTNAME} & \text{ warbanow} \vee \text{ barbanow} \\ \text{person} \end{bmatrix}$$

Some internet search engines do not allow for searches with ‘or’. In these cases, one has to carry out two distinct search operations: one for “Julia Warbanow” and then another for “Julia Barbanow”. This corresponds to the two following disjunctively connected descriptions:

$$(13) \begin{bmatrix} \text{FIRSTNAME} & \text{ julia} \\ \text{LASTNAME} & \text{ warbanow} \\ \text{person} \end{bmatrix} \vee \begin{bmatrix} \text{FIRSTNAME} & \text{ julia} \\ \text{LASTNAME} & \text{ barbanow} \\ \text{person} \end{bmatrix}$$

Since we have type hierarchies as a means of expression, we can sometimes do without disjunctive specification of values and instead state the subordinate type: For *printer* \vee *photocopier*, one can simply write *printing appliance* if one assumes the type hierarchy in Figure 6.1 on the preceding page.

6.4 Structure sharing

Structure sharing is an important part of the formalism. It serves to express the notion that certain parts of a structure are identical. A linguistic example for the identity of values is agreement. In sentences such as (14), the number value of the noun phrase has to be identical to that of the verb:

- (14) a. Der Mann schläft.
the man sleeps
‘The man is sleeping.’
b. Die Männer schlafen.
the men sleep
‘The men are sleeping.’
c. *Der Mann schlafen.
the man sleep
Intended: ‘The man are sleeping.’

The identity of values is indicated by boxes containing numbers. The boxes can also be viewed as variables.

When describing objects we can make claims about equal values or claims about identical values. A claim about the identity of values is stronger. Let us take the following feature description containing information about the children that Max’s father and mother have as an example:

- (15)
$$\left[\begin{array}{ll} \text{FIRSTNAME} & \text{max} \\ \text{LASTNAME} & \text{meier} \\ \text{DATE-OF-BIRTH} & 10.10.1985 \\ \text{FATHER} & \left[\begin{array}{ll} \text{FIRSTNAME} & \text{peter} \\ \text{LASTNAME} & \text{meier} \\ \text{CHILDREN} & \left\langle \left[\begin{array}{ll} \text{FIRSTNAME} & \text{klaus} \\ & \text{person} \end{array} \right], \dots \right\rangle \\ & \text{person} \end{array} \right] \\ \text{MOTHER} & \left[\begin{array}{ll} \text{FIRSTNAME} & \text{anna} \\ \text{LASTNAME} & \text{meier} \\ \text{CHILDREN} & \left\langle \left[\begin{array}{ll} \text{FIRSTNAME} & \text{klaus} \\ & \text{person} \end{array} \right], \dots \right\rangle \\ & \text{person} \end{array} \right] \\ & \text{person} \end{array} \right]$$

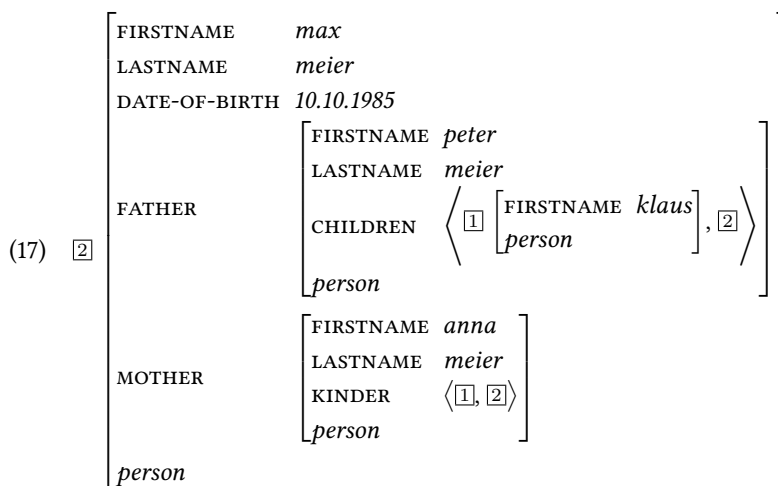
Notice that under the paths FATHER|CHILDREN and MOTHER|CHILDREN, we find a list of a description of a person with the first name Klaus. The question of whether the feature description is of one or two children of Peter and Anna cannot be answered. It is certainly possible that we are dealing with two different children from previous partnerships who both happen to be called Klaus.

By using structure sharing, it is possible to specify the identity of the two values:

209

We have introduced structure sharing in order to be able to express the fact that Max's parents both have a son Klaus together. It would not be enough to list Max in the child-lists of his parents separately. We want to capture the fact that it is the same Max which appears in each of these lists and furthermore, we have to ensure that the child being described is identical to the entire object being described. Otherwise, the description would permit a situation where Max's parents could have a second child also called Max. The description given in (17) can capture all facts correctly.

6.5 Cyclic structures



Structures such as those described in (17) are called cyclic because one ends up going in a circle if one follows a particular path: e.g. the path FATHER|CHILDREN|...|FATHER|CHILDREN|...² can be potentially repeated an infinite number of times.

6.6 Unification

Grammatical rules are written exactly like lexical entries in HPSG and Construction Grammar and are done so with the help of feature descriptions. Since a word or a larger phrasal entity can be used as daughter in a phrase licensed by some grammatical rule, the word or phrase must also have properties which are compatible with the description of the daughters in the grammatical rule. If this kind of compatibility exists, then we can say that the respective items are *unifiable*.³ If one unifies two descriptions, the result is a description which contains information from both descriptions but no additional information.

The way unification works can be demonstrated with feature descriptions describing people. One can imagine that Bettina Kant goes to the private detective Max Müller and wants to find a specific person. Normally, those who go to a detective's office only come with a partial description of the person they are looking for, e.g. the gender, hair color

² The dots here stand for the path to [2] in the list which is the value of CHILDREN. See Exercise 3.

³ The term *unification* should be used with care. It can only be properly used under certain assumptions with regard to the formal basis of linguistic theories. Informally, the term is often used in formalisms where unification is not technically defined. In HPSG, it mostly means that the constraints of two descriptions lead to a single description. What one wants to say here, intuitively, is that the objects described have to satisfy the constraints of both descriptions at the same time (*constraint satisfaction*). Since the term *unification* is so broadly-used, it will also be further used in this section. The term will not play a role in the remaining discussions of theories with the exception of explicitly unification-based approaches. In contrast, the concept for the satisfaction of constraints presented here is very important for the comprehension of the following chapters.

or date of birth. Perhaps even the registration number of the car belonging to the person is known.

It is then expected of the detective that he or she provides information fitting the description. If we are looking for a blonde female named Meier (18a), then we do not want to get descriptions of a male red-head (18b). The descriptions in (18) are incompatible and cannot be unified:

- (18) a.
$$\begin{bmatrix} \text{LASTNAME} & \textit{meier} \\ \text{GENDER} & \textit{female} \\ \text{HAIRCOLOR} & \textit{blonde} \\ \textit{person} \end{bmatrix}$$
- b.
$$\begin{bmatrix} \text{LASTNAME} & \textit{meier} \\ \text{GENDER} & \textit{male} \\ \text{HAIRCOLOR} & \textit{red} \\ \textit{person} \end{bmatrix}$$

The description in (19) would be a possible result for a search for a blonde, female individual called Meier:

- (19)
$$\begin{bmatrix} \text{FIRSTNAME} & \textit{katharina} \\ \text{LASTNAME} & \textit{meier} \\ \text{GENDER} & \textit{female} \\ \text{DATE-OF-BIRTH} & \textit{15.10.1965} \\ \text{HAIRCOLOR} & \textit{blonde} \\ \textit{person} \end{bmatrix}$$

Katharina Meier could also have other properties unknown to the detective. The important thing is that the properties known to the detective match those that the client is looking for. Furthermore, it is important that the detective uses reliable information and not make up any information about the sought object. The unification of the search in (18a) and the information accessible to the detective in (19) is in fact (19) and not (20), for example:

- (20)
$$\begin{bmatrix} \text{FIRSTNAME} & \textit{katharina} \\ \text{LASTNAME} & \textit{meier} \\ \text{GENDER} & \textit{female} \\ \text{DATE-OF-BIRTH} & \textit{15.10.1965} \\ \text{HAIRCOLOR} & \textit{blond} \\ \text{CHILDREN} & \langle \rangle \\ \textit{person} \end{bmatrix}$$

It could indeed be the case that Katharina Meier has children, but there are perhaps several people called Katharina Meier with the otherwise same properties. With this invented information, we would rule out one or more possible candidates.

It is possible that our detective Max Müller does not have any information about hair color in his files. His files could contain the following information:

(21)	[FIRSTNAME	<i>katharina</i>
		LASTNAME	<i>meier</i>
		GENDER	<i>weiblich</i>
		DATE-OF-BIRTH	<i>15.10.1965</i>
		<i>person</i>	

These data are compatible with the search criteria. If we were to unify the descriptions in (18a) and (21), we would get (19). If we assume that the detective has done all his work, then Bettina Kant now knows that the person she is looking for has the properties of her original search plus the newly discovered properties.

6.7 Phenomena, models and formal theories

In the previous sections, we introduced feature descriptions with types. These feature descriptions describe typed feature structures, which are models of observable linguistic structures. In the definitions of types, one determines which properties of linguistic objects should be described. The type hierarchy together with type definitions is also referred to as a *signature*. As a grammarian, one typically uses types in feature descriptions. These descriptions contain constraints which must hold for linguistic objects. If no constraints are given, all values that are compatible with the specification in the signature are possible values. For example, one can omit the case description of a linguistic object such as *Frau* ‘woman’ since *Frau* can – as shown in (22) – appear in all four cases:

- | | | | | | |
|------|----|---------|-------------|----------|--------------|
| (22) | a. | Die | Frau | schläft. | (nominative) |
| | | the.NOM | woman | sleeps | |
| | b. | Wir | gedenken | der | Frau. |
| | | we | commemorate | the.GEN | woman |
| | c. | Er | hilft | der | Frau. |
| | | he | helps | the.DAT | woman |
| | d. | Er | liebt | die | Frau. |
| | | he | loves | the | woman |
| | | | | | (accusative) |

In a given model, there are only fully specified representations, that is, the model contains four forms of *Frau*, each with a different case. For masculine nouns such as *Mann* ‘man’, one would have to say something about case in the description since the genitive-singular form *Mann-es* differs from other singular forms, which can be seen by adding *Mann* into the examples in (22). (23) shows the feature descriptions for *Frau* ‘woman’ and *Mann* ‘man’:

- (23) a. Frau ‘woman’

- $$\left[\begin{array}{l} \text{GENDER } fem \end{array} \right]$$
- b. Mann ‘man’
- $$\left[\begin{array}{l} \text{GENDER } mas \\ \text{CASE } \textit{nominative} \vee \textit{dative} \vee \textit{accusativ} \end{array} \right]$$

Unlike (23b), (23a) does not contain a case feature since we do not need to say anything special about case in the description of *Frau*. Since all nominal objects require a case feature, it becomes clear that the structures for *Frau* must actually also have a case feature. The value of the case feature is of the type *case*. *case* is a general type which subsumes the subtypes *nominative*, *genitive*, *dative* and *accusative*. Concrete linguistic objects always have exactly one of these maximally specified types as their case value. The feature structures belonging to (23) are given in Figure 6.2 and Figure 6.3.

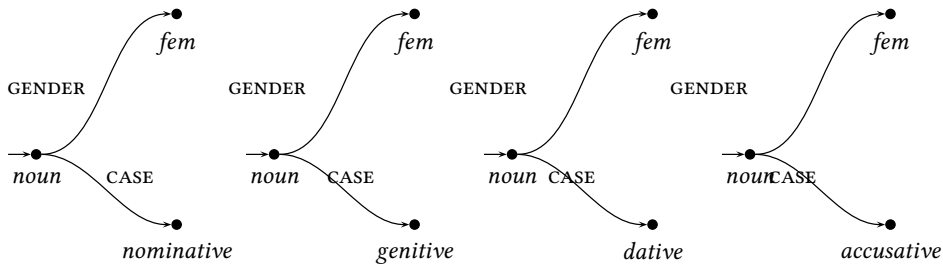


Figure 6.2: Feature structures for the description of *Frau* ‘woman’ in (23a)

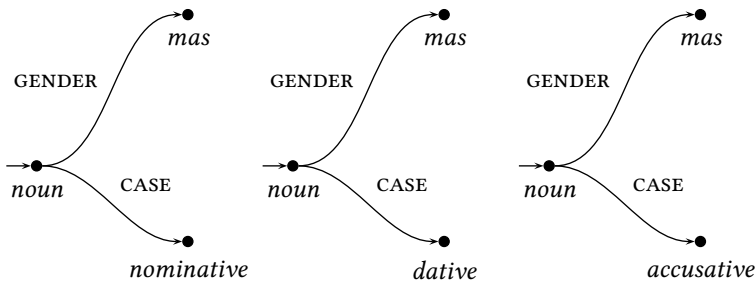


Figure 6.3: Feature structures for the description of *Mann* ‘man’ in (23b)

In these representations, a certain type (*noun*, *fem*, *nominative*, ...) belongs to each node, whereby the types in feature structures are always maximally specified, that is, they do not have any further sub-types. There is always an entry node (*noun* in the example above) and the other nodes are connected with arrows with the feature labels (*GENDER*, *CASE*).

If we return to the example with people from the previous sections, we can capture the difference between a model and a description as follows: if we have a model of

people that includes first name, last name, date of birth, gender and hair color, then it follows that every object we model also has a birthday. We can, however, decide to omit these details from our descriptions if they do not play a role for stating constraints or formulating searches.

The connection between linguistic phenomena, the model and the formal theory is shown in Figure 6.4. The model is designed to model linguistic phenomena. Further-

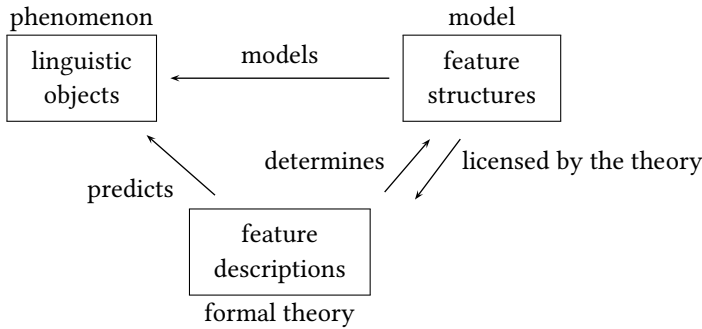


Figure 6.4: Phenomenon, model and formal theory

more, it must be licensed by our theory. The theory determines the model and makes predictions with regard to possible phenomena.

Comprehension questions

1. What are the reasons for using types?
2. What is inheritance? What is special about multiple inheritance?
3. Are the following structures compatible, that is, can they be used to describe the same object?

$$(24) \quad \left[\begin{array}{l} \text{FIRSTNAME } max \\ \text{LASTNAME } meier \\ \text{FATHER } \left[\begin{array}{l} \text{FIRSTNAME } peter \\ \text{LASTNAME } meier \\ person \end{array} \right] \end{array} \right] \quad \left[\begin{array}{l} \text{FIRSTNAME } max \\ \text{LASTNAME } meier \\ \text{FATHER } \left[\begin{array}{l} \text{FIRSTNAME } peter \\ \text{LASTNAME } m\ddot{u}ller \\ person \end{array} \right] \end{array} \right]$$

$$(25) \quad \left[\begin{array}{l} \text{FIRSTNAME } \textit{max} \\ \text{LASTNAME } \textit{meier} \\ \text{FATHER } \left[\begin{array}{l} \text{FIRSTNAME } \textit{peter} \\ \text{LASTNAME } \textit{meier} \\ \textit{person} \end{array} \right] \end{array} \right] \quad \left[\begin{array}{l} \text{FIRSTNAME } \textit{max} \\ \text{LASTNAME } \textit{meier} \\ \text{MOTHER } \left[\begin{array}{l} \text{FIRSTNAME } \textit{ursula} \\ \text{LASTNAME } \textit{müller} \\ \textit{person} \end{array} \right] \end{array} \right]$$

Exercises

1. Think about how one could describe musical instruments using feature descriptions.
2. Come up with a type hierarchy for the word classes (*det*, *comp*, *noun*, *verb*, *adj*, *prep*). Think about the ways in which one can formulate a type hierarchy so that one can express the feature divisions in Table 3.1 on page 96.
3. In this chapter, we introduced lists. This may look like an extension of the formalism, but it is not as it is possible to convert the list notation into a notation which only requires feature-value pairs. Think about how one could do this.
4. (Additional exercise) The relation *append* will play a role in Chapter 9. This relation serves to combine two lists to form a third. Relational constraints such as *append* do in fact constitute an expansion of the formalism. Using relational constraints, it is possible to combine any number of features to other values, that is, one can write programs, which calculate a particular value depending on other values. This poses the question as to whether one needs such powerful descriptive tools in a linguistic theory and if we do allow them, what kind of complexity we afford them. A theory which can do without relational constraints should be preferred over one that does use relational constraints (see Müller (2007b: Chapter 20) for a comparison of theories).

For the concatenation of lists, there is a possible implementation in feature structures without recourse to relational constraints. Find out what this is. Give your sources and document how you went about finding the solution.

Further reading

This chapter was designed to give the reader an easy-to-follow introduction to typed feature structures. The mathematical properties of the structures, type hierarchies and the combinatorial possibilities of such structures could not be discussed in detail here. Knowledge of at least part of these properties is important for work in computational linguistics and in developing one's own analyses. I refer the interested reader to the following publications: Shieber (1986) is a short introduction to the theory of Unification Grammar. There is a relatively general overview followed by the discussion of important

grammar types such as DCG, LFG, GPSG, HPSG, PATR-II. Johnson (1988) describes the formalism of untyped feature structures in a mathematically precise way. Carpenter (1992) goes into the detail about the mathematical aspects of typed feature structures. The formalism developed by King (1999) for HPSG-grammars forms the basis for the formalism by Richter (2004), which still counts as the standard analysis.

7 Lexical Functional Grammar

Lexical Functional Grammar (LFG) was developed in the 80s by Joan Bresnan and Ron Kaplan (Bresnan & Kaplan 1982). LFG forms part of so-called West-Coast linguistics: Unlike MIT, where Chomsky works and teaches, the institutes of researchers such as Joan Bresnan and Ron Kaplan are on the west coast of the USA (Joan Bresnan in Stanford and Ron Kaplan at Xerox in Palo Alto and now at the language technology firm Nuance Communications in Bay Area in California).

Bresnan & Kaplan (1982) view LFG explicitly as a psycholinguistically plausible alternative to transformation-based approaches. For a discussion of the demands of grammatical theories from a psycholinguistic point of view, see Chapter 15.

The more in-depth works on German are Berman (1996; 2003a); Cook (2001).

LFG has well-designed formal foundations (Kaplan & Bresnan 1982; Kaplan 1995), and hence first implementations were available rather quickly (Frey & Reyle 1983a;b; Yasukawa 1984; Block & Hunze 1986; Eisele & Dorre 1986; Kohl 1992; Kohl & Momma 1992; Wada & Asher 1986; Delmonte 1990; Her, Higinbotham & Pentheroudakis 1991; Kaplan & Maxwell III 1996; Mayo 1997; 1999; Boullier & Sagot 2005a;b; Clément 2009; Clément & Kinyon 2001).

The following is a list of languages with implemented LFG fragments, probably incomplete:

- Arrernte (Dras, Lareau, Börschinger, Dale, Motazedi, Rambow, Turpin & Ulinski 2014),
- Arabic (Attia 2008),
- Bengali (Sengupta & Chaudhuri 1997),
- Danish (Ørsnes 2002; Ørsnes & Wedekind 2003; 2004),
- German (Rohrer 1996; Berman 1996; Kuhn & Rohrer 1997; Butt et al. 1999a; Dipper 2003; Rohrer & Forst 2006; Forst 2006; Frank 2006; Forst & Rohrer 2009),
- English (Her, Higinbotham & Pentheroudakis 1991; Butt, Dipper, Frank & King 1999a; Riezler, King, Kaplan, Crouch, Maxwell III & Johnson 2002; King & Maxwell III 2007),
- French (Zweigenbaum 1991; Frank 1996; Frank & Zaenen 2002; Butt, Dipper, Frank & King 1999a; Clément & Kinyon 2001; Boullier, Sagot & Clément 2005),
- Georgian (Meurer 2009),

- Indonesian (Arka, Andrews, Dalrymple, Mistica & Simpson 2009),
- Italian (Delmonte 1990; Mayo 1999; Quaglia 2014),
- Irish (Sulger 2009; 2010),
- Japanese (Her, Higinbotham & Pentheroudakis 1991; Masuichi & Ohkuma 2003; Umemoto 2006),
- Korean (Her, Higinbotham & Pentheroudakis 1991),
- Malagasy (Randriamasimanana 2006; Dalrymple, Liakata & Mackie 2006),
- Mandarin Chinese (Her, Higinbotham & Pentheroudakis 1991; Fang & King 2007),
- Murrinh-Patha (Seiss & Nordlinger 2012),
- Norwegian (Dyvik, Meurer & Rosén 2005),
- Polish (Patejuk & Przepiórkowski 2012),
- Portuguese (Alencar 2004),
- Spanish (Mayo 1999),
- Tigrinya (Kifle 2012),
- Turkish (Çetinoğlu & Oflazer 2006),
- Hungarian (Laczkó, Rákosi & Tóth 2010; Rákosi, Laczkó & Csernyi 2011),
- Urdu/Hindi (Butt, King & Roth 2007; Bögel, Butt & Sulger 2008),
- Welsh (Mittendorf & Sadler 2005) and
- Wolof (Dione 2014; 2013).

Many of these grammars were developed in the ParGram consortium¹ (Butt, King, Niño & Segond 1999b; Butt, Dyvik, King, Masuichi & Rohrer 2002). Apart from these grammars there is a small fragment of Northern Sotho, which is currently being expanded (Faaß 2010).

Many of the LFG systems combine linguistically motivated grammars with a statistical component. Such a component can help to find preferred readings of a sentence first, it can increase the efficiency of processing and make the complete processing robust (for instance Kaplan et al. 2004; Riezler et al. 2002). Josef van Genabith's group in Dublin is working on the induction of LFG grammars from corpora (e.g. Johnson et al. 1999; O'Donovan et al. 2005; Cahill et al. 2005; Chrupala & van Genabith 2006; Guo et al. 2007; Cahill et al. 2008; Schluter & van Genabith 2009).

Some of the systems can be tested online:

¹ <http://pargram.b.uib.no/research-groups/>. 01.09.2014.

- <http://iness.uib.no/xle-web/xle-web>
- <http://lfg-demo.computing.dcu.ie/lfgparser.html>
- <http://www.xlfg.org/>

7.1 General remarks on the representational format

LFG assumes multiple levels of representation.² The most important are c-structure and f-structure. c-structure is the constituent structure and it is licensed by a phrase structure grammar. This phrase structure grammar uses \bar{X} structures for languages for which this is appropriate. f-structure stands for functional structure. Functional structure contains information about the predicates involved and about the grammatical functions (subject, object, ...) which occur in a constituent. Mappings mediate between these representational levels.

7.1.1 Functional structure

In LFG, grammatical functions such as subject and object play a very important role. Unlike in all other theories discussed in this book, they are primitives of the theory. A sentence such as (1a) will be assigned a functional structure as in (1b):

- (1) a. David devoured a sandwich.
- b.
- | | | | | |
|------|------|------------|--------------|---|
| [| PRED | ‘DEVOUR | (SUBJ, OBJ)’ |] |
| SUBJ | [| PRED | ‘DAVID’ |] |
| OBJ | [| SPEC | A |] |
| | PRED | ‘SANDWICH’ |] | |

Every lexical item contributes a PRED feature with a corresponding value. The grammatical functions governed by a head (government = subcategorization) are determined in the specification of PRED.³ Corresponding functions are called *governable grammatical functions*. Examples of this are shown in Table 7.1 on the following page (Dalrymple 2006). The PRED specification corresponds to the theta grid in GB theory. The valency of a head is specified by the PRED value.

The non-governable grammatical functions are given in Table 7.2 on the next page. Topic and focus are information-structural terms. There are a number of works on their exact definition, which differ to varying degrees (Kruijff-Korbayová & Steedman 2003: 253–254). Broadly speaking, one can say that the focus of an utterance constitutes new

² The English examples discussed in this section are taken from Dalrymple (2001) and Dalrymple (2006).

³ In the structure in (1b), the SUBJ and OBJ in the list following *devour* are identical to the values of SUBJ and OBJ in the structure. For reasons of presentation, this will not be explicitly indicated in this structure and following structures.

Table 7.1: Governable grammatical functions

SUBJ:	subject
OBJ:	object
COMP:	sentential complement or closed (non-predicative) infinitival complement
XCOMP:	open (predicative) complement, often infinitival, the SUBJ function is externally controlled
OBJ _θ :	secondary OBJ functions that are related to a special, language specific set of grammatical roles; English has OBJ _{THEME} only.
OBL _θ :	a group of thematically restricted oblique functions as for instance OBL _{GOAL} or OBL _{AGENT} . These often correspond to adpositional phrases in c-structure.

Table 7.2: Non-governable grammatical functions

ADJ:	adjuncts
TOPIC:	the topic of an utterance
FOCUS:	the focus of an utterance

information and that the topic is old or given information. Bresnan (2001: 97) uses the following question tests in order to determine topic and focus:

- (2) Q: What did you name your cat?
A: Rosie I named her. (*Rosie* = FOCUS)
- (3) Q: What did you name your pets?
A: My dog, I named Harold. My cat, I named Rosie. (*my dog, my cat* = TOPIC)

f-structures are characterized using functional descriptions, for example, one can refer to a value of the feature TENSE in the functional structure f using the following expression:

- (4) (f TENSE)

It is possible to say something about the value which this feature should have in the feature description. The following descriptions express the fact that in the structure f , the feature TENSE must have the value PAST.

- (5) (f TENSE) = PAST

We can also say that a feature has a specific f-structure as its value. The expression in (6) ensures that the SUBJ feature in f is the f-structure g :

$$(6) \quad (f \text{ SUBJ}) = g$$

For the analysis of (7a), we get the constraints in (7b):

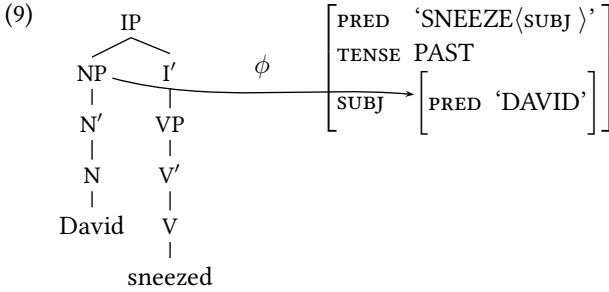
- (7) a. David sneezed.
 b. $(f \text{ PRED}) = \text{'SNEEZE'}\langle \text{SUBJ} \rangle$
 $(f \text{ TENSE}) = \text{PAST}$
 $(f \text{ SUBJ}) = g$
 $(g \text{ PRED}) = \text{'DAVID'}$

The description in (7) is true for the following structure:

$$(8) \quad f: \begin{bmatrix} \text{PRED} & \text{'SNEEZE'}\langle \text{SUBJ} \rangle \\ \text{TENSE} & \text{PAST} \\ \text{SUBJ} & g: \begin{bmatrix} \text{PRED} & \text{'DAVID'} \end{bmatrix} \end{bmatrix}$$

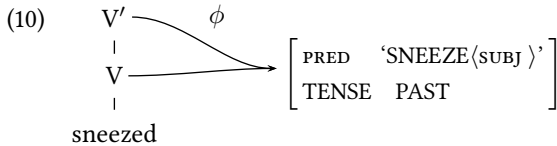
This description is true for many other structures which contain further features. We are only interested in structures that contain the information provided in the description.

(9) shows how a node in the c-structure can be connected to the f-structure for the entire sentence:



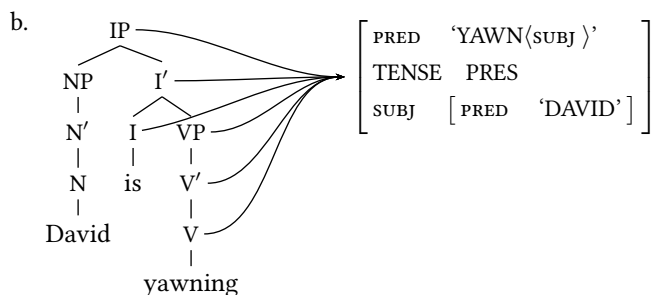
The function ϕ from the NP-node to the f-structure corresponding to the NP is notated with the arrow marked ϕ .

A phrase and its head always correspond to the same f-structure:



In LFG grammars of English, the CP/IP system is assumed as in GB theory (see Section 3.1.8). IP, I' and I (also VP) are mapped onto the same f-structure.

- (11) a. David is yawning.



f-structures have to fulfill two well-formedness conditions: They have to be both *complete* and *coherent*. Both these conditions will be discussed in the following sections.

7.1.2 Completeness

Every head adds another restriction for the PRED value of the corresponding f-structure. In determining completeness, one has to check that the elements required in the PRED are actually realized. In (12b), OBJ is missing a value, which is why (12a) is ruled out by the theory.

(12) a. *David devoured.

b.
$$\left[\begin{array}{ll} \text{PRED} & \text{'DEVOUR<SUBJ,OBJ>'} \\ \text{SUBJ} & [\text{PRED 'DAVID'}] \end{array} \right]$$

7.1.3 Coherence

The Coherence Constraint requires that all argument functions in a given f-structure have to be selected in the value of the local PRED-attributes. (13a) is ruled out because COMP does not appear under the arguments of *devour*.

(13) a. *David devoured a sandwich that Peter sleeps.

b.
$$\left[\begin{array}{ll} \text{PRED} & \text{'DEVOUR<SUBJ,OBJ>'} \\ \text{SUBJ} & [\text{PRED 'DAVID'}] \\ \text{OBJ} & \left[\begin{array}{ll} \text{SPEC} & \text{A} \\ \text{PRED} & \text{'SANDWICH'} \end{array} \right] \\ \text{COMP} & \left[\begin{array}{ll} \text{PRED} & \text{'SLEEP<SUBJ>'} \\ \text{SUBJ} & [\text{PRED 'PETER'}] \end{array} \right] \end{array} \right]$$

The constraints on completeness and coherence together ensure that only those arguments required in the PRED-specification are actually realized. Both of those constraints

taken together correspond to the Theta Criterion in GB theory (see page 93).⁴

7.1.4 Restrictions on the c-structure/f-structure relation

Symbols in c-structures are assigned restrictions for f-structures. The following symbols are used: ‘ \uparrow ’ refers to the f-structure of the immediately dominating node and ‘ \downarrow ’ refers to the f-structure of the c-structure node bearing the annotation. A common annotation is ‘ $\uparrow = \downarrow$ ’. This constraint states that the f-structure of the mother node is identical to that of the annotated category:

$$(14) \quad V' \rightarrow \begin{array}{c} V \\ \uparrow = \downarrow \end{array}$$

f-structure of the mother = own f-structure

The annotation ‘ $\uparrow = \downarrow$ ’ is below the head of a structure.

Phrases which are licensed by the annotated c-structure in (14) can be visualized as follows:

$$(15) \quad \begin{array}{c} V' \\ | \\ V \end{array} \rightarrow []$$

(16) shows a V' rule with an object:

$$(16) \quad V' \rightarrow \begin{array}{cc} V & NP \\ \uparrow = \downarrow & (\uparrow \text{ OBJ}) = \downarrow \end{array}$$

The annotation on the NP signals that the OBJ value in the f-structure of the mother ($\uparrow \text{ OBJ}$) is identical to the f-structure of the NP node, that is, to everything that is contributed from the material below the NP node (\downarrow). This is shown in the following diagram in (17):

$$(17) \quad \begin{array}{c} V' \\ \swarrow \quad \searrow \\ V \quad NP \end{array} \rightarrow \begin{array}{c} [\text{OBJ} \rightarrow []] \end{array}$$

In the equation $(\uparrow \text{ OBJ}) = \downarrow$, the arrows ‘ \uparrow ’ and ‘ \downarrow ’ correspond to feature structures. ‘ \uparrow ’ and ‘ \downarrow ’ stand for the f and g in equations such as (6). (18) is an example with an intransitive verb and the corresponding visualization is in (19):

$$(18) \quad \text{sneezed} \quad V \quad \begin{array}{l} (\uparrow \text{ PRED}) = \text{'SNEEZE' } \langle \text{SUBJ} \rangle \text{' } \\ (\uparrow \text{ TENSE}) = \text{PAST} \end{array}$$

$$(19) \quad \begin{array}{c} V \\ | \\ \text{sneezed} \end{array} \rightarrow \begin{bmatrix} \text{PRED} & \text{'SNEEZE' } \langle \text{SUBJ} \rangle \text{' } \\ \text{TENSE} & \text{PAST} \end{bmatrix}$$

⁴ For the differences between predicate-argument structures in LFG and the Deep Structure oriented Theta Criterion, see Bresnan & Kaplan (1982: xxvi–xxviii).

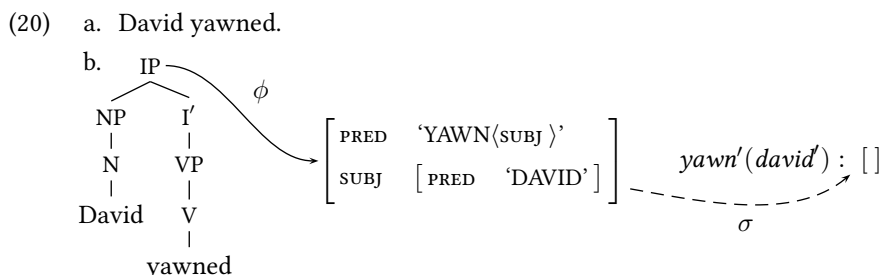
7.1.5 Semantics

Following Dalrymple (2006), *glue semantics* is the dominant approach to semantic interpretation in LFG (Dalrymple, Lamping & Saraswat 1993; Dalrymple 2001: Chapter 8). There are, however, other variants where Kamp's discourse representation structures (Kamp & Reyle 1993) are used (Frey & Reyle 1983a;b).

In the following, glue semantics will be presented in more detail.⁵ Under a glue-based approach, it is assumed that f-structure is the level of syntactic representation which is crucial for the semantic interpretation of a phrase, that is, unlike GB theory, it is not the position of arguments in the tree which play a role the composition of meaning, but rather functional relations such as *SUBJ* and *OBJ*. Glue semantics assumes that each sub-structure of the f-structure corresponds to a semantic resource connected to a meaning and furthermore, that the meaning of a given f-structure comes from the sum of these parts, although certain instructions for the combination of semantic resources must be respected. These instructions are given as a set of logic premises written in linear logic as *glue language*. The calculation of the meaning of an utterance corresponds to a logical conclusion.

This conclusion is reached on the basis of logical premises contributed by the words in an expression or possibly even by a syntactic construction itself. The requirements on how the meaning of the parts can be combined to yield the full meaning are expressed in linear logic, a resource-based logic. Linear logic is different from classic logic in that it does not allow the use of premises of conclusions more than once in a derivation. Hence, in linear logic, premises are resources which have to be used. This corresponds directly to the use of words in an expression: words contribute to the entire meaning exactly once. It is not possible to ignore them or to use their meaning more than once. A sentence such as *Peter knocked twice*. does not mean the same as *Peter knocked*. The meaning of *twice* must be included in the full meaning of the sentences. Similarly, the sentence cannot mean the same as *Peter knocked twice twice.*, since the semantic contribution of a given word cannot be used twice.

The syntactic structure for the sentence in (20a) together with its semantic representation is given in (20b):



The semantic structure of this sentence is connected to the f-structure via the correspon-

⁵ The following discussion heavily draws from the corresponding section of Dalrymple (2006). (It is a translation of my translation of the original material into German.)

dence function σ (depicted here as a dashed line). The semantic representation is derived from the following lexical information for the verb *yawned*:

$$(21) \quad \lambda x. \text{yawn}'(x) : (\uparrow \text{SUBJ})_\sigma \multimap \uparrow_\sigma$$

This formula is referred to as the *meaning constructor*. Its job is to combine the meaning of *yawned* – a one place predicate $\lambda x. \text{yawn}'(x)$ – with the formula $(\uparrow \text{SUBJ})_\sigma \multimap \uparrow_\sigma$ in linear logic. Here, the connective \multimap is the *linear implication* symbol of linear logic. The symbol contains the meaning that *if* a semantic resource $(\uparrow \text{SUBJ})_\sigma$ for the meaning of the subject is available, *then* a semantic resource for \uparrow_σ must be created which will stand for the entire meaning of the sentence. Unlike the implication operator of classic logic, the linear implication must consume and produce semantic resources: the formula $(\uparrow \text{SUBJ})_\sigma \multimap \uparrow_\sigma$ states that if a semantic resource $(\uparrow \text{SUBJ})_\sigma$ is found, it is consumed and the semantic resource \uparrow_σ is produced.

Furthermore, it is assumed that a proper name such as *David* contributes its own semantic structure as a semantic resource. In an utterance such as *David yawned*, this resource is consumed by the verb *yawned*, which requires a resource for its SUBJ in order to produce the resource for the entire sentence. This corresponds to the intuition that a verb in any given sentence requires the meaning of its arguments in order for the entire sentence to be understood.

The f-structure of *David yawned* with the instantiated meaning construction contributed by *David* and *yawned* is given in (22):

$$(22) \quad y : \left[\begin{array}{ll} \text{PRED} & \text{'YAWN'(\text{SUBJ})'} \\ \text{SUBJ} & d : \left[\begin{array}{ll} \text{PRED} & \text{'DAVID'} \end{array} \right] \end{array} \right]$$

$$[\text{David}] \quad \text{david}' : d_\sigma$$

$$[\text{yawn}] \quad \lambda x. \text{yawn}'(x) : d_\sigma \multimap y_\sigma$$

The left side of the meaning constructor marked by **[David]** is the meaning of the proper name *David*, *david'* to be precise. The left-hand side of the meaning constructor **[yawn]** is the meaning of the intransitive verb – a one-place predicate $\lambda x. \text{yawn}'(x)$.

Furthermore, one must still postulate further rules to determine the exact relation between the right-hand side (the glue) of the meaning constructors in (22) and the left-hand side (the meaning). For simple, non-implicational meaning constructors such as **[David]** in (22), the meaning on the left is the same as the meaning of the semantic structure on the right. Meaning constructors such as **[yawn]** have a λ -expression on the left, which has to be combined with another expression via Functional Application (see Section 2.3). The linear implication on the right-hand side must also be applied in parallel. This combined process is shown in (23).

$$(23) \quad \frac{x : f_\sigma \quad P : f_\sigma \multimap g_\sigma}{P(x) : g_\sigma}$$

The right-hand side of the rule corresponds to a logical conclusion following the *modus ponens* rule. With these correspondences between expressions in linear logic and the meanings themselves, we can proceed as shown in (24):

(24) $david' : d_\sigma$ The meaning $david'$ is combined with the semantic structure of SUBJ d_σ .

$\lambda x.yawn'(x) : d_\sigma \multimap y_\sigma$ If we find the semantic resource for the SUBJ d_σ on the glue side, this resource is consumed and the semantic resource for the entire sentence y_σ is produced. On the meaning side, we apply the function $\lambda x.yawn'(x)$ to the meaning connected to d_σ .

$yawn'(david') : y_\sigma$ We have created the semantic structure y_σ for the entire sentence, associated with the meaning of $yawn'(david')$.

After combining the respective meanings of *yawned* and *David* and then carrying out β -reduction, we arrive at the desired result of $yawn'(david')$ as the meaning of *David yawned*.

Glue analyses of quantification, modification and other phenomena have been investigated in a volume on glue semantics (Dalrymple 1999). Particularly problematic for these approaches are cases where there appear to be too many or too few resources for the production of utterances. These kinds of cases have been discussed by Asudeh (2004).

7.1.6 Adjuncts

Adjuncts are not selected by their head. The grammatical function ADJ is a non-governable grammatical function. Unlike arguments, where every grammatical function can only be realized once, a sentence can contain multiple adjuncts. The value of ADJ in the f-structure is therefore not a simple structure as with the other grammatical functions, but rather a set.

(25) a. David devoured a sandwich at noon yesterday.

b.
$$\left[\begin{array}{l} \text{PRED 'DEVOUR' \langle \text{SUBJ, OBJ} \rangle} \\ \text{SUBJ } \left[\begin{array}{l} \text{PRED 'DAVID'} \end{array} \right] \\ \text{OBJ } \left[\begin{array}{l} \text{SPEC A} \\ \text{PRED 'SANDWICH'} \end{array} \right] \\ \text{ADJ } \left\{ \left[\begin{array}{l} \text{PRED 'YESTERDAY'} \end{array} \right], \left[\begin{array}{l} \text{PRED 'AT' \langle \text{OBJ} \rangle} \\ \text{OBJ } \left[\begin{array}{l} \text{PRED 'NOON'} \end{array} \right] \end{array} \right] \right\} \end{array} \right]$$

The annotation on the c-structure rule for adjuncts requires that the f-structure of the adjuncts be part of the ADJ set of the mother's f-structure:

$$(26) \quad V' \rightarrow \begin{array}{cc} V' & PP \\ \uparrow = \downarrow & \downarrow \in (\uparrow \text{ ADJ}) \end{array}$$

The representation of adjuncts in a set is not sufficient to characterize the meaning of an utterance with scope-bearing adjuncts. In order to determine this kind of meaning, one has to refer to the linear order of the adjuncts, that is, their c-structure. For linearization restrictions in LFG, see Zaenen & Kaplan (1995).

7.2 Passive

Bresnan & Mchombo (1995) argue that one should view words as “atoms” of which syntactic structure is comprised (*lexical integrity*⁶).

Syntactic rules cannot create new words or make reference to the internal structure of words. Every terminal node (each “leaf” of the tree) is a word. It follows from this that analyses such as the GB analysis of Pollock (1989) in Figure 7.1 on the following page for the French example in (27) are ruled out (the figure is taken from Kuhn (2007: 617)):

- (27) Marie ne parlerait pas
 Marie NEG speak.COND.3SG NEG
 ‘Marie would not speak.’

In Pollock's analysis, the various morphemes are in specific positions in the tree and are combined only after certain movements have been carried out.

The assumption of lexical integrity is made by all theories discussed in this book with the exception of GB and Minimalism. However, formally, this is not a must as it is also possible to connect morphemes to complex syntactic structures in theories such as Categorical Grammar, GPSG, HPSG, CxG and TAG. As far as I know, this kind of analysis has never been proposed.

Bresnan noticed that, as well as passivized verbs, there are passivized adjectives, which show the same morphological idiosyncrasies as the corresponding particles (Bresnan 1982b: 21; Bresnan 2001: 31). Some examples are given in (28):

- (28) a. a well-written novel (write – written)
 b. a recently given talk (give – given)
 c. my broken heart (break – broken)
 d. an uninhabited island (inhabit – inhabited)
 e. split wood (split – split)

⁶ See Anderson (1992: 84) for more on lexical integrity.

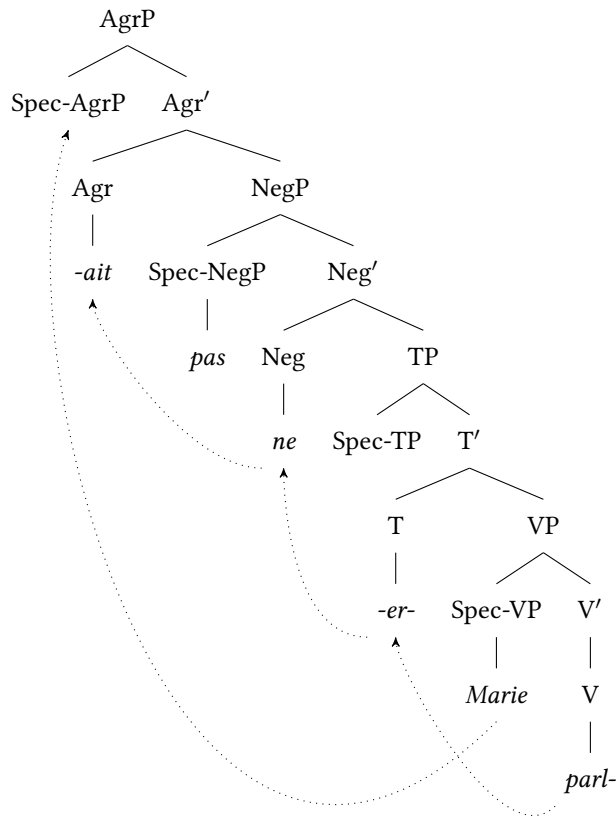


Figure 7.1: Pollock's analysis of *Marie ne parlerait pas* 'Marie would not speak.' according to Kuhn (2007: 617)

If one assumes lexical integrity, then adjectives would have to be derived in the lexicon. If the verbal passive were not a lexical process, but rather a phrase-structural one, then the form identity would remain unexplained.

In LFG, grammatical functions are primitives, that is, they are not derived from a position in the tree (e.g. Subject = SpecIP). Words (fully inflected word-forms) determine the grammatical function of their arguments. Furthermore, there is a hierarchy of grammatical functions. During participial formation in morphology, the highest verbal argument is suppressed. The next highest argument moves up and is not realized as the OBJECT but rather as the SUBJECT. This was explicitly encoded in earlier work (Bresnan 1982b: 8):

- (29) Passivization rule:
 (SUBJ) $\mapsto \emptyset / (\text{OBL})$
 (OBJ) $\mapsto (\text{SUBJ})$

The first rule states that the subject is either not realized (\emptyset) or it is realized as an oblique element (the *by* PP in English). The second rule states that if there is an accusative object, this becomes the subject.

In later work, the assignment of grammatical functions was taken over by Lexical Mapping Theory (Bresnan & Kanerva 1989). It is assumed that thematic roles are ordered in a universally valid hierarchy (Bresnan & Kanerva 1989; Bresnan 2001: 307): agent > beneficiary > experiencer/goal > instrument > patient/theme > locative. Patient-like roles are marked as unrestricted ($[-r]$) in a corresponding representation, the so-called a-structure. Secondary patient-like roles are marked as *objective*, $[+o]$ and all other roles marked as non-objective ($[-o]$). For the transitive verb *schlagen* ‘to beat’, we have the following:

(30)		Agent	Patient
	a-structure	<i>schlagen</i> ‘beat’	$\langle x \quad y \rangle$
		$[-o]$	$[-r]$

The mapping of a-structure to f-structure is governed by the following restrictions:

- (31) a. Subject-Mapping-Principle: The most prominent role marked with $[-o]$ is mapped to SUBJ if it is initial in the a-structure. Otherwise, the role marked with $[-r]$ is mapped to SUBJ.
- b. The argument roles are connected to grammatical functions as shown in the following table. Non-specified values for o and r are to be understood as ‘+’:
- | | | |
|--------|--------|------------------------------------|
| | $[-r]$ | $[+r]$ |
| $[-o]$ | SUBJ | OBL _{θ} |
| $[+o]$ | OBJ | OBJ _{θ} |
- c. *Function-Argument Bi-uniqueness*: Every a-structure role must be associated to exactly one function and vice versa.

For the argument structure in (30), the principle in (31a) ensures that the agent x receives the grammatical function SUBJ. (31b) adds an o-feature with the value ‘+’ so that the patient y is associated with OBJ:

(32)		Agent	Patient
	a-structure	<i>schlagen</i> ‘beat’	$\langle x \quad y \rangle$
		$[-o]$	$[-r]$
		SUBJ	OBJ

Under passivization, the most prominent role is suppressed so that only the $[-r]$ marked patient role remains. Following a, this role will then be mapped to the subject.

(33)		Agent	Patient
	a-structure	<i>schlagen</i> ‘beat’	$\langle x \quad y \rangle$
		$[-o]$	$[-r]$
		\emptyset	SUBJ

Unlike the objects of transitive verbs, the objects of verbs such as *helfen* ‘help’ are marked as [+o] (Berman 1999). The lexical case of the objects is given in the a-structure, since this case (dative) is linked to a semantic role (Zaenen, Maling & Thráinsson 1985: 465). The corresponding semantic roles are obligatorily mapped to the grammatical function OBJ_θ .

(34)	a-structure	<i>helfen</i> ‘help’	Agent	Beneficiary
			$\langle x$	$y \rangle$
			$[-o]$	$[+o]/\text{DAT}$
			SUBJ	OBJ_θ

Passivization will yield the following:

(35)	a-structure	<i>helfen</i> ‘help’	Agent	Beneficiary
			$\langle x$	$y \rangle$
			$[-o]$	$[+o]/\text{DAT}$
			\emptyset	OBJ_θ

Since there is neither a $[-o]$ nor a $[-r]$ argument, no argument is connected to the subject function. The result is an association of arguments and grammatical functions, which corresponds to that found by impersonal passives.

These mapping principles may seem complex at first glance, but they play a role in analyzing an entire range of phenomena, e.g. the analysis of unaccusative verbs (Bresnan & Zaenen 1990). For the analysis of the passive, we can now say that the passive suppresses the highest $[-o]$ role. Any mention of some other possibly present object in the passive rule is no longer necessary.

7.3 Verb position

There are two possibilities for the analysis of verb placement in German.

- a trace in verb-final position (as in GB) (see Choi (1999), Berman (1996: Section 2.1.4)) and
- so-called *extended head domains* (see Berman (2003a)).

In the analysis of extended head domains, the verb is simply omitted from the verb phrase. The following preliminary variant of the VP rule is used:⁷

(36)	$\text{VP} \rightarrow (\text{NP}) (\text{NP}) (\text{V})$
------	--

All components of the VP are optional as indicated by the brackets. As in GB analyses, the verb in verb-first clauses is in C. The I projection is – as in a number of GB works (Haider 1993; 1995; 1997a; Sternefeld 2006: Section IV.3) – omitted, since it is difficult to

⁷ See Bresnan (2001: 110) and Dalrymple (2006: Section 2.2) for a corresponding rule with optional constituents on the right-hand side of the rule.

motivate its existence for German (Berman 2003a: Section 3.2.2). The verb contributes its f-structure information from the C position. Figure 7.2 contains a simplified version of the analysis proposed by Berman (2003a: 41).

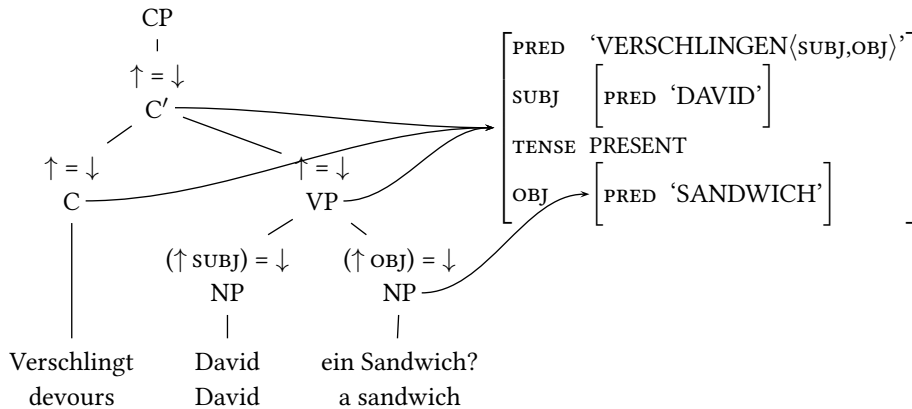


Figure 7.2: Analysis of verb placement following Berman (2003a: 41)

After what we learned about phrase structure rules in Chapters 2 and 5, it may seem strange to allow VPs without V. This is not a problem in LFG, however, since for the analysis of a given sentence, it only has to be ensured that all the necessary parts (and only these) are present. This is ensured by the constraints on completeness and coherence. Where exactly the information comes from is not important. In Figure 7.2, the verb information does not come from the VP, but rather from the C node. C' is licensed by a special rule:

$$(37) \quad C' \rightarrow \begin{array}{cc} C & VP \\ \uparrow = \downarrow & \uparrow = \downarrow \end{array}$$

In LFG rules, there is normally only one element listed under ' $\uparrow = \downarrow$ ', namely the head. In (37), there are two such elements, which is why both equally contribute to the f-structure of the mother. The head domain of V has been extended to C. The information about SUBJ and OBJ comes from the VP and the information about PRED from C.

7.4 Local reordering

Two possibilities for treating local reordering have been discussed in the literature:⁸

- Movement of arguments from a base configuration as in GB (see Choi (1999))

⁸ Kaplan (1995: 20–21) shows how one can write grammars in the ID/LP-format in LFG. A GPSG-like analysis of German constituent order has not been proposed in the LFG framework.

- Direct derivation via phrase structure rules (see Berman: 1996: Section 2.1.3.1; 2003a)

If one assumes that traces are relevant for the semantic interpretation of a given structure, then the first option has the same problems as movement-based GB analyses. These have already been discussed in Section 3.5.

In what follows, I will present the analysis proposed by Berman (1996: Section 2.1.3) in a somewhat simplified form. Case and grammatical functions of verbal arguments are determined in the lexicon (Berman 1996: 22). (38) shows the lexical entry for the verb *verschlengen* ‘devour’:^{9,10}

- (38) *verschlingt* V (↑ PRED) = ‘VERSCHLINGEN(⟨SUBJ, OBJ⟩)’
 (↑ SUBJ AGR CAS) = NOM
 (↑ OBJ AGR CAS) = ACC
 (↑ TENSE) = PRESENT

Berman proposes an analysis that does not combine the verb with all its arguments and adjuncts at the same time, as was the case in GPSG. Instead, she chooses the other

⁹ The four cases in German can be represented using two binary features (GOV, OBL) (Berman 1996: 22). Nominative corresponds to GOV– and OBL– and accusative to GOV+ and OBL–. This kind of encoding allows one to leave case partially underspecified. If one does not provide a value for GOV then an element with OBL– is compatible with both nominative and accusative. Since this underspecification is not needed in the following discussion, I will omit this feature decomposition and insert the case values directly.

¹⁰ Alternative analyses derive the grammatical function of an NP from its case (Berman (2003a: 37) for German; Bresnan (2001: 187, 201) for German and Russian).

- (i) (↓ CASE) = ACC ⇒ (↑ OBJ) = ↓

Karttunen (1989: Section 2.1) makes a similar suggestion for Finnish in the framework of Categorical Grammar. Such analyses are not entirely unproblematic as case cannot always be reliably paired with grammatical functions. In German, as well as temporal accusatives (ii.a), there are also verbs with two accusative objects (ii.b–c) and predicative accusatives (ii.d).

- (ii) a. Er arbeitete den ganzen Tag.
 he worked the.ACC whole.ACC day
 b. Er lehrte ihn den Ententanz.
 he taught him.ACC the.ACC duck.dance
 c. Das kostet ihn einen Taler.
 that costs him.ACC a.ACC taler
 d. Sie nannte ihn einen Lügner.
 she called him..ACC a.ACC liar

All of these accusatives can occur in long-distance dependencies (see Section 7.5):

- (iii) Wen glaubst du, dass ich getroffen habe.
 who believe you that I met have
 ‘Who do you think I met?’

wen is not the object of *glauben* ‘believe’ and as such cannot be included in the f-structure of *glauben* ‘believe’. One would have to reformulate the implication in (i) as a disjunction of all possible grammatical functions of the accusative and still account for the fact that accusatives can come from a more deeply embedded f-structure.

extreme and assumes that the verb is not combined with an adjunct or an argument, but rather forms a VP directly. The rule for this is shown in (39):

$$(39) \quad \text{VP} \rightarrow \quad (\text{V})$$

$$\quad \quad \quad \uparrow = \downarrow$$

At first sight, it may seem odd that a V such as *verschlingen* ‘devour’ does not have the same distribution as a verb with its arguments. However, one should recall that the constraints pertaining to coherence and completeness of f-structures play an important role so that the theory cannot make incorrect predictions.

Since the verb can occur in initial position, the rule in (39) is marked as optional (see Section 7.3).

The following rule can be used additionally to combine the verb with its subject or object.

$$(40) \quad \text{VP} \rightarrow \quad \text{NP} \quad \text{VP}$$

$$\quad \quad \quad (\uparrow \text{SUBJ} \mid \text{OBJ} \mid \text{OBJ}_\theta) = \downarrow \quad \uparrow = \downarrow$$

The ‘|’ here stands for a disjunction, that is, the NP can either be the subject or the object of the superordinate f-structure. Since VP occurs both on the left and right-hand side of the rule in (40), it can be applied multiple times. The rule is not complete, however. One has to account for prepositional objects and adjuncts.

Figure 7.3 on the following page shows the analysis for (41a).

- (41) a. [dass] David den Apfel verschlingt
 that David the apple devours
- b. [dass] den Apfel David verschlingt
 that the apple David devours
 ‘that David is devouring the apple.’

The analysis of (41b) is shown in Figure 7.4 on the next page. The analysis of (41b) differs from the one of (41a) only in the order of the replacement of the NP node by the subject or object.

One further fact must be discussed: In the rule (39), the verb is optional. If it is omitted, the VP is empty. In this way, the VP-rule in (40) can have an empty VP on the right-hand side of the rule. This is also simply omitted even though the VP symbol in the right-hand side of rule (40) is not marked as optional. That is, the corresponding symbol then also becomes optional as a result of taking the rest of the grammar into consideration as well as possible interactions with other rules.

7.5 Long-distance dependencies and functional uncertainty

We have seen that LFG can explain phenomena such as passivization, local reordering as well as verb placement without transformations. In Chapter 5 on GPSG, we already

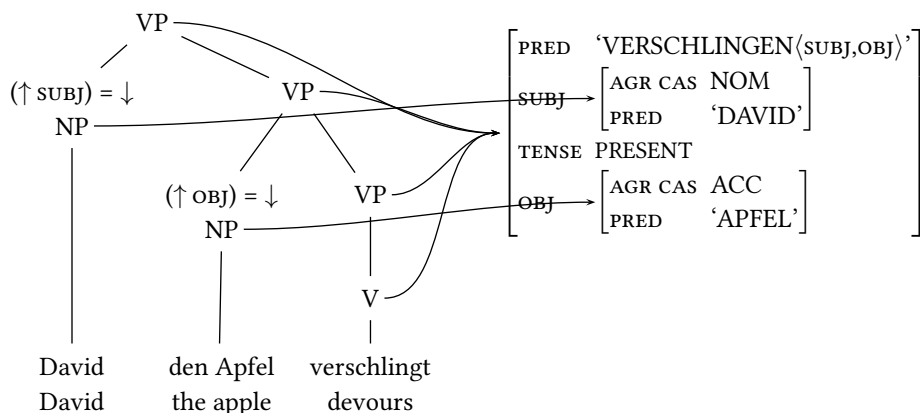


Figure 7.3: Analysis of SOV order following Berman (1996)

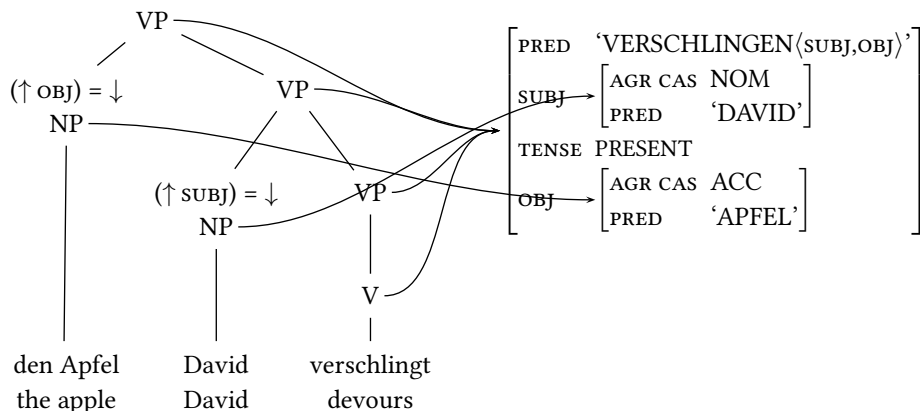


Figure 7.4: Analysis of OSV order following Berman (1996)

saw that the development of a transformation-less analysis for long-distance dependencies constitutes a real achievement. In LFG, Kaplan & Zaenen (1989) proposed another transformation-less analysis of long-distance dependencies, which we will consider in further detail in what follows.

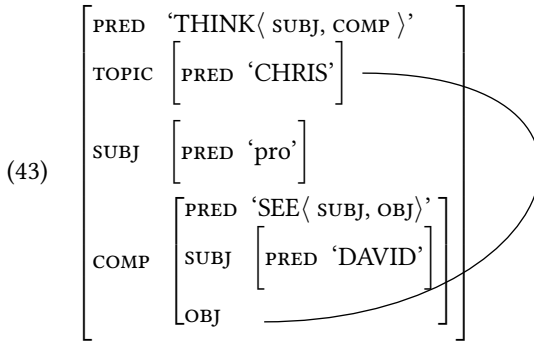
In example (42), the displaced constituent *Chris* is characterized by two functions:

- (42) Chris, we think that David saw.

For one, it has an argument function which is normally realized in a different position (the OBJ function of *saw* in the above example) and additionally it has a discourse function: a certain emphasis of the information-structural status in this construction (TOPIC in the matrix clause). In LFG, TOPIC and FOCUS are assumed to be grammaticalized dis-

course functions (furthermore, SUBJ is classified as the default discourse function). Only grammaticalized discourse functions are represented on the level of f-structure, that is, those created by a fixed syntactic mechanism and interact with the rest of the syntax.

Unlike argument functions, the discourse functions TOPIC and FOCUS are not lexically subcategorized and are therefore not subject to the completeness and coherence conditions. TOPIC and FOCUS are identified with an f-structure which bears an argument function. (43) gives the f-structure for the sentence in (42):

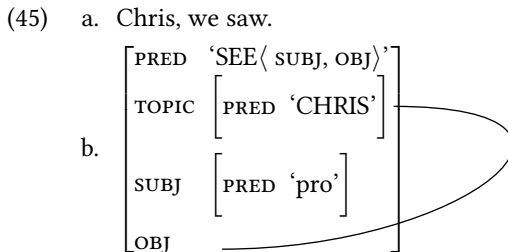


The connecting line means that the value of TOPIC is identical to the value of COMP|OBJ. In Chapter 6 on feature descriptions, I used boxes for structure sharing rather than connecting lines, since boxes are more common across frameworks. It is possible to formulate this as a constraint as in (44):

$$(44) \quad (\uparrow \text{TOPIC}) = (\uparrow \text{COMP OBJ})$$

Fronting operations such as (42) are also possible from various levels of embedding. (45a) shows an example with less embedding. The object is located in the same f-structure as the topic. However, the object in (42) comes from a clause embedded under *think*.

The f-structure corresponding to (45a) is given in (45b):

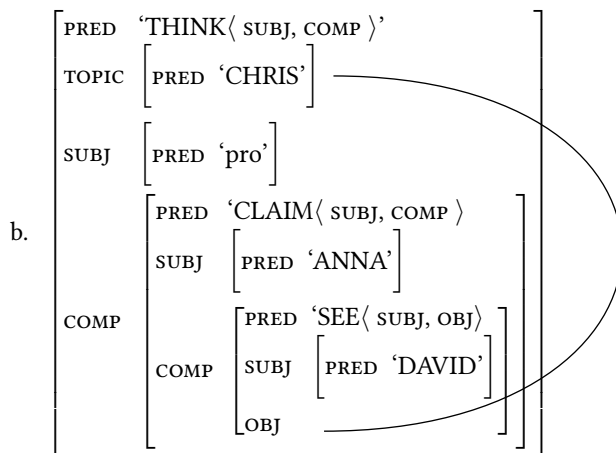


The identity restriction for TOPIC and object can be formulated in this case as in (46):

$$(46) \quad (\uparrow \text{TOPIC}) = (\uparrow \text{OBJ})$$

Example (47a) shows a case of even deeper embedding than in (42) and (47b,c) show the corresponding f-structure and the respective restriction.

- (47) a. Chris, we think Anna claims that David saw.



- c.
- $(\uparrow \text{TOPIC}) = (\uparrow \text{COMP COMP OBJ})$

The restrictions in (44), (46) and (47c) are c-structure constraints. The combination of a c-structure with (44) is given in (48):

- (48) CP \rightarrow XP C'
- $(\uparrow \text{TOPIC}) = \downarrow$ $\uparrow = \downarrow$
- $(\uparrow \text{TOPIC}) = (\uparrow \text{COMP OBJ})$

(48) states that the first constituent contributes to the TOPIC value in the f-structure of the mother and furthermore that this topic value has to be identical to that of the object in the complement clause. We have also seen examples of other embeddings of various depths. We therefore require restrictions of the following kind as in (49):

- (49) a. $(\uparrow \text{TOPIC}) = (\uparrow \text{OBJ})$
 b. $(\uparrow \text{TOPIC}) = (\uparrow \text{COMP OBJ})$
 c. $(\uparrow \text{TOPIC}) = (\uparrow \text{COMP COMP OBJ})$
 d. ...

The generalization emerging from these equations is given in (50):

- (50) $(\uparrow \text{TOPIC}) = (\uparrow \text{COMP}^* \text{OBJ})$

Here, '*' stands for an unrestricted number of occurrences of COMP. This means of leaving the possible identification of discourse and grammatical function open is known as *functional uncertainty*, see Kaplan & Zaenen (1989).

As was shown in the discussion of examples (2) and (3) on page 220, it is not the case that only a TOPIC can be placed in the specifier position of CP in English as FOCUS can occur there too. One can use disjunctions in LFG equations and express the corresponding condition as follows:

$$(51) \quad (\uparrow \text{TOPIC} | \text{FOCUS}) = (\uparrow \text{COMP}^* \text{OBJ})$$

One can introduce a special symbol for TOPIC|FOCUS: DF stands for a disjunction of discourse functions. (51) can then be abbreviated as in (52):

$$(52) \quad (\uparrow \text{DF}) = (\uparrow \text{COMP}^* \text{OBJ})$$

The final version of the c-structure rule for fronting in English will therefore have the form of (53):

$$(53) \quad \text{CP} \rightarrow \begin{array}{cc} \text{XP} & \text{C}' \\ (\uparrow \text{DF}) = \downarrow & \uparrow = \downarrow \\ (\uparrow \text{DF}) = (\uparrow \text{COMP}^* \text{OBJ}) & \end{array}$$

In German, as well as objects, nearly any other constituent (e.g. subjects, sentential complements, adjuncts) can be fronted. The c-structure rule for this is shown in (54):¹¹

$$(54) \quad \text{CP} \rightarrow \begin{array}{cc} \text{XP} & \text{C}' \\ (\uparrow \text{DF}) = \downarrow & \uparrow = \downarrow \\ (\uparrow \text{DF}) = (\uparrow \text{COMP}^* \text{GF}) & \end{array}$$

Here, GF is an abbreviation for a disjunction of grammatical functions which can occur in the prefield.

7.6 Summary and classification

LFG is a constraint-based theory and utilizes feature descriptions and PSG rules. Grammatical functions are treated as primitives of the theory, which sets LFG apart from all other theories covered in this book. They are not defined structurally (as in GB). LFG is a lexicalist theory. Like GPSG, LFG can do without transformations. Processes affecting argument structure such as passivization are analyzed by means of lexical rules. Whereas GPSG treated long-distance dependencies using the percolation of information in trees, LFG uses functional uncertainty: a part of the f-structure is identified with an f-structure with an arbitrary depth of embedding. Coherence and completeness ensure that the long-distance dependency can be correctly resolved, that is, it ensures that a fronted object is not assigned to an f-structure which already contains an object or one in which no object may occur.

While LFG does contain phrase-structural component, this plays a significantly less important role compared to other models of grammar. There are rules in which all constituents are optional and it has even been proposed for some languages that there are rules where the part of speech of the constituents is not specified (see Section 13.1.2).

¹¹ Berman (1996) uses the symbol ZP for symbols in the prefield rather than XP in (54). She formulates various phrase structure rules for ZPs, which replace ZP with NP, PP, AP and various adjuncts. Following Berman, ZPs can also be combined with the verb in the middle field. For reasons of exposition, I refrained from using ZPs symbols in the formulation of the VP rule (40) in Section 7.4 and instead used NP directly.

In these kinds of grammars, f-structure, coherence and completeness work together to ensure that the grammar only allows well-formed structures.

LFG differs from other theories such as HPSG and variants of Construction Grammar in that feature structures are untyped. Generalizations can therefore not be represented in type hierarchies. Until a few years ago, the hierarchical organization of knowledge in inheritance hierarchies did not form part of theoretical analyses. In computer implementations, there were macros but these were viewed as abbreviations without any theoretical status. It is possible to organize macros into hierarchies and macros were discussed explicitly in Dalrymple, Kaplan & King (2004) with reference to capturing linguistic generalizations. Asudeh, Dalrymple & Toivonen (2008) suggest using macros not only for the organization of lexical items but also for capturing generalizations regarding c-structure annotations. Because of these developments, there was a greater convergence between LFG and other theories such as HPSG and CxG.

Williams (1984) compares analyses in LFG with GB. He shows that many analyses are in fact transferable. The function that f-structure has in LFG is handled by the Theta Criterion and Case Theory in GB. LFG can explicitly differentiate between subjects and non-subjects. In GB, on the other hand, a clear distinction is made between external and internal arguments (see Williams 1984: Section 1.2). In some variants of GB, as well as in HPSG and CxG, the argument with subject properties (if there is one) is marked explicitly (Haider 1986a; Heinz & Matiassek 1994; Müller 2003b; Michaelis & Ruppenhofer 2001). This special argument is referred to as the *designated argument*. In infinitival constructions, subjects are often not expressed inside the infinitival phrase. Nevertheless, the unexpressed subject is mostly coreferential with an argument of the matrix verb:

- (55) a. Er versucht, [das Buch zu lesen].
 he tries the book to read
 ‘He is trying to read the book.’
 b. Er zwingt ihn, [das Buch zu lesen].
 he forces him the book to read
 ‘He is forcing him to read the book.’

This is a fact that every theory needs to be able to capture, that is, every theory must be able to differentiate between subjects and non-subjects.

For a comparison of GB/Minimalism and LFG/HPSG, see Kuhn (2007).

Comprehension questions

1. What do the terms *coherence* and *completeness* mean?
2. What are extended head domains?
3. What does lexical integrity mean?

Exercises

1. Give the lexical entry for *kannte* ‘knew’.
2. How could one analyze the following sentence?

(56) Den Apfel verschlingt David.
 the apple devours David
 ‘David devours the apple.’

Provide the necessary c-structure rules. What kind of f-structure is licensed? Draw a syntactic tree with corresponding references to the f-structure. For fronted constituents, simply write NP rather than expanding the XP node. The c-structure rule for the NP can also be omitted and a triangle can be drawn in the tree.

Further reading

Section 7.1 was based extensively on the textbook and introductory article of Dalrymple (2001; 2006). Additionally, I have drawn from teaching materials of Jonas Kuhn from 2007. Bresnan (2001) is a comprehensive textbook in English for the advanced reader. Some of the more in-depth analyses of German in LFG are Berman (1996; 2003a).

Levelt (1989) developed a model of language production based on LFG. Pinker (1984) – one of the best-known researchers on language acquisition – used LFG as the model for his theory of acquisition. For another theory on first and second language acquisition that uses LFG, see Pienemann (2005).

8 Categorical Grammar

Categorical Grammar is the oldest of the theories discussed here. It was developed in the 30s by the Polish logician Kazimierz Ajdukiewicz (Ajdukiewicz 1935). Since syntactic and semantic descriptions are tightly connected and all syntactic combinations correspond to semantic ones, Categorical Grammar is popular amongst logicians and semanticists. Some exceptional works in the field of semantics are those of Richard Montague (1974).

Other important works come from David Dowty in Columbus, Ohio (1979), Michael Moortgat in Utrecht (1989), Glyn Morrill in Barcelona (1994), Bob Carpenter in New York (1998) and Mark Steedman in Edinburgh (1991; 1997; 2000). A large fragment for German using Montague Grammar has been developed by von Stechow (1979). The 2569-page grammar of the *Institut für Deutsche Sprache* in Mannheim (Eroms, Stickel & Zifonun 1997) contains Categorical Grammar analyses in the relevant chapters. Fanselow (1981) has investigated morphology in the framework of Categorical Grammar. Uszkoreit (1986a), Karttunen (1986; 1989) and Calder et al. (1988) developed combinations of unification-based approaches and Categorical Grammar.

The basic operations for combining linguistic objects are rather simple and well-understood, so that it is no surprise that there are many systems for the development and processing of Categorical Grammars (Yampol & Karttunen 1990; Carpenter 1994; Bouma & van Noord 1994; Lloré 1995; König 1999; Moot 2002; White & Baldridge 2003; Baldridge et al. 2007; Morrill 2012). An important role is played by Mark Steedman's group (see for instance Clark et al. (2002); Clark & Curran (2007)).

Implemented fragments exist for the following languages:

- German (Uszkoreit 1986a; König 1999; Vierhuff, Hildebrandt & Eikmeyer 2003; Vancoppenolle, Tabbert, Bouma & Stede 2011)
- English (Villavicencio 2002; Baldridge 2002; Beavers 2003; 2004)
- Finnish (Karttunen 1989)
- French (Baschung, Bes, Corluy & Guillotin 1987)
- Dutch (Bouma & van Noord 1994; Baldridge 2002)
- Tagalog (Baldridge 2002)
- Turkish (Hoffman 1995; Baldridge 2002)

In addition, Baldridge et al. (2007: 15) mention an implementation for Classical Arabic.

Some of the systems for the processing of Categorical Grammars were augmented by statistics components, so that lexical items can be extracted from corpora and the

processing is robust (Osborne & Briscoe 1997; Clark, Hockenmaier & Steedman 2002). Briscoe (2000) and Villavicencio (2002) use statistical information in their UG-based language acquisition models.

8.1 General remarks on the representational format

8.1.1 Representation of valency information

In Categorial Grammar, complex categories replace the SUBCAT-feature that is used in GPSG to ensure that a head can only be used with suitable grammatical rules. Simple phrase structure rules can be replaced with complex categories as follows:

(1) Rule	Category in the lexicon
$vp \rightarrow v(\text{ditrans})\ np\ np$	$(vp/np)/np$
$vp \rightarrow v(\text{trans})\ np$	vp/np
$vp \rightarrow v(\text{np_and_pp})\ np\ pp(\text{to})$	$(vp/pp)/np$

vp/np stands for something that needs an np in order for it to form a vp .

In Categorial Grammar, there are only a few very abstract rules. One of these is forward application, also referred to as the multiplication rule:

- (2) forward application:
 $X/Y * Y = X$

This rule combines an X looking for a Y with a Y and requires that Y occurs to the right of X/Y . The result of this combination is an X that no longer requires a Y . X/Y is called the *functor* and Y is the *argument* of the functor.

Valency is encoded only once in Categorial Grammar, as in GB-Theory, in the lexicon. In GPSG, valency information was present in grammatical rules and in the SUBCAT-feature of the lexical entry.

Figure 8.1 shows how a lexical entry for a transitive verb is combined with its internal argument. A derivation in CG is basically a binary branching tree, it is however mostly

$$\frac{\frac{\text{chased}}{vp/np} \quad \frac{\text{Mary}}{np}}{vp} \rightarrow$$

Figure 8.1: Combination of a verb and its object (preliminary)

represented as follows: An arrow under a pair of categories indicates that these have been combined via a combinatorial rule. The direction of this arrow indicates the direction of this combination. The result is given beneath the arrow. Figure 8.2 on the next page shows the tree corresponding to Figure 8.1. One usually assumes left associativity for '/', that is, $(vp/pp)/np = vp/pp/np$.

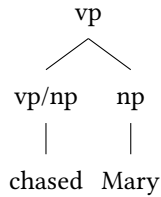


Figure 8.2: Derivation in Figure 8.1 as a tree diagram

If we look at the lexical entries in (1), it becomes apparent that the category v does not appear. The lexicon only determines what the product of combination of a lexical entry with its arguments is. The symbol for vp can also be eliminated: an (English) vp is something that requires an NP to its left in order to form a complete sentence. This can be represented as follows: $s \backslash np$. Using the rule for backward application, it is possible to calculate derivations such as the one in Figure 8.3.

(3) Backward application:

$$Y * X \backslash Y = X$$

$$\begin{array}{ccccccc}
 the & cat & chased & Mary \\
 \hline
 np/n & n & (s \backslash np)/np & np \\
 \hline
 & > & & > \\
 np & & s \backslash np & \\
 \hline
 & & s & <
 \end{array}$$

Figure 8.3: Analysis of a sentence with a transitive verb

In Categorical Grammar, there is no explicit difference made between phrases and words: An intransitive verb is described in the same way as a verb phrase with an object: $s \backslash np$. In the same way, proper nouns are complete noun phrases, which are assigned the symbol np .

8.1.2 Semantics

As already mentioned, Categorical Grammar is particularly popular among semanticists as syntactic combinations always result in parallel semantic combinations and this is also the case with complex combinations such as those we will see in more detail in the following sections, where meaning composition is defined precisely. In the following, we will take a closer look at the representational formats discussed in Steedman (1997: Section 2.1.2).

Steedman proposes the following lexical entry for the verb *eats*:¹

¹ I have adapted his notation to correspond to the one used in this book.

$$(4) \text{ eats} := (s : \text{eat}'(x, y) \backslash np_{3S} : x) / np : y$$

In (4), the meaning of each category is given after the colon. Since nothing is known about the meaning of the arguments in the lexical entry of *eat*, the meaning is represented by the variables x and y . When the verb combines with an NP, the denotation of the NP is inserted. An example is given in (5):²

$$(5) \frac{(s : \text{eat}'(x, y) \backslash np_{3S} : x) / np : y \quad np : \text{apples}'}{s : \text{eat}'(x, \text{apples}') \backslash np_{3S} : x} >$$

When combining a functor with an argument, it must be ensured that the argument fits the functor, that is, it must be unifiable with it (for more on unification see Section 6.6). The unification of $np : y$ with $np : \text{apples}'$ results in $np : \text{apples}'$ since *apples'* is more specific than the variable y . Apart from its occurrence in the term $np : y$, y occurs in the description of the verb in another position ($s : \text{eat}'(x, y) \backslash np_{3S} : x$) and therefore also receives the value *apples'* there. Thus, the result of this combination is $s : \text{eat}'(x, \text{apples}') \backslash np_{3S} : x$ as shown in (5).

Steedman notes that this notation becomes less readable with more complex derivations and instead uses the more standard λ -notation:

$$(6) \text{ eats} := (s \backslash np_{3S}) / np : \lambda y. \lambda x. \text{eat}'(x, y)$$

Lambdas are used to allow access to open positions in complex semantic representations (see Section 2.3). A semantic representation such as $\lambda y. \lambda x. \text{eat}'(x, y)$ can be combined with the representation of *apples* by removing the first lambda expression and inserting the denotation of *apples* in all the positions where the corresponding variable (in this case, y) appears (see Section 2.3 for more on this point):

$$(7) \frac{\lambda y. \lambda x. \text{eat}'(x, y) \quad \text{apples}'}{\lambda x. \text{eat}'(x, \text{apples}')}$$

This removal of lambda expressions is called β -reduction.

If we use the notation in (6), the combinatorial rules must be modified as follows:

$$(8) \begin{aligned} X/Y : f * Y : a &= X : f \ a \\ Y : a * X \backslash Y : f &= X : f \ a \end{aligned}$$

In the rules, the semantic contribution of the argument (a) is written after the semantic denotation of the functor (f). The open positions in the denotation of the functor are represented using lambdas. The argument can be combined with the first lambda expression using β -reduction.

Figure 8.4 on the next page shows the derivation of a simple sentence with a transitive verb. After forward and backward application, β -reduction is immediately applied.

² The assumption that *apples* means *apples'* and not *apples'(z)* minus the quantifier contribution is a simplification here.

$$\begin{array}{c}
\frac{\text{Jacob}}{np : jacob'} \quad \frac{\text{eats}}{(s \backslash np) / np : \lambda y. \lambda x. eat'(x, y)} \quad \frac{\text{apples}}{np : apples'} \\
\hline
\frac{s \backslash np : \lambda y. \lambda x. eat'(x, y) \quad apples'}{= \lambda x. eat'(x, apples')} > \\
\hline
\frac{}{s : \lambda x. eat'(x, apples') \quad jacob'} < \\
\hline
= eat'(jacob', apples')
\end{array}$$

Figure 8.4: Meaning composition in Categorical Grammar

8.1.3 Adjuncts

As noted in Section 1.6, adjuncts are optional. In phrase structure grammars, this is captured by the fact that an element on the left-hand side of a rule occurs with an adjunct on the right-hand side. Since the symbol on the left is the same as the one on the right, this rule can be applied iteratively. (9) shows some examples of this:

- (9) a. $VP \rightarrow VP \text{ PP}$
b. $\text{Noun} \rightarrow \text{Noun PP}$

One can analyse an arbitrary amount of PPs following a VP or noun using these rules.

In Categorical Grammar, adjuncts have the following general form: $X \backslash X$ or X / X . Adjectives are modifiers, which must occur before the noun. They have the category n / n . Modifiers occurring after nouns (prepositional phrases and relative clauses) have the category $n \backslash n$ instead.³ For VP-modifiers, X is replaced by the symbol for the VP ($s \backslash np$) and this yields the relatively complex expression $(s \backslash np) \backslash (s \backslash np)$. Adverbials in English are VP-modifiers and have this category. Prepositions that can be used in a PP modifying a verb require a NP in order to form a complete PP and therefore have the category $((s \backslash np) \backslash (s \backslash np)) / np$. Figure 8.5 gives an example of an adverb (*quickly*) and a preposition (*round*). Note that the result of the combination of *round* and *the garden* corresponds to

$$\begin{array}{ccccccccccc}
\text{The} & \text{small} & \text{cat} & \text{chased} & \text{Mary} & \text{quickly} & \text{round} & \text{the} & \text{garden} \\
np/n & n/n & n & (s \backslash np) / np & np & (s \backslash np) \backslash (s \backslash np) & (s \backslash np) \backslash (s \backslash np) / np & np / n & n \\
\hline
& n & & s \backslash np & & & & & \\
\hline
np & & & s \backslash np & & & & & \\
\hline
& & & & & & np & & \\
\hline
& & & & & & (s \backslash np) \backslash (s \backslash np) & & \\
\hline
& & & & & & (s \backslash np) & & \\
\hline
& & & & & & s & &
\end{array}$$

Figure 8.5: Example of analysis with adjuncts in Categorical Grammar

³ In Categorical Grammar, there is no category symbol like \overline{X} for intermediate projections of \overline{X} Theory. See Exercise 2.

the category of the adverb $((s\backslash np)\backslash(s\backslash np))$. In GB theory, adverbs and prepositions were also placed into a single class (See page 96). This overarching class was then divided into subclasses based on the valency of the elements in question.

8.2 Passive

In Categorial Grammar, the passive is analyzed by means of lexical rule (Dowty 1978: 412; Dowty 2003: Section 3.4). (10) shows the rule in Dowty (2003: 49).

- (10) Syntax: $\alpha \in (s\backslash np)/np \rightarrow \text{PST-PART}(\alpha) \in \text{PstP}/np_{by}$
 Semantics: $\alpha' \rightarrow \lambda y \lambda x \alpha'(y)(x)$

Here, PstP stands for past participle and np_{by} is an abbreviation for a verb phrase modifier of the form $vp\backslash vp$ or $(s\backslash np)\backslash(s\backslash np)$. The rule says the following: If a word belongs to the set of words with the category $(s\backslash np)/np$, then the word with past participle morphology also belongs in the set of words with the category PstP/np_{by} .

(11a) shows the lexical entry for the transitive verb *touch* and (11b) the result of rule application:

- (11) a. *touch*: $(s\backslash np)/np$
 b. *touched*: PstP/np_{by}

The auxiliary *was* has the category $(s\backslash np)/\text{PstP}$ and the preposition *by* has the category np_{by}/np , or its unabbreviated form $((s\backslash np)\backslash(s\backslash np))/np$. In this way, (12) can be analyzed as in Figure 8.6.

- (12) John was touched by Mary.

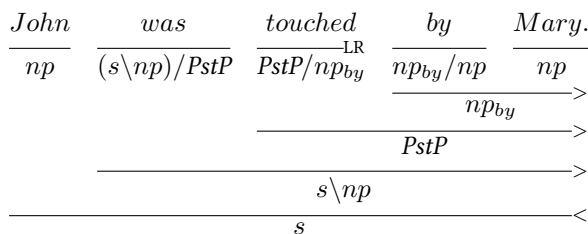


Figure 8.6: Analysis of the passive using lexical rules

The question as to how to analyze the pair of sentences in (13) still remains unanswered.⁴

- (13) a. He gave the book to Mary.
 b. The book was given to Mary.

⁴ Thanks to Roland Schäfer for pointing out these data to me.

gave has the category $((s\backslash np)/pp)/np$, that is, the verb must first combine with an NP (*the book*) and a PP (*to Mary*) before it can be combined with the a subject. The problem is that the rule in (10) cannot be applied to *gave* with a *to*-PP since the pp argument is sandwiched between both np arguments in $((s\backslash np)/pp)/np$. One would have to generalize the rule in (10) somehow by introducing new technical means⁵ or assume additional rules for cases such as (13b).

8.3 Verb position

Steedman (2000: 159) proposed an analysis with variable branching for Dutch, that is, there are two lexical entries for *at* ('eat'): an initial one with its arguments to the right, and another occupying final position with its arguments to its left.

- (14) a. *at* ('eat') in verb-final position: $(s_{+SUB} \backslash np) \backslash np$
 b. *at* ('eat') in verb-initial position: $(s_{-SUB}/np)/np$

Steedman uses the feature SUB to differentiate between subordinate and non-subordinate sentences. Both lexical items are connected via lexical rules.

One should note here that the NPs are combined with the verb in different orders. The normal order is:

- (15) a. in verb-final position: $(s_{+SUB} \backslash np[nom]) \backslash np[acc]$
 b. in verb-initial position: $(s_{-SUB}/np[acc])/np[nom]$

The corresponding derivations for German sentences with a bivalent verb are shown in Figures 8.7 and 8.8.

$$\begin{array}{c}
 \begin{array}{ccc}
 \textit{er} & \textit{ihn} & \textit{isst} \\
 \hline
 np[nom] & np[acc] & (s_{+SUB} \backslash np[nom]) \backslash np[acc]
 \end{array} \\
 \hline
 s_{+SUB} \backslash np[nom] < \\
 \hline
 s_{+SUB} <
 \end{array}$$

Figure 8.7: Analysis of verb-final sentences following Steedman

$$\begin{array}{c}
 \begin{array}{ccc}
 \textit{isst} & \textit{er} & \textit{ihn} \\
 \hline
 ((s_{-SUB}/np[acc])/np[nom]) & np[nom] & np[acc]
 \end{array} \\
 \hline
 s_{-SUB}/np[acc] > \\
 \hline
 s_{-SUB} >
 \end{array}$$

Figure 8.8: Analysis of verb-initial sentences following Steedman

⁵ Baldridge (p. M. 2010) suggests using regular expressions in a general lexical rule for passive.

In Figure 8.7, the verb is first combined with subject. For criticism of these kinds of analyses with variable branching, see Netter (1992) and Müller (2005b).

Jacobs (1991) developed an analysis which corresponds to the verb movement analysis in GB. He assumes verb-final structures, that is, there is a lexical entry for verbs where arguments are selected to the left of the verb. A transitive verb would therefore have the entry in (16a). Additionally, there is a trace in verb-final position, which requires the arguments of the verb and the verb itself in initial position. (16b) shows what the verb trace looks like for a transitive verb in initial position:

- (16) a. Verb in final position:
 $(s \backslash np[nom]) \backslash np[acc]$
 b. Verb trace for the analysis of verb-first:
 $((s \backslash ((s \backslash np[nom]) \backslash np[acc])) \backslash np[nom]) \backslash np[acc]$

The entry for the verb trace is very complex. It is much simpler to refer to the analysis in Figure 8.9.

<i>isst</i>	<i>er</i>	<i>ihn</i>	–
$(s \backslash np[nom]) \backslash np[acc]$	$np[nom]$	$np[acc]$	$((s \backslash (s \backslash np[nom]) \backslash np[acc]) \backslash np[nom]) \backslash np[acc]$
			$(s \backslash (s \backslash np[nom]) \backslash np[acc]) \backslash np[nom]$
			$s \backslash ((s \backslash np[nom]) \backslash np[acc])$
			s

Figure 8.9: Analysis of verb-initial sentences following Jacobs

The trace is the head in the entire analysis: It is first combined with the accusative object and then with the subject. In a final step, it is combined with the transitive verb in initial-position.⁶ A problem with this kind of analysis is that the verb *isst* (‘eats’), as well as *er* (‘he’) and *ihn* (‘him/it’), are arguments of the verb trace in (17).

- (17) Morgen [isst [er [ihn _]]]
 tomorrow eats he him
 ‘He will eat it tomorrow.’

Since adjuncts can occur before, after or between arguments of the verb in German, one would expect that *morgen* (‘tomorrow’) can occur before the verb *isst*. As adjuncts do not change the categorial status of a projection, the phrase *morgen isst er ihn* (‘tomorrow he eats him’) should be able to occur in the same positions as *isst er ihn*. This is not the case, however:

- (18) a. Deshalb isst er ihn.
 therefore eats he him
 b. *Deshalb morgen isst er ihn.
 therefore tomorrow eats he him

⁶ See Netter (1992) for a similar analysis in HPSG.

An approach, which avoids this problem, comes from Kiss & Wesche (1991) (see Section 9.3). Here, the authors assume that there is a verb in initial position, which selects the projection containing the verb trace. If adverbials are only combined with verbs in final-position, then a direct combination of *morgen* and *isst er ihn* is ruled out. If one assumes that the verb in first-position is the functor, then it is possible to capture the parallels between complementizers and verbs in initial position (Höhle 1997): Finite verbs in initial position differ from complementizers only by requiring the projection of a verb trace, whereas complementizers requires projections of overt verbs:

- (19) a. dass [er ihn isst]
 that he him eats
 b. Isst [er ihn _]
 eats he him

This description of verb position in German captures the central insights of the GB analysis in Section 3.2.

8.4 Local reordering

Up to now, we have seen combinations of functors and arguments where the arguments were either to the left or to the right of the functor. The saturation of arguments always took place in a fixed order: The argument furthest to the right was combined first with the functor, e.g. (s\NP)/PP first combined with the PP and the result of this combination was combined with the NP.

There are a number of possibilities for the analysis of ordering variants in German: Uszkoreit (1986b) suggests accounting for possible orders lexically, that is, that each possible order corresponds to a lexical entry. One would therefore have at least six lexical entries for a ditransitive verb. Briscoe (2000: 257) and Villavicencio (2002: 96–98) propose a variant of this analysis where the order of arguments is modified in the syntax: A syntactic rule can, for example, change the order (S/PRT)/NP into (S/NP)/PRT.

A different approach is suggested by Steedman & Baldridge (2006). They discuss various options of ordering arguments attested in the languages of the world. This includes languages in which the order of the combination is free, as well as languages where the direction of combination does not make a difference. Steedman and Baldridge introduce the following convention for representing categories: Elements in curly brackets can be discharged in any order. ‘|’ in place of ‘\’ or ‘/’ serves to indicate that the direction does not matter. Some prototypical examples are shown in (20):

- | | | | |
|------|----------|------------------------|--------------------------|
| (20) | English | (S\NP)/NP | S(VO) |
| | Latin | S{ NP[nom], NP[acc] } | free order |
| | Tagalog | S{/NP[nom], /NP[acc] } | free order, verb-initial |
| | Japanese | S{ NP[nom], \NP[acc] } | free order, verb-final |

Hoffman (1995: Section 3.1) has proposed an analysis analagous to that of Japanese for Turkish and this could also be used in conjunction with an analysis of verb position for

German. This would correspond to the GB/MP analysis of Fanselow (2001) or the HPSG analysis presented in Section 9.4.

8.5 Long-distance dependencies

Steedman (1989: Section 1.2.4) proposes an analysis of long-distance dependencies without movement or empty elements. For examples such as (21), he assumes that the category of *Harry must have been eating* or *Harry devours* is s/np .

- (21) a. These apples, Harry must have been eating.
b. apples which Harry devours

The fronted NP *these apples* and the relative pronoun *which* are both functors in the analysis of (21) which take s/np as their argument. Using the machinery introduced up to now, we cannot assign the strings in (21) the category s/np although it is intuitively the case that *Harry devours* is a sentence missing an NP. We still require two further extensions of Categorial Grammar: type raising and forward and backward composition. Both of these operations will be introduced in the following sections.

8.5.1 Type Raising

The category np can be transformed into the category $(s/(s\backslash np))$ by *type raising*. If we combine this category with $(s\backslash np)$, then we get the same result as if had combined np and $(s\backslash np)$ with the forward application rule in (2). (22a) shows the combination of an NP with a VP (a sentence missing an NP to its left). The combination of the NP with its raised type and the VP is given in (22b).

- (22) a. $np * s\backslash np = s$
b. $s/(s\backslash np) * s\backslash np = s$

In (22a), a verb or verb phrase selects an NP to its left ($s\backslash np$). In (22b), an NP having undergone type raising selects a verb or verb phrase to its right which in turn requires an NP to its left ($s\backslash np$).

Type raising simply reverses the direction of selection: The VP in (22a) is the functor and the NP is the argument, whereas in (22b), it is the type raised NP, which acts as the functor, and the VP is the argument. In each case, the resulting combination is the same. This change of selectional direction may just seem like a trick at first glance, but as we will see, this trick can be extremely useful. First, however, we will introduce forward and backward composition.

8.5.2 Forward and backward composition

(23) shows the rules for forward and backward composition.

- (23) a. Forward composition ($> B$)
 $X/Y * Y/Z = X/Z$
 b. Backward composition ($< B$)
 $Y \setminus Z * X \setminus Y = X \setminus Z$

These rules will be explained using forward composition as an example. (23a) can be understood as follows: X/Y more or less means; if I find a Y , then I am a complete X . In the combinatorial rule, X/Y is combined with Y/Z . Y/Z stands for a Y that is not complete yet but still missing a Z . The requirement that Y must find a Z in order to be complete is postponed: we pretend that Y was complete and use it anyway, but we remember that something is still missing. Hence, if we combine X/Y with Y/Z , we get something which becomes an X when combined with a Z .

8.5.3 Analysis of long-distance dependencies

By using forward composition, we can assign *Harry must have been eating* the category s/np . Figure 8.10 shows how this works. *must* is a verb which requires an unmarked

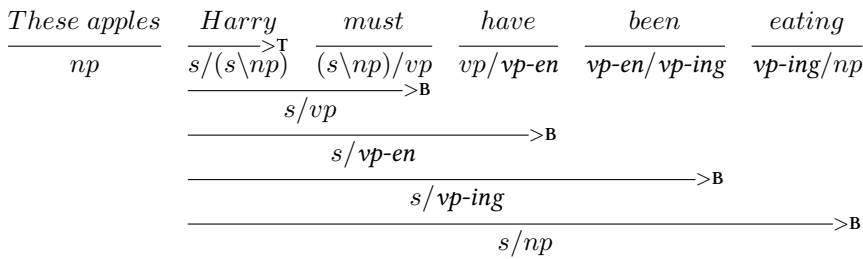


Figure 8.10: Application of forward composition to VP-chains

infinitive form, *have* requires a participial and *been* must combine with a gerund. In the above figure, the arrows with a small ‘T’ stand for type raising, whereas the arrows with a ‘B’ indicate composition. The direction of composition is indicated by the direction of the arrow.

For the analysis of (21a), we are still missing one small detail, a rule that turns the NP at the beginning of the sentence into a functor which can be combined with s/np . Normal type raising cannot handle this because this would produce $s/(s \setminus np)$ when $s/(s \setminus np)$ is required.

Steedman (1989: 217) suggests the rule in (24):

- (24) Topicalization (\uparrow):
 $X \Rightarrow st/(s/X)$
 where $X \in \{ NP, PP, VP, AP, S' \}$

st stands for a particular type of sentence, namely one with topicalization.

If we replace X with np , we can turn *these apples* into $st/(s/np)$ and complete the analysis of (21a) as shown in Figure 8.11 on the following page.

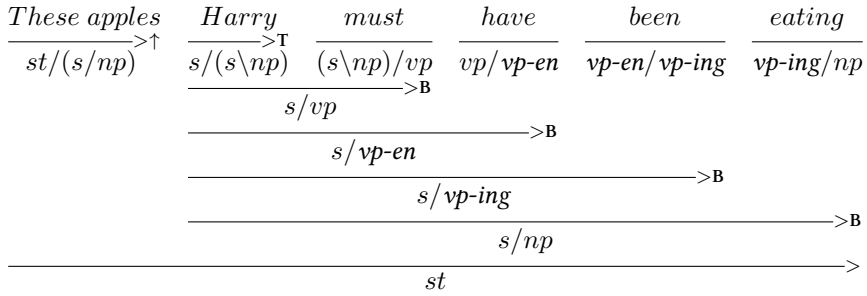


Figure 8.11: Analysis of long-distance dependencies in Categorical Grammar

The mechanism presented here will of course also work for dependencies that cross sentence boundaries. Figure 8.12 shows the analysis for (25):

(25) Apples, I believe that Harry eats.

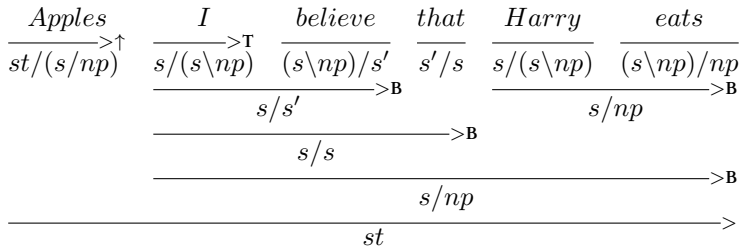


Figure 8.12: Analysis of long-distance dependencies across sentence boundaries

Using the previously described tools, it is, however, only possible to describe extractions where the fronted element in the sentence would have occurred at the right edge of the phrase without fronting. This means it is not possible to analyze sentences where the middle argument of a ditransitive verb has been extracted (Steedman 1985: 532). Pollard (1988: 406) provides the derivation in Figure 8.13 on the facing page for (26).

(26) Fido we put downstairs.

In this analysis, it is not possible to combine *we* and *put* using the rule in (23a) since $(s \backslash np)$ is not directly accessible: breaking down $((s \backslash np)/pp)/np$ into functor and argument gives us $((s \backslash np)/pp)$ and np . In order to deal with such cases, we need another variant of composition:

(27) Forward composition for $n=2$ ($> BB$)

$$X/Y * (Y/Z1)/Z2 = (X/Z1)/Z2$$

With this addition, it is now possible to combine the type-raised *we* with *put*. The result is $(s/pp)/np$. The topicalization rule in (24), however, requires an element to the right

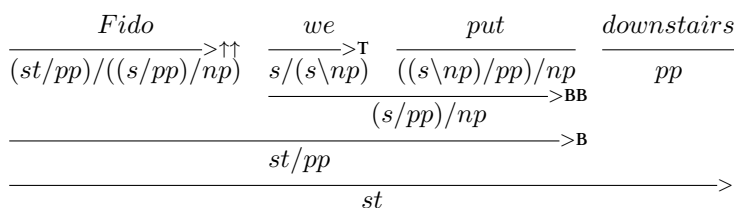


Figure 8.13: Analysis of long-distance dependencies across sentence boundaries

of *st* with the form (s/X) . This is not the case in Figure 8.13. For the NP *Fido*, we need a functor category which allows that the argument itself is complex. The rule which is needed for the case in (26) is given in (28).

- (28) Topicalization for $n=2$ ($\uparrow\uparrow$):
 $X2 \Rightarrow (st/X1)/((s/X1)/X2)$
 where $X1$ and $X2 \in \{ NP, PP, VP, AP, S' \}$

If we assume that verbs can have up to four arguments (z.B. *buy*: buyer, seller, goods, price), then it would be necessary to assume a further rule for composition as well as another topicalization rule. Furthermore, one requires a topicalization rule for subject extraction (Pollard 1988: 405). Steedman has developed a notation, which provides a compact notation of the previously discussed rules, but if one considers what exactly these representations stand for, one still arrives at the same number of rules that have been discussed here.

8.6 Summary and classification

The operations of Combinatory Categorical Grammar, which go beyond those of standard Categorical Grammar, allow for so much flexibility that it is even possible to assign a category to sequences of words that would not normally be treated as a constituent. This is certainly an advantage for the analysis of coordination (see Section 23.6.2) and furthermore, Steedman (1991) has argued that intonation data support the constituent status of these strings. See Section 15.2 for a direct model of incremental language processing in Categorical Grammar. In phrase structure grammars, it is possible to use GPSG mechanisms to pass information about relative pronouns contained in a phrase up the tree. These techniques are not used in CG and this leads to a large number of recategorization rules for topicalization and furthermore leads to inadequate analyses of pied-piping constructions in relative clauses. As the topicalization analysis was already discussed in Section 8.5, I will briefly go into relative clauses here.

Steedman & Baldridge (2006: 614) present an analysis of long-distance dependencies using the following relative clause in (29):

- (29) the man that Manny says Anna married

The relative pronoun is the object of *married* but occurs outside the clause with *Anna married*.

Steedman assumes the lexical entry in (30) for relative pronouns:

$$(30) \quad (n \backslash n) / (s / np)$$

This means the following: if there is a sentence missing an NP to the right of a relative pronoun, then the relative pronoun can form an N-modifier ($n \backslash n$) with this sentence. The relative pronoun is the head (functor) in this analysis.

Utilizing both additional operations of type raising and composition, the examples with relative clauses can be analyzed as shown in Figure 8.14. The lexical entry for the

$$\begin{array}{c}
 \frac{that}{(n \backslash n) / (s / np)} \quad \frac{Manny \quad says}{s / (s \backslash np)} \quad \frac{Anna \quad married}{s / (s \backslash np)} \\
 \frac{s / (s \backslash np)}{s / s} \quad \frac{s / (s \backslash np)}{s / np} \\
 \frac{s / s}{s / np} \\
 n \backslash n
 \end{array}$$

Figure 8.14: Categorical Grammar analysis of a relative clause with long-distance dependency

verbs corresponds to what was discussed in the previous chapter: *married* is a normal transitive verb and *says* is a verb that requires a sentential complement and forms a VP ($s \backslash np$) with it. This VP yields a sentence when combined with an NP. The noun phrases in Figure 8.14 have been type raised. Using forward composition, it is possible to combine *Anna* and *married* to yield s / np . This is the desired result: a sentence missing an NP to its right. *Manny* and *says* and then *Manny says* and *Anna married* can also be combined via forward composition and we then have the category s / np for *Manny says Anna married*. This category can be combined with the relative pronoun using forward application and we then arrive at $n \backslash n$, which is exactly the category for postnominal modifiers.

However, the assumption that the relative pronoun constitutes the head is problematic since one has to then go to some lengths to explain pied-piping constructions such as (31).

- (31) a. Here's the minister [[in [the middle [of [whose sermon]]]] the dog barked].⁷
 b. Reports [the height of the lettering on the covers of which] the government prescribes should be abolished.⁸

In (31), the relative pronoun is embedded in a phrase that has been extracted from the rest of the relative clause. The relative pronoun in (31a) is the determiner of *sermon*. Depending on the analysis, *whose* is the head of the phrase *whose sermon*. The NP is

⁷ Pollard & Sag (1994: 212).

⁸ Ross (1967: 109).

embedded under *of* and the phrase *of whose sermon* depends on *middle*. The entire NP *the middle of the sermon* is a complement of the preposition *in*. We would have to go a long way to be able to claim that *whose* is the head of the relative clause in (31a). The relative pronoun in (31b) is even more deeply embedded. Steedman (1997: 50) gives the following lexical entries for *who*, *whom* and *which*:

- (32) a. $((n \setminus n) / (s \setminus np)) \setminus (np / np)$ (complex subject-relative phrase)
 b. $((n \setminus n) / (s \setminus pp)) \setminus (pp / np)$ (complex extracted PP-relative phrase)
 c. $((n \setminus n) / (s \setminus np)) \setminus (np / np)$ (complex extracted NP-relative phrase)

Using (32b) and (32c), it is possible to analyze (33a) and (33b):

- (33) a. a report the cover of which Keats (expects that Chapman) will design
 b. a subject on which Keats (expects that Chapman) will speak

In the analysis of (33b), *which* requires a preposition to its left (*pp/np*) so it can form the category $(n \setminus n) / (s \setminus pp)$. This category needs a sentence lacking a PP to its right in order to form a post-nominal modifier ($n \setminus n$). In the analysis of (33a), *the cover of* becomes *np/np* by means of composition and *which* with the lexical entry (32c) can combine with *the cover of* to its left. The result is the category $(n \setminus n) / (s \setminus np)$, that is, something that requires a sentence missing a NP.

Ross' examples (31b) can also be analyzed as follows (32c):

- (34) Reports [the height of the lettering on the covers of]_{np/np} which]_{(n \setminus n) / (s \setminus np)} the government prescribes

The complex expression *the height of the lettering on the covers of* becomes *np/np* after composition and the rest of the analysis proceeds as that of (33a).

In addition to entries such as those in (32), we also need further entries to analyze sentences such as (35), where the relative phrase has been extracted from the middle of the clause (see Pollard (1988: 410)):

- (35) Fido is the dog which we put downstairs.

The problem here is similar to what we saw with topicalization: *we put* does not have the category *s/np* but rather *(s/pp)/np* and as such, cannot be directly combined with the relative pronoun in (30).

Morrill (1995: 204) discusses the lexical entry in (32b) for the relative pronoun in (36):

- (36) about which John talked

In the lexical entry (32b), *which* requires something to the left of it, which requires a noun phrase in order to form a complete prepositional phrase, that is, *which* selects a preposition. Morrill saw that there is a need to postulate further entries for cases where the relative pronoun occurs in the middle of the fronted phrase.

- (37) the contract [the loss of which after so much wrangling] John would finally have to pay for

These and other cases can be dealt with by additional lexical stipulations. Morrill instead proposes additional types of combination of functors and arguments, which allow a functor $B \uparrow A$ to enclose its argument A and produce B , or a functor $A \downarrow B$ to enclose its argument to then yield B (p. 190). Even with these additional operations, he still needs the two lexical entries in (38) for the derivation of a pied-piping construction with an argument NP or a PP:

- (38) a. $(NP \uparrow NP) \downarrow (N \backslash N)/(S/NP)$
 b. $(PP \uparrow NP) \downarrow (N \backslash N)/(S/PP)$

These lexical entries are still not enough, however, as (38b) contains a PP but this PP corresponds to an argument PP, which is required for (36). To analyze (31a), we need to assume the category $(s \backslash np)/(s \backslash np)$ for the prepositional phrase. We therefore also require at least three additional entries for relative pronouns.

By introducing new operations, Morrill manages to reduce the number of lexical entries for *which*, however, the fact remains that he has to mention the categories which can occur in pied-piping constructions in the lexical entry of the relative pronoun.

Furthermore, the observation that relative clauses consist of a phrase with a relative pronoun plus a sentence missing a relative phrase is lost. This insight can be kept if one assumes a GPSG-style analysis where information about whether there is a relative pronoun in the relative phrase can be passed up to the highest node of the relative phrase. The relative clause can then be analyzed as the combination of a sentence with a gap and a correspondingly marked relative phrase. For the discussion of such analyses in the framework of GB theory and HPSG/CxG, see Section 23.10.3.

Comprehension questions

1. Identify the functors and arguments in Figures 8.1 and 8.3.
2. Which combinatorial possibilities do you know?
3. What is composition used for?

Exercises

1. Analyze the following sentence:

(39) The children in the room laugh loudly.

2. Analyze the noun phrase in (40):

(40) the picture of Mary

Compare your resulting analysis with the structure given in Figure 2.4 on page 69 and think about which categories of \bar{X} syntax the categories in Categorial Grammar correspond to.

Further reading

Mark Steedman discusses a variant of Categorical Grammar, *Combinatory Categorical Grammar*, in a series of books and articles: Steedman (1991; 2000); Steedman & Baldridge (2006).

Lobin (2003) compares Categorical Grammar with Dependency Grammar and Pickering & Barry (1993) suggest a combination of Dependency Grammar and Categorical Grammar, which they call Dependency Categorical Grammar.

Briscoe (2000) and Villavicencio (2002) discuss UG-based acquisition models in the framework of Categorical Grammar.

9 Head-Driven Phrase Structure Grammar

Head-Driven Phrase Structure Grammar (HPSG) was developed by Carl Pollard and Ivan Sag in the mid-80's in Stanford and in the Hewlett Packard research laboratories in Palo Alto (Pollard & Sag 1987; 1994). Like LFG, HPSG is part of so-called West Coast linguistics. Another similarity to LFG is that HPSG aims to provide a theory of competence which is compatible with performance (Sag & Wasow (2011), also see Chapter 15).

The formal properties of the description language for HPSG grammars are well-understood and there are many systems for processing such grammars (Dörre & Seiffert 1991; Dörre & Dorna 1993; Popowich & Vogel 1991; Uszkoreit et al. 1994; Erbach 1995; Schütz 1996; Schmidt et al. 1996b;a; Uszkoreit et al. 1996; Müller 1996c; 2004c; Carpenter & Penn 1996; Penn & Carpenter 1999; Götz et al. 1997; Copestake 2002; Callmeier 2000; Dahllöf 2003; Meurers et al. 2002; Penn 2004; Müller 2007a; Sato 2008; Kaufmann 2009).¹ Currently, the LKB system by Ann Copestake and the TRALE system, that was developed by Gerald Penn (Meurers, Penn & Richter 2002; Penn 2004), have the most users. The DELPH-IN-Konsortium and various TRALE users developed many small and some large fragments of various languages. The following is a list of implementations in different systems:

- Arabic (Haddar et al. 2010; Hahn 2011; Masum et al. 2012; Boukedi & Haddar 2014),
- Bengali (Paul 2004; Islam et al. 2012),
- Bulgarian (Simov, Osenova, Simov & Kouylekov 2004; Osenova 2010a;b; 2011),
- Danish (Ørsnes 1995; 2009b; Neville & Paggio 2004; Müller 2009c; Müller & Ørsnes 2011; Müller 2012; Müller & Ørsnes 2015),
- German (Kiss 1991; Netter 1993; 1996; Meurers 1994; Hinrichs et al. 1997; Kordoni 1999; Tseng 2000; Geißler & Kiss 1994; Keller 1994; Müller 1996c; 1999a; Müller & Kasper 2000; Crysmann 2003; 2005b;c; Müller 2007b; Kaufmann & Pfister 2007; 2008; Kaufmann 2009; Fokkens 2011),
- English (Copestake & Flickinger 2000; Flickinger, Copestake & Sag 2000; Flickinger 2000; Dahllöf 2002; 2003; De Kuthy & Meurers 2003a; Meurers, De Kuthy & Metcalf 2003; De Kuthy, Metcalf & Meurers 2004),

¹ Uszkoreit et al. (1996); Bolc et al. (1996) compare systems that were available or were developed at the beginnings of the 1990s. Melnik (2007) compares LKB and TRALE. See also Müller (2015a)

- Esperanto (Li 1996),
- French (Tseng 2003),
- Ga (Kropp Dakubu, Hellan & Beermann 2007; Hellan 2007),
- Sign Language (German, French, British, Greek) (Sáfár & Marshall 2002; Marshall & Sáfár 2004; Sáfár & Glauert 2010),
- Georgianh (Abzianidze 2011),
- Greek (Kordoni & Neu 2005),
- Hausa (Crysmann 2005a; 2009; 2011; 2012),
- Hebrew (Melnik 2007),
- Japanisch (Siegel 2000; Siegel & Bender 2002; Bender & Siegel 2005),
- Yiddish (Müller & Ørsnes 2011),
- Korean (Kim & Yang 2003; 2004; 2006; 2009; Kim, Sells & Yang 2007; Song, Kim, Bond & Yang 2010; Kim, Yang, Song & Bond 2011),
- Maltese (Müller 2009b),
- Mandarin Chinese (Liu 1997; Ng 1997; Müller & Lipenkova 2009; 2013),
- Dutch (van Noord & Bouma 1994; Bouma, van Noord & Malouf 2001b; Fokkens 2011),
- Norwegian (Hellan & Haugereid 2003; Beermann & Hellan 2004; Hellan & Beermann 2006),
- Persian (Müller 2010b; Müller & Ghayoomi 2010),
- Polish (Przepiórkowski, Kupść, Marciniak & Mykowiecka 2002; Mykowiecka, Marciniak, Przepiórkowski & Kupść 2003),
- Portuguese (Branco & Costa 2008a;b; Costa & Branco 2010),
- Russian (Avgustinova & Zhang 2009),
- Sahaptin (Drellishak 2009),
- Spanish (Pineda & Meza 2005a;b; Bildhauer 2008; Marimon 2013),
- Turkish (Fokkens, Poulson & Bender 2009),
- Wambaya (Bender 2008a;c; 2010).

The first implemented HPSG grammar was a grammar of English that was developed in the Hewlet-Packard labs in Palo Alto (Flickinger, Pollard & Wasow 1985; Flickinger 1987). Grammars for German were developed in Heidelberg, Stuttgart and Saarbrücken in the LILOG project. After this grammars for German, English and Japanese were developed in Heidelberg, Saarbrücken and Stanford in the *Verbmobil* project. *Verbmobil* was the largest AI project in Germany ever. It was a machine translation project for spoken language in the domains of trip planning and appointment scheduling (Wahlster 2000).

Currently there are two larger groups that are working on grammar development: the DELPH-IN consortium (Deep Linguistic Processing with HPSG)² and the network CoGETI (Constraintbasierte Grammatik: Empirie, Theorie und Implementierung)³. Many of the grammar fragments that are listed above were developed by members of DELPH-IN and some were derived from the Grammar Matrix which was developed for the LKB to provide grammar writers with a typologically motivated initial grammar that corresponds to the properties of the language under development (Bender, Flickinger & Oepen 2002). The CoreGram project⁴ is a similar project that is run at the Freie Universität Berlin. It is developing grammars for German, Danish, Persian, Maltese, Mandarin Chinese, Spanish, French and Yiddish that share a common core. Constraints that hold for all languages are represented in one place and used by all grammars. Furthermore there are constraints that hold for certain language classes and again they are represented together and used by the respective grammars. So while the Grammar Matrix is used to derive grammars that individual grammar writers can use, adapt and modify to suit their needs, CoreGram really develops grammars for various languages that are used simultaneously and have to stay in sync. A description of the CoreGram can be found in Müller (2013a; 2015a).

There are systems that combine linguistically motivated analyses with statistics components (Brew 1995; Miyao et al. 2005; Miyao & Tsujii 2008) or learn grammars or lexicons from corpora (Fouvy 2003; Cramer & Zhang 2009).

The following URLs point to pages on which grammars can be tested:

- <http://www.delph-in.net/erg/>
- <http://hpsg.fu-berlin.de/Demos/>

9.1 General remarks on the representational format

HPSG has the following characteristics: It a lexically-based theory, that is, the majority of linguistic constraints are situated in the descriptions of words or roots. HPSG is sign-based in the sense of Saussure (1916b): The form and meaning of linguistic signs are always represented together. Typed feature structures are used to model all relevant information. These structures can be described with feature descriptions such as in (1). Lexical entries, phrases and principles are always modeled and described with the same

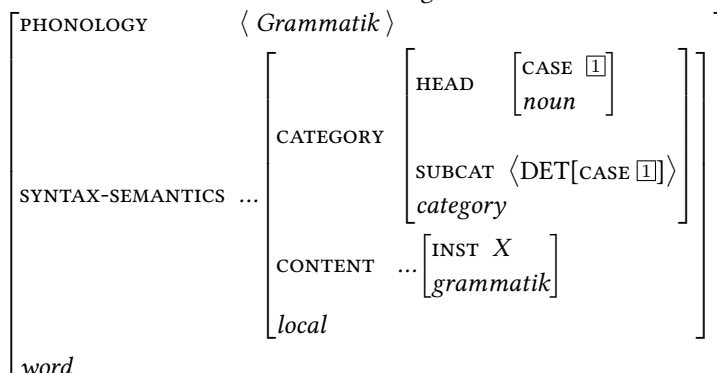
² <http://www.delph-in.net/>. 25.03.2013.

³ <http://www.user.gwdg.de/~cogeti/>. 25.03.2013. Supported by the DFG under the grant number HO3279/3-1.

⁴ <http://hpsg.fu-berlin.de/Projects/CoreGram.html>. 20.05.2015.

formal means. Generalizations about word classes or rule schemata are captured with inheritance hierarchies. Phonology, syntax and semantics are represented in a single structure. There are no separate levels of representation such as PF or LF in Government & Binding Theory. (1) shows an excerpt from the representation of a word such as *Grammatik* ‘grammar’.

- (1) Lexical item for the word *Grammatik* ‘grammar’:



One can see that this feature description contains information about the phonology, syntactic category and semantic content of the word *Grammatik*. To keep things simple, the value of PHONOLOGY (PHON) is mostly given as orthographic representation. In fully spelled-out theories, the PHON value is a complex structure that contains information about metrical grids and weak or strong accents. See Bird & Klein (1994), Orgun (1996), Höhle (1999), Walther (1999), Crysmann (2002: Chapter 6), and Bildhauer (2008) for phonology in the framework of HPSG. The details of the description in (1) will be explained in the following sections.

HPSG has adopted various insights from other theories and newer analyses have been influenced by developments in other theoretical frameworks. Functor-argument structures, the treatment of valency information and functional composition have been adopted from Categorical Grammar. Functional composition plays an important role in the analysis of verbal complexes. The ID/LP format as well as the Slash mechanism for long-distance dependencies both come from GPSG. The analysis assumed here for verb position in German was developed in the framework of Government & Binding.

9.1.1 Representation of valency information

The phrase structure grammars discussed in Chapter 2 have the disadvantage that one requires a great number of different rules for the various valency types. (2) shows some examples of this kind of rules and the corresponding verbs.

- | | | |
|-----|--|---|
| (2) | $S \rightarrow \text{NP}[\textit{nom}], V$ | $X \textit{schläft}$ ‘X is sleeping’ |
| | $S \rightarrow \text{NP}[\textit{nom}], \text{NP}[\textit{acc}], V$ | $X Y \textit{erwartet}$ ‘X expects Y’ |
| | $S \rightarrow \text{NP}[\textit{nom}], \text{PP}[\textit{über}], V$ | $X \textit{über} Y \textit{spricht}$ ‘X talks about Y’ |
| | $S \rightarrow \text{NP}[\textit{nom}], \text{NP}[\textit{dat}], \text{NP}[\textit{acc}], V$ | $X Y Z \textit{gibt}$ ‘X gives Y to Z’ |
| | $S \rightarrow \text{NP}[\textit{nom}], \text{NP}[\textit{dat}], \text{PP}[\textit{mit}], V$ | $X Y \textit{mit} Z \textit{dient}$ ‘X serves Y with Z’ |

In order for the grammar not to create any incorrect sentences, one has to ensure that verbs are only used with appropriate rules.

- (3)
- a. * dass Peter das Buch schläft
that Peter the book sleeps
 - b. * dass Peter erwartet
that Peter expects
 - c. * dass Peter über den Mann erwartet
that Peter about the man expects

Therefore, verbs (and heads in general) have to be divided into valency classes. These valency classes have to then be assigned to grammatical rules. One has to therefore further specify the rule for transitive verbs in (2) as follows:

- (4) $S \rightarrow \text{NP}[\textit{nom}], \text{NP}[\textit{acc}], V[\textit{nom_acc}]$

Here, the valency has been encoded twice. First, we have said something in the rules about what kind of elements can or must occur, and then we have stated in the lexicon which valency class the verb belongs to. In Section 5.5, it was pointed out that morphological processes need to refer to valency information. Hence, it is desirable to remove redundant valency information from the grammatical rules. For this reason, HPSG – like Categorical Grammar – adopts descriptions of the arguments of a head in the lexical entry of that head. There is a feature with a list-value, the SUBCAT feature, which contains descriptions of the objects that must combine with a head in order to yield a complete phrase. (5) gives some examples for the verbs in (2):

- | | | |
|-----|-----------------------------|---|
| (5) | Verb | SUBCAT |
| | <i>schlafen</i> ‘to sleep’ | $\langle \text{NP}[\textit{nom}] \rangle$ |
| | <i>erwarten</i> ‘to expect’ | $\langle \text{NP}[\textit{nom}], \text{NP}[\textit{acc}] \rangle$ |
| | <i>sprechen</i> ‘to speak’ | $\langle \text{NP}[\textit{nom}], \text{PP}[\textit{über}] \rangle$ |
| | <i>geben</i> ‘to give’ | $\langle \text{NP}[\textit{nom}], \text{NP}[\textit{dat}], \text{NP}[\textit{acc}] \rangle$ |
| | <i>dienen</i> ‘to serve’ | $\langle \text{NP}[\textit{nom}], \text{NP}[\textit{dat}], \text{PP}[\textit{mit}] \rangle$ |

SUBCAT is an abbreviation for subcategorization. It is often said that a head subcategorizes for certain arguments. See page 93 for more on the term *subcategorization*.

Figure 9.1 shows the analysis for (6a) and the analysis for (6b) is in Figure 9.2 on the facing page:

- (6) a. [dass] Peter schläft
 that Peter sleeps
 b. [dass] Peter Maria erwartet
 that Peter Maria expects
 ‘that Peter expects Maria’

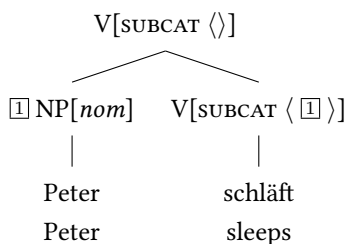


Figure 9.1: Analysis of *Peter schläft* ‘Peter sleeps’ in *dass Peter schläft* ‘that Peter sleeps’

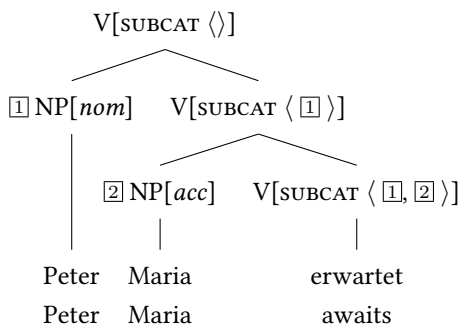
In Figures 9.1 and 9.2, an element of the SUBCAT list is combined with its head in each local tree. The elements that are combined with the selecting head are then no longer present in the SUBCAT list of the mother node. $V[\text{SUBCAT } \langle \rangle]$ corresponds to a complete phrase (VP or S). The boxes with numbers show the structure sharing (see Section 6.4). Structure sharing is the most important means of expression in HPSG. It plays a central role for phenomena such as valency, agreement and long-distance dependencies. In the examples above, [1] indicates that the description in the SUBCAT list is identical to another daughter in the tree. The descriptions contained in valence lists are usually partial descriptions, that is, not all properties of the argument are exhaustively described. Therefore, it is possible that a verb such as *schläft* ‘sleeps’ can be combined with various kinds of linguistic objects: The subject can be a pronoun, a proper name or a complex noun phrase, it only matters that the linguistic object in question has an empty SUBCAT list and bears the correct case.⁵

9.1.2 Representation of constituent structure

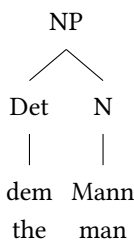
As already noted, feature descriptions in HPSG serve as a single descriptive inventory for morphological rules, lexical entries and syntactic rules. The trees we have seen thus far are only visualizations of the relations and do not have any theoretical status. There are also no rewrite rules in HPSG.⁶ The job of phrase structure rules is handled by feature descriptions. Information about dominance is represented using DTR-features (head

⁵ Furthermore, it must agree with the verb. This is not shown here.

⁶ However, phrase structure rules are used in some computer implementations of HPSG in order to improve the efficiency of processing.

Figure 9.2: Analysis of *Peter Maria erwartet* ‘Peter awaits Maria’

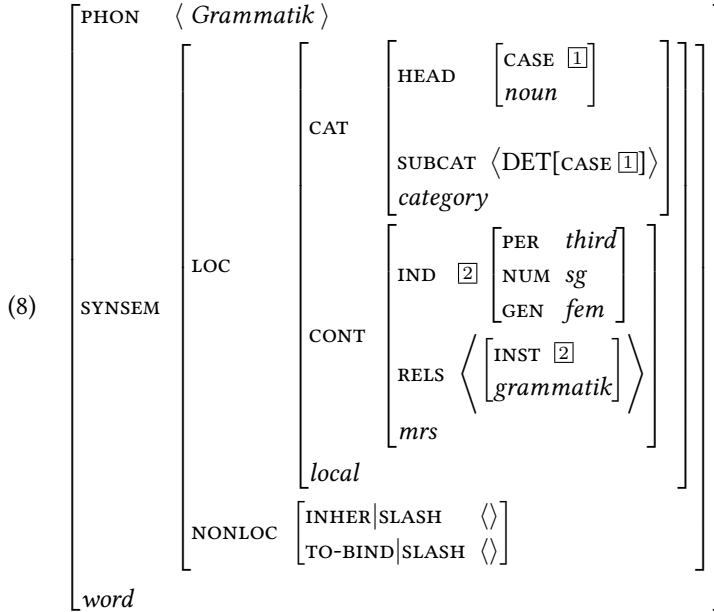
daughter and non-head daughter), information about precedence is implicitly contained in PHON. (7) shows the representation of PHON values in a feature description corresponding to the tree in Figure 9.3.

Figure 9.3: *dem Mann* ‘the man’

$$(7) \left[\begin{array}{l} \text{PHON} \\ \text{HEAD-DTR} \\ \text{NON-HEAD-DTRS} \end{array} \left[\begin{array}{l} \langle \text{dem Mann} \rangle \\ \left[\text{PHON} \langle \text{Mann} \rangle \right] \\ \left\langle \left[\text{PHON} \langle \text{dem} \rangle \right] \right\rangle \end{array} \right] \right]$$

In (7), there is exactly one head daughter (HEAD-DTR). The head daughter is always the daughter containing the head. In a structure with the daughters *das* and *Bild von Maria*, the latter would be the head daughter. In principle, there can be multiple non-head daughters. If we were to assume a flat structure for a sentence with a ditransitive verb, as in Figure 2.1 on page 56, we would have three non-head daughters. It also makes sense to assume binary branching structures without heads (see Müller (2007b: Chapter 11) for an analysis of relative clauses).

Before it is shown how it is ensured that only those head-argument structures are licensed in which the argument matches the requirements of the head, I will present the general structure of feature descriptions in HPSG. The structure presented at the start of this chapter is repeated in (8) with all the details relevant to the present discussion:



In the outer layer, there are the features PHON and SYNSEM. As previously mentioned, PHON contains the phonological representation of a linguistic object. The value of SYNSEM is a feature structure which contains syntactic and semantic information that can be selected by other heads. The daughters of phrasal signs are represented outside of SYNSEM. This ensures that there is a certain degree of locality involved in selection: A head cannot look into its daughters (Pollard und Sag 1987: 143–145; 1994: 23). Inside SYNSEM, there is information relevant in local contexts (LOCAL, abbreviated to LOC) as well as information important for long-distance dependencies (NONLOCAL or NONLOC for short). Locally relevant information includes syntactic (CATEGORY or CAT), and semantic (CONTENT or CONT) information. Syntactic information includes information that determines the central characteristics of a phrase, that is, the head information. This is represented under HEAD. Further details of this will be discussed in Section 9.1.4. Among other things, the part of speech of a linguistic object belongs to the head properties of a phrase. As well as HEAD, SUBCAT belongs to the information contained inside CAT. The semantic content of a sign is present under CONT. The type of the CONT value is *mrs*, which stands for *Minimal Recursion Semantics* (Copestake, Flickinger, Pollard & Sag 2005). An MRS structure is comprised of an index and a list of relations which restrict this index. Of the NONLOCAL features, only SLASH is given here. There are further features for dealing with relative and interrogative clauses (Pollard & Sag 1994; Sag 1997; Ginzburg & Sag 2000;

Holler 2005), which will not be discussed here.

As can be seen, the description of the word *Grammatik* ‘grammar’ becomes relatively complicated. In theory, it would be possible to list all properties of a given object directly in a single list. This would, however, have the downside that the identity of groups of feature-value pairs could not be as easily expressed. Using the feature geometry (8), it is possible to express the fact that the CAT values of both conjuncts in symmetric coordinations such as those in (9) are identical.

- (9) a. [der Mann] und [die Frau]
 the man and the woman
 b. Er [kennt] und [liebt] diese Schallplatte.
 he knows and loves this record
 c. Er ist [dumm] und [arrogant].
 he is dumb and arrogant

If valency and the part of speech information were not represented in the sub-structure, we would have to state separately that utterances such as (9) require that both conjuncts have the same valency and part of speech.

After this general introduction to feature geometry, we can now turn to the head-argument schema:

Schema 1 (Head-Argument Schema (binary branching, preliminary version))

head-argument-phrase \Rightarrow

$$\left[\begin{array}{l} \text{SYNSEM|LOC|CAT|SUBCAT } \boxed{1} \\ \text{HEAD-DTR|SYNSEM|LOC|CAT|SUBCAT } \boxed{1} \oplus \langle \boxed{2} \rangle \\ \text{NON-HEAD-DTRS } \langle [\text{SYNSEM } \boxed{2}] \rangle \end{array} \right]$$

Schema 1 states the properties a linguistic object of the type *head-argument-phrase* must have. The arrow in Schema 1 stands for a logical implication and not for the arrow of rewrite rules as we know it from phrase structure grammars. ‘ \oplus ’ (*append*) is a relation which combines two lists. (10) shows a possible division of a list that contains two elements:

$$(10) \quad \langle x, y \rangle = \langle x \rangle \oplus \langle y \rangle \text{ or } \\ \langle \rangle \oplus \langle x, y \rangle \text{ or } \\ \langle x, y \rangle \oplus \langle \rangle$$

The list $\langle x, y \rangle$ can be subdivided into two lists each containing one element, or alternatively into an empty list and $\langle x, y \rangle$.

Schema 1 can be read as follows: If an object is of the type *head-argument-phrase* then it must have the properties on the right-hand side of the implication. In concrete terms, this means that these objects always have a valence list which corresponds to $\boxed{1}$, that they have a head daughter with a valence list that can be divided into two sublists $\boxed{1}$ and $\langle \boxed{2} \rangle$ and also that they have a non-head daughter whose syntactic and semantic

properties (SYNSEM value) is compatible with the last element of the SUBCAT list of the head daughter (2). (11) shows the corresponding feature description for the example in (6a).

$$(11) \left[\begin{array}{l} \text{PHON} \langle \textit{Peter schläft} \rangle \\ \text{SYNSEM} | \text{LOC} | \text{CAT} | \text{SUBCAT} \langle \rangle \\ \text{HEAD-DTR} \left[\begin{array}{l} \text{PHON} \langle \textit{schläft} \rangle \\ \text{SYNSEM} | \text{LOC} | \text{CAT} | \text{SUBCAT} \langle \boxed{1} \text{ NP}[\textit{nom}] \rangle \end{array} \right] \\ \text{NON-HEAD-DTRS} \left\langle \left[\begin{array}{l} \text{PHON} \langle \textit{Peter} \rangle \\ \text{SYNSEM} \boxed{1} \end{array} \right] \right\rangle \\ \textit{head-argument-phrase} \end{array} \right]$$

NP[*nom*] is an abbreviation for a complex feature description. Schema 1 divides the SUBCAT list of the head daughter into a single-element list and the remainder. Since *schläft* ‘sleeps’ only has one element in its SUBCAT list, the remainder is an empty list. The remainder is also the SUBCAT value of the mother.

9.1.3 Linearization rules

Dominance schemata do not say anything about the order of the daughters. As in GPSG, linearization rules are specified separately. Linearization rules can make reference to the properties of daughters, their function in a schema (head, argument, adjunct, ...) or both. If we assume a feature INITIAL for all heads, then heads which precede their arguments would have the INITIAL value ‘+’ and heads following their arguments would have the value ‘−’. The linearization rules in (12) ensure that ungrammatical orderings such as (13b,d) are ruled out.⁷

- (12) a. Head[INITIAL +] < Argument
b. Argument < Head[INITIAL −]

Prepositions have an INITIAL value ‘+’ and therefore have to precede arguments. Verbs in final position bear the value ‘−’ and have to follow their arguments.

- (13) a. [in [den Schrank]]
in the cupboard
b. * [[den Schrank] in]
the cupboard in
c. dass [er [ihn umfüllt]]
that he it refills

⁷ Noun phrases pose a problem for (12): Determiners have been treated as argument until now and were included in the SUBCAT list of the head noun. Determiners occur to the left of noun, whereas all other arguments of the noun are to the right. This problem can be solved either by refining linearization rules (Müller 1999a: 164–165) or by introducing a special valence feature for determiners (Pollard & Sag 1994: Section 9.4). For an approach using such a feature, see Section 9.6.

- d. * dass [er [umfüllt ihn]]
that he refills him

9.1.4 Projection of head properties

The verb form belongs to the features that are important for the distribution of verbal projections. Certain verbs require a verbal argument with a particular form:

- (14) a. [Dem Mann helfen] will er nicht.
the man help wants he not
'He doesn't want to help the man.'
- b. [Dem Mann geholfen] hat er nicht.
the man helped has he not
'He hasn't helped the man.'
- c. * [Dem Mann geholfen] will er nicht.
the man helped wants he not
- d. * [Dem Mann helfen] hat er nicht.
the man help has he not

wollen 'to want' always needs an infinitive without *zu* 'to', *haben* 'have' on the other hand requires a verb in participle form. *glauben* 'believe' can occur with a finite clause, but not with an infinitive without *zu*:

- (15) a. Ich glaube, Peter kommt morgen.
I believe Peter comes tomorrow
'I think Peter is coming tomorrow.'
- b. * Ich glaube, Peter morgen kommen.
I believe Peter tomorrow come
- c. * Ich glaube, morgen kommen.
I believe tomorrow come

As well as information about the part of speech, projections of verbs must also contain information about the verb form. Figure 9.4 on the next page shows this on the basis of the finite verb *gibt* 'gives'.

GPSG has the Head Feature Convention that ensures that head features on the mother node are identical to those on the node of the head daughter. In HPSG, there is a similar principle. Unlike GPSG, head features are explicitly contained as a group of features in the feature structures. They are listed under the path `SYNSEM|LOC|CAT|HEAD`. (16) shows the lexical entry for *gibt* 'gives':

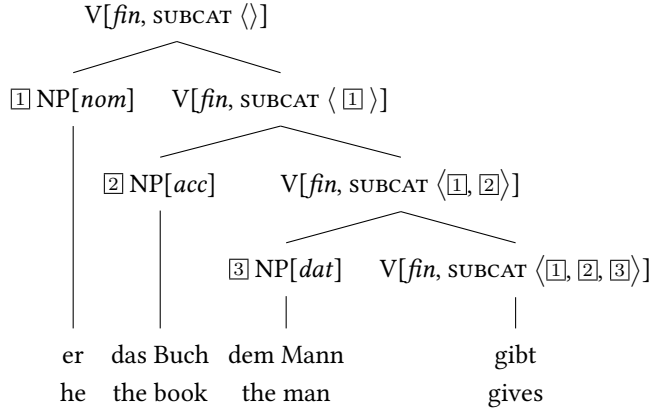


Figure 9.4: Projection of the head features of the verb

- (16) *gibt* ‘gives’:
- | | | | | | | | | | |
|------------------------------------|--|---|---|------------|-------------|-------------|--|--------|--|
| PHON $\langle \text{gibt} \rangle$ | <table style="border-collapse: collapse; width: 100%;"> <tr> <td style="padding: 5px;">HEAD</td> <td style="border-left: 1px solid black; padding: 5px; vertical-align: top;"> <table style="border-collapse: collapse;"> <tr><td>VFORM</td><td><i>fin</i></td></tr> <tr><td><i>verb</i></td><td></td></tr> </table> </td> </tr> <tr> <td style="padding: 5px;">SUBCAT</td> <td style="border-left: 1px solid black; padding: 5px; vertical-align: top;"> $\langle \text{NP}[nom], \text{NP}[acc], \text{NP}[dat] \rangle$ </td> </tr> </table> | HEAD | <table style="border-collapse: collapse;"> <tr><td>VFORM</td><td><i>fin</i></td></tr> <tr><td><i>verb</i></td><td></td></tr> </table> | VFORM | <i>fin</i> | <i>verb</i> | | SUBCAT | $\langle \text{NP}[nom], \text{NP}[acc], \text{NP}[dat] \rangle$ |
| HEAD | | <table style="border-collapse: collapse;"> <tr><td>VFORM</td><td><i>fin</i></td></tr> <tr><td><i>verb</i></td><td></td></tr> </table> | VFORM | <i>fin</i> | <i>verb</i> | | | | |
| VFORM | | <i>fin</i> | | | | | | | |
| <i>verb</i> | | | | | | | | | |
| SUBCAT | $\langle \text{NP}[nom], \text{NP}[acc], \text{NP}[dat] \rangle$ | | | | | | | | |
| SYNSEM LOC CAT | | | | | | | | | |
| <i>word</i> | | | | | | | | | |

The *Head Feature Principle* takes the following form:

Principle 4 (Head Feature Principle)

The HEAD value of any headed phrase is structure-shared with the HEAD value of the head daughter.

Figure 9.5 on the facing page contains a variant of Figure 9.4 with the structure sharing made explicit.

The following section will deal with how this principle is formalized as well as how it can be integrated into the architecture of HPSG.

9.1.5 Inheritance hierarchies and generalizations

Up to now, we have seen one example of a dominance schema and more will follow in the coming sections, e. g. schemata for head-adjunct structures as well as for the resolution of long-distance dependencies. The Head Feature Principle is a general principle which must be met by all structures licensed by these schemata. As mention above, it must be met by all structures with a head. Formally, this can be captured by categorizing syntactic structures into those with and those without heads and assign the type

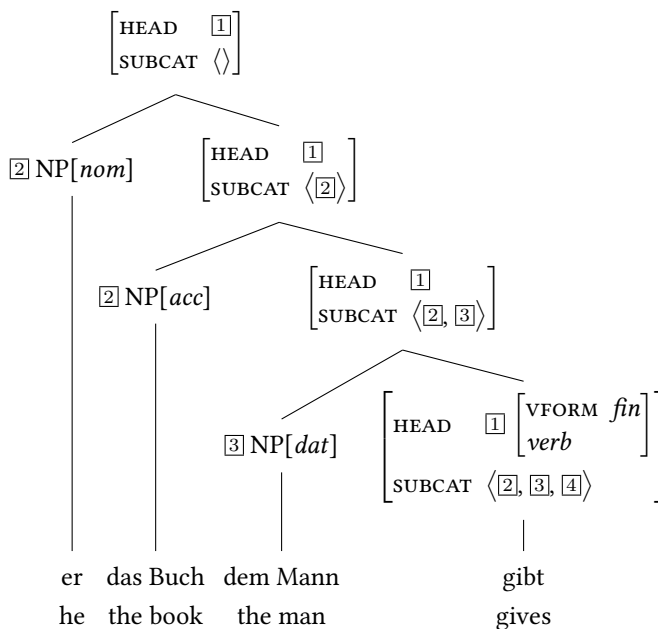


Figure 9.5: Projection of head features of a verb with structure sharing

headed-phrase to those with a head. The type *head-argument-phrase* – the type which the description in Schema 1 on page 267 has – is a subtype of *headed-phrase*. Objects of a certain type always have the same properties that objects of a supertype have. Recall the example from Section 6.2: An object of the type *female person* has all the properties of the type *person*. Furthermore, objects of type *female person* have additional, more specific properties not shared with other subtypes of *person*.

If one formulates a restriction on a supertype, this automatically affects all of its subtypes. The Head Feature Principle hence can be formalized as follows:

$$(17) \text{ headed-phrase} \Rightarrow \left[\begin{array}{l} \text{SYNSEM|LOC|CAT|HEAD } \boxed{1} \\ \text{HEAD-DTR|SYNSEM|LOC|CAT|HEAD } \boxed{1} \end{array} \right]$$

The arrow is a logical implication. (17) can be read as follows: If a structure is of type *headed-phrase*, then it must hold that the value of `SYNSEM|LOC|CAT|HEAD` is identical to the value of `HEAD-DTR|SYNSEM|LOC|CAT|HEAD`.

An extract from the type hierarchy under *sign* is given in Figure 9.6 on the next page. *word* and *phrase* are subclasses of linguistic signs. Phrases can be divided into phrases with heads (*headed-phrase*) and those without (*non-headed-phrase*). There are also subtypes for phrases of type *non-headed-phrase* and *headed-phrase*. We have already discussed *head-argument-phrase*, and other subtypes of *headed-phrase* will be discussed in the later sections. As well as *word* and *phrase*, there are the types *root* and *stem*, which

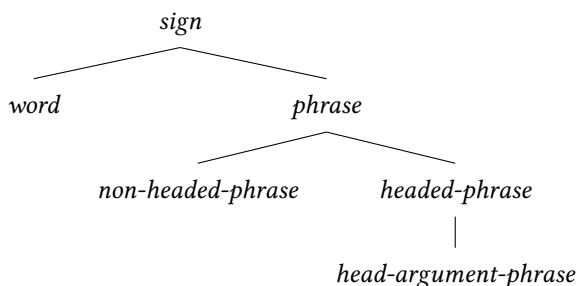


Figure 9.6: Type hierarchy for *sign*: all subtypes of *headed-phrase* inherit constraints

play an important role for the structure of the lexicon and the morphological component. Due to space considerations, it is not possible to further discuss these types here, but see Chapter 24.

The description in (18) shows the head-argument schema from page 267 together with the restrictions that the type *head-argument-phrase* inherits from *headed-phrase*.

(18) Head-Argument-Schema + Head Feature Principle:

$$\left[\begin{array}{l} \text{SYNSEM|LOC|CAT} \left[\begin{array}{l} \text{HEAD} \quad \boxed{1} \\ \text{SUBCAT} \quad \boxed{2} \end{array} \right] \\ \\ \text{HEAD-DTR|SYNSEM|LOC|CAT} \left[\begin{array}{l} \text{HEAD} \quad \boxed{1} \\ \text{SUBCAT} \quad \boxed{2} \oplus \langle \boxed{3} \rangle \end{array} \right] \\ \\ \text{NON-HEAD-DTRS} \langle [\text{SYNSEM} \boxed{3}] \rangle \\ \text{head-argument-phrase} \end{array} \right]$$

(19) gives a description of a structure that is licensed by Schema 1. As well as valency information, the head information is specified in (19) and it is also apparent how the Head Feature Principle forces the projection of features: The head value of the entire structure ($\boxed{1}$) corresponds to the head value of the verb *gibt* ‘gives’.

$$\begin{array}{c}
 (19) \quad \left[\begin{array}{l}
 \text{PHON } \langle \text{dem Mann gibt} \rangle \\
 \text{SYNSEM|LOC|CAT} \left[\begin{array}{l}
 \text{HEAD } \boxed{1} \\
 \text{SUBCAT } \boxed{2} \langle \text{NP}[\text{nom}], \text{NP}[\text{acc}] \rangle
 \end{array} \right] \\
 \text{HEAD-DTR} \left[\begin{array}{l}
 \text{PHON } \langle \text{gibt} \rangle \\
 \text{SYNSEM|LOC|CAT} \left[\begin{array}{l}
 \text{HEAD } \boxed{1} \left[\begin{array}{l} \text{VFORM } \textit{fin} \\ \textit{verb} \end{array} \right] \\
 \text{SUBCAT } \boxed{2} \oplus \langle \boxed{3} \rangle
 \end{array} \right] \\
 \text{word}
 \end{array} \right] \\
 \text{NON-HEAD-DTRS} \left\langle \begin{array}{l}
 \left[\begin{array}{l}
 \text{PHON } \langle \text{dem Mann} \rangle \\
 \text{SYNSEM } \boxed{3} \left[\begin{array}{l} \text{LOC|CAT} \left[\begin{array}{l} \text{HEAD } \left[\begin{array}{l} \text{CAS } \textit{dat} \\ \textit{noun} \end{array} \right] \\ \text{SUBCAT } \langle \rangle \end{array} \right] \end{array} \right] \\
 \text{HEAD-DTR } \dots \\
 \text{NON-HEAD-DTRS } \dots
 \end{array} \right] \right\rangle \\
 \text{head-argument-phrase}
 \end{array} \right]
 \end{array}$$

For the entire sentence *er das Buch dem Mann gibt* ‘he the book to the man gives’, we arrive at a structure (already shown in Figure 9.5) described by (20):

$$(20) \quad \left[\begin{array}{l}
 \text{SYNSEM|LOC|CAT} \left[\begin{array}{l}
 \text{HEAD } \left[\begin{array}{l} \text{VFORM } \textit{fin} \\ \textit{verb} \end{array} \right] \\
 \text{SUBCAT } \langle \rangle
 \end{array} \right]
 \end{array} \right]$$

This description corresponds to the sentence symbol S in the phrase structure grammar on page 55, however (20) additionally contains information about the form of the verb.

Using dominance schemata as an example, we have shown how generalizations about linguistic objects can be captured, however, we also want to be able to capture generalizations in other areas of the theory: Like Categorical Grammar, the HPSG lexicon contains a very large amount of information. Lexical entries (roots and words) can also be divided into classes, which can then be assigned types. In this way, it is possible to capture what all verbs, intransitive verbs and transitive verbs have in common. See Figure 24.1 on page 645.

Now that some fundamental concepts of HPSG have been introduced, the following section will show how the semantic contribution of words is represented and how the meaning of a phrase can be determined compositionally.

9.1.6 Semantics

An important difference between theories such as GB, LFG and TAG, on the one hand, and HPSG and CxG on the other is that the semantic content of a linguistic object is modeled in a feature structure just like all its other properties. As previously mentioned,

semantic information is found under the path `SYNSEM|LOC|CONT`. (21) gives an example of the `CONT` value for *Buch* ‘book’. The representation is based on Minimal Recursion Semantics (MRS):⁸

$$(21) \left[\begin{array}{l} \text{IND} \left[\begin{array}{l} \boxed{1} \left[\begin{array}{l} \text{PER} \ 3 \\ \text{NUM} \ \text{sg} \\ \text{GEN} \ \text{neu} \end{array} \right] \end{array} \right] \\ \text{RELS} \left\langle \left[\begin{array}{l} \text{INST} \ \boxed{1} \\ \text{buch} \end{array} \right] \right\rangle \\ \text{mrs} \end{array} \right]$$

`IND` stands for index and `RELS` is a list of relations. Features such as person, number and gender are part of a nominal index. These are important in determining reference or coreference. For example, *sie* ‘she’ in (22) can refer to *Frau* ‘woman’ but not to *Buch* ‘book’. On the other hand, *es* ‘it’ cannot refer to *Frau* ‘woman’.

- (22) Die Frau_i kauft ein Buch_j. Sie_i liest es_j.
 the woman buys a book she reads it
 ‘The woman buys a book. She reads it.’

In general, pronouns have to agree in person, number and gender with the element they refer to. Indices are then identified accordingly. In HPSG, this is done by means of structure sharing. It is also common to speak of *coindexation*. (23) provides some examples of coindexation of reflexive pronouns:

- (23) a. Ich_i sehe mich_i.
 I see myself
 b. Du_i siehst dich_i.
 you see yourself
 c. Er_i sieht sich_i.
 he sees himself
 d. Wir_i sehen uns_i.
 we see ourselves
 e. Ihr_i seht euch_i.
 you see yourselves
 f. Sie_i sehen sich_i.
 they see themselves

The question of which instances of coindexation are possible and which are necessary is determined by Binding Theory. Pollard & Sag (1992; 1994) have shown that Binding

⁸ Pollard & Sag (1994) and Ginzburg & Sag (2000) make use of situation semantics (Barwise & Perry 1983; Cooper, Mukai & Perry 1990; Devlin 1992). An alternative approach which has already been used in HPSG is Lexical Resource Semantics (Richter & Sailer 2004). For an earlier underspecification analysis in HPSG, see Nerbonne (1993).

Theory in HPSG does not have many of the problems that arise when implementing binding in GB with reference to tree configurations. There are, however, a number of open questions for Binding Theory in HPSG (Müller 1999a: Section 20.4).

(24) shows the CONT value for the verb *geben* ‘give’:

$$(24) \left[\begin{array}{l} \text{IND } \boxed{1} \text{ event} \\ \text{RELS } \left\langle \begin{array}{l} \text{EVENT } \boxed{1} \\ \text{AGENT } \textit{index} \\ \text{THEME } \textit{index} \\ \text{GOAL } \textit{index} \\ \textit{geben} \end{array} \right\rangle \\ \textit{mrs} \end{array} \right]$$

It is assumed that verbs have an event variable for the type *event*, which is represented under IND just as with indices for nominal objects. Until now, we did not assign elements in the valence list to argument roles in the semantic representation. This connection is referred to as *linking*. (25) shows how linking works in HPSG. The referential indices of the argument noun phrases are structure shared with one of the semantic roles of the relation contributed by the head.

(25) *gibt* ‘gives’:

$$\left[\begin{array}{l} \text{CAT } \left[\begin{array}{l} \text{HEAD } \left[\begin{array}{l} \text{VFORM } \textit{fin} \\ \textit{verb} \end{array} \right] \\ \text{SUBCAT } \left\langle \text{NP}[\textit{nom}]_{\boxed{1}}, \text{NP}[\textit{acc}]_{\boxed{2}}, \text{NP}[\textit{dat}]_{\boxed{3}} \right\rangle \end{array} \right] \\ \text{CONT } \left[\begin{array}{l} \text{IND } \boxed{4} \text{ event} \\ \text{RELS } \left\langle \begin{array}{l} \text{EVENT } \boxed{4} \\ \text{AGENT } \boxed{1} \\ \text{THEME } \boxed{2} \\ \text{GOAL } \boxed{3} \\ \textit{geben} \end{array} \right\rangle \\ \textit{mrs} \end{array} \right] \end{array} \right]$$

Since we can use general terms such as AGENT and PATIENT for argument roles, it is possible to state generalizations about valency classes and the realization of argument roles. For example, one can divide verbs into verbs taking an agent, verbs with an agent and theme, verbs with agent and patient etc. These various valency/linking patterns can be represented in type hierarchies and the specific lexical entries can be assigned to these classes, that is, one can let them inherit constraints from the respective types. A type constraint for verbs with agent, theme and goal would take the form of (26):

$$(26) \left[\begin{array}{l} \text{CAT|SUBCAT} \langle \boxed{1}, \boxed{2}, \boxed{3} \rangle \\ \text{CONT} \left[\begin{array}{l} \text{IND} \boxed{4} \text{ event} \\ \text{RELS} \left\langle \begin{array}{l} \text{EVENT} \boxed{4} \\ \text{AGENT} \boxed{1} \\ \text{THEME} \boxed{2} \\ \text{GOAL} \boxed{3} \\ \text{agent-theme-goal-rel} \end{array} \right\rangle \\ \text{mrs} \end{array} \right] \end{array} \right]$$

$\boxed{1}$ stands for an object of any syntactic category with the index $\boxed{1}$. The type for the relation *geben* is a subtype of *agent-theme-goal-rel*. The lexical entry for the word *geben* ‘give’ or the root *geb-* has the linking pattern in (26).

For more on theories of linking in HPSG, see Davis (1996), Wechsler (1991) und Davis & Koenig (2000).

Up to now, we have only seen how the meaning of lexical entries can be represented. The Semantic Principle determines the calculation of the semantic contribution of phrases: The index of the entire expression corresponds to the index of the head daughter, and the RELS value of the entire sign corresponds to the combination of the RELS values of the daughters plus any relations introduced by the dominance schema. The last point is important because the assumption that schemata can add something to meaning can capture the fact that there are some cases where the entire meaning of a phrase is more than simply the sum of its parts. Pertinent examples are often discussed as part of Construction Grammar. Semantic composition in HPSG is organized such that the contribution of phrases can be integrated into the complete meaning of an utterance. For examples, see Section 23.10.

The connection between the semantic contribution of the verb and its arguments is established in the lexical entry. As such, we ensure that the argument roles of the verb are assigned to the correct argument in the sentence. This is, however, not the only thing that the semantics is responsible for. It has to be able to generate the various readings associated with quantifier scope ambiguities (see page 92) as well as deal with semantic embedding of predicates under other predicates. All these requirements are fulfilled by MRS. Due to space considerations, we cannot go into detail here. The reader is referred to the article by Copestake et al. (2005) and to Section 21.3 in the discussion chapter.

9.1.7 Adjuncts

Analogous to the selection of arguments by heads via SUBCAT, adjuncts can also select their heads using a feature (MODIFIED). Adjectives, prepositional phrases that modify nouns, and relative clauses select an almost complete nominal projection, that is, a noun that only still needs to be combined with a determiner to yield a complete NP. (27) shows part of the lexical entry for *interessantes* ‘interesting’:

(27) CAT value for *interessantes* ‘interesting’:

$$\begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{MOD } \bar{N} \\ \text{adj} \end{bmatrix} \\ \text{SUBCAT} & \langle \rangle \end{bmatrix}$$

interessantes is an adjective that does not take any arguments and therefore has an empty SUBCAT list. Adjectives such as *treu* ‘faithful’ would have a dative NP in their SUBCAT list.

- (28) ein dem König treues Mädchen
 a the.DAT king faithful girl
 ‘a girl faithful to the king’

The CAT value is given in (29):

(29) CAT value für *treues* ‘faithful’:

$$\begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{MOD } \bar{N} \\ \text{adj} \end{bmatrix} \\ \text{SUBCAT} & \langle \text{NP}[\text{dat}] \rangle \end{bmatrix}$$

dem König treues ‘faithful to the king’ forms an adjective phrase, which modifies *Mädchen*.

Unlike the selectional feature SUBCAT that belongs to the features under CAT, MOD is a head feature. The reason for this is that the feature that selects the modifying head has to be present on the maximal projection of the adjunct. The \bar{N} -modifying property of the adjective phrase *dem König treues* has to be included in the representation of the entire AP just like it is present in the lexical entry for adjectives in (27) on the lexical level. The adjectival phrase *dem König treues* has the same syntactic properties as the basic adjective *interessantes*:

(30) CAT value für *dem König treues*:

$$\begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{MOD } \bar{N} \\ \text{adj} \end{bmatrix} \\ \text{SUBCAT} & \langle \rangle \end{bmatrix}$$

Since MOD is a head feature, the Head Feature Principle (see page 270) ensures that the MOD-Wert of the entire projection is identical to the MOD-Wert of the lexical entry for *treues* ‘loyal’.

As an alternative to the selection of the head by the modifier, one could assume a description of all possible adjuncts on the head itself. This was suggested by Pollard & Sag (1987: 161). Pollard & Sag (1994: Section 1.9) revised the earlier analysis since the semantics of modification could not be captured.⁹

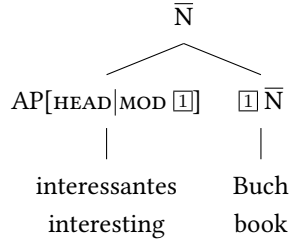


Figure 9.7: Head-adjunct structure (selection)

Figure 9.7 demonstrates selection in head-adjunct structures. Head-adjunct structures are licensed by the following schema:

Schema 2 (Head-Adjunct Schema (preliminary version))

head-adjunct-phrase \Rightarrow

$$\left[\begin{array}{l} \text{HEAD-DTR} | \text{SYNSEM } [1] \\ \text{NON-HEAD-DTRS } \left\langle \left[\begin{array}{l} \text{SYNSEM} | \text{LOC} | \text{CAT } \left[\begin{array}{l} \text{HEAD} | \text{MOD } [1] \\ \text{SUBCAT } \langle \rangle \end{array} \right] \end{array} \right] \right\rangle \end{array} \right]$$

The value of the selectional feature on the adjunct ($[1]$) is identified with the *SYNSEM* value of the head daughter, thereby ensuring that the head daughter has the properties specified by the adjunct. The *SUBCAT* list of the non-head daughter is an empty list, which is why only completely saturated adjuncts are allowed in head-adjunct structures. Phrases such as (31b) are therefore correctly ruled out:

- (31) a. die Wurst in der Speisekammer
the sausage in the pantry
- b. * die Wurst in
the sausage in

Example (31a) requires some further explanation. For the preposition *in* (as used in (31a)) has the following *CAT* value:

- (32) CAT value für *in*:
- $$\left[\begin{array}{l} \text{HEAD } \left[\begin{array}{l} \text{MOD } \overline{N} \\ \text{prep} \end{array} \right] \\ \text{SUBCAT } \langle \text{NP}[\text{dat}] \rangle \end{array} \right]$$

⁹ See Bouma, Malouf & Sag (2001a), however. Bouma, Malouf & Sag (2001a) pursue a hybrid analysis where there are adjuncts which select heads and also adjuncts that are selected by a head. *Minimal Recursion Semantics* (MRS) is the semantic theory underlying this analysis. Using this semantics, the problems arising for Pollard & Sag (1987) with regard to the semantics of modifiers are avoided.

After combining it with the nominal phrase *der Speisekammer* ‘the pantry’ one gets:

(33) CAT value for *in der Speisekammer* ‘in the pantry’:

$$\begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{MOD} & \bar{N} \\ & \text{prep} \end{bmatrix} \\ \text{SUBCAT} & \langle \rangle \end{bmatrix}$$

This representation corresponds to that of the adjective *interessantes* and can – ignoring the position of the PP – also be used in the same way: The PP modifies a \bar{N} .

Heads that can only be used as arguments but do not modify anything have a MOD-Wert of *none*. They can therefore not occur in the position of the non-head daughter in head-adjunct structure since the MOD-Wert of the non-head daughter has to be compatible with the SYNSEM value of the head daughter.

9.2 Passive

HPSG follows Bresnan’s argumentation (see Section 7.2) and takes care of the passive in the lexicon.¹⁰ A lexical rule takes the verb stem as its input and licenses the participle form and the most prominent argument (the so-called designated argument) is suppressed.¹¹ Since grammatical functions are not part of theory in HPSG, we do not require any mapping principles that map objects to subjects. Nevertheless, one still has to explain the change of case under passivization. If one fully specifies the case of a particular argument in the lexical entries, one has to ensure that the accusative argument of a transitive verb is realized as nominative in the passive. (34) shows what such a lexical would look like:

(34) Lexical rule for personal passives following Kiss (1992):

$$\begin{bmatrix} \text{PHON} & \boxed{1} \\ \text{SYNSEM|LOC|CAT} & \begin{bmatrix} \text{HEAD} & \text{verb} \\ \text{SUBCAT} & \langle \text{NP}[\text{nom}], \text{NP}[\text{acc}]_{\boxed{2}} \rangle \oplus \boxed{3} \end{bmatrix} \\ & \text{stem} \end{bmatrix} \mapsto$$

¹⁰ Some exceptions to this are analyses influenced by Construction Grammar such as Tseng (2007) and Haugereid (2007). These approaches are problematic, however, as they cannot account for Bresnan’s adjectival passives. For other problems with Haugereid’s analysis, see Müller (2007c) and Section 23.3.6.

¹¹ For more on the designated argument, see Haider (1986a). HPSG analyses of the passive in German have been considerably influenced by Haider. Unaccusative verbs differ from unergatives and transitives in that they do not have a designated argument. We cannot go into the literature on unaccusativity here. The reader is referred to the original works by Haider and the chapter on the passive in Müller (2007b).

$$\left[\begin{array}{l} \text{PHON } f(\boxed{1}) \\ \text{SYNSEM|LOC|CAT} \left[\begin{array}{l} \text{HEAD} \left[\begin{array}{l} \text{VFORM } \textit{passive-part} \end{array} \right] \\ \text{SUBCAT} \left\langle \text{NP}[\textit{nom}]_{\boxed{2}} \right\rangle \oplus \boxed{3} \end{array} \right] \\ \textit{word} \end{array} \right]$$

This lexical rule takes the verb stem as its input¹², which requires a nominative argument, an accusative argument and possibly further arguments (if $\boxed{3}$ is not the empty list) and licenses a lexical entry that requires a nominative argument and possibly the arguments in $\boxed{3}$. The output of the lexical rule specifies the *VFORM* value of the output word. This is important as the auxiliary and the main verb must go together. For example, it is not possible to use the perfect participle instead of the passive participle since these differ in their valency in Kiss' approach:

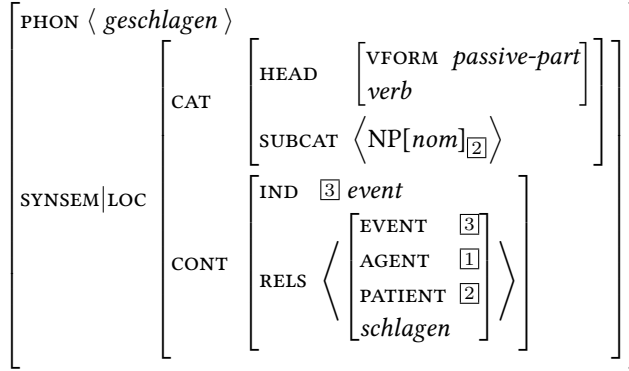
- (35) a. Der Mann hat den Weltmeister geschlagen.
 the man has the world.champion beaten
 'The man has beaten the world champion.'
- b. * Der Mann wird den Weltmeister geschlagen.
 the man is the world.champion beaten
- c. Der Weltmeister wird geschlagen.
 the world.champion is beaten
 'The world champion is (being) beaten.'

There are a few conventions for the meaning of lexical rules: All information that is not mentioned in the output sign is taken over from the input sign. Thus, the meaning of the verb is not mentioned in the passive rule, which makes sense as the passive rule is a meaning preserving rule. The *CONT* values of the input and output are identical. It is important here that the linking information its retained. As an example consider the application of the rule to the verb stem *schlag-* 'hit':

- (36) a. Input:
- $$\left[\begin{array}{l} \text{PHON } \langle \textit{schlag} \rangle \\ \text{SYNSEM|LOC} \left[\begin{array}{l} \text{CAT} \left[\begin{array}{l} \text{HEAD } \textit{verb} \\ \text{SUBCAT } \left\langle \text{NP}[\textit{nom}]_{\boxed{1}}, \text{NP}[\textit{acc}]_{\boxed{2}} \right\rangle \end{array} \right] \\ \text{CONT} \left[\begin{array}{l} \text{IND } \boxed{3} \textit{event} \\ \text{RELS } \left\langle \begin{array}{l} \text{EVENT } \boxed{3} \\ \text{AGENT } \boxed{1} \\ \text{PATIENT } \boxed{2} \\ \textit{schlagen} \end{array} \right\rangle \end{array} \right] \end{array} \right] \end{array} \right]$$

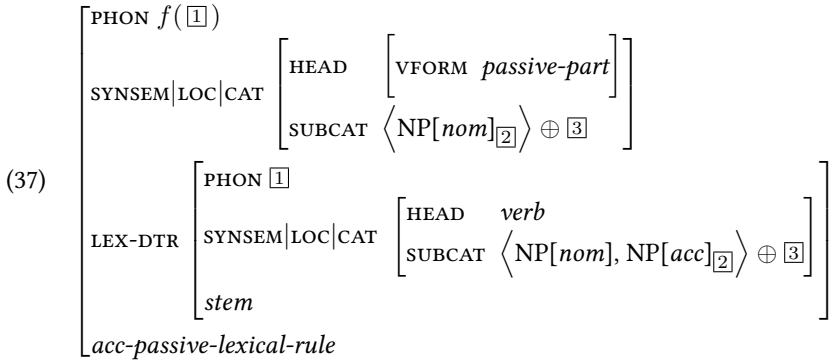
¹² The term *stem* includes roots (*helf-* 'help-', products of derivation (*besing-* 'to extol') and compounds. The lexical rule can therefore also be applied to stems *helf-* and derived forms such as *besing-*.

b. Output:



The agent role is connected to the subject of *schlag-*. After passivization, the subject is suppressed and the argument that is connected to the patient role of *schlag-* becomes the subject of the participle. Argument linking is not affected by this and thus the nominative argument is correctly assigned to the patient role.

Lexical rules such as (34) can also be captured with feature descriptions (Meurers 2001):



What is on the left-hand side of the rule in (34), is contained in the value of LEX-DTR in (37). Since this kind of lexical rule is fully integrated into the formalism, feature structures corresponding to these lexical rules also have their own type. If the result of the application of a given rule is an inflected word, then the type of the lexical rule (*acc-passive-lexical-rule* in our example) is a subtype of *word*. Since lexical rules have a type, it is possible to state generalizations over lexical rules.

The lexical rules discussed thus far work well for the personal passive. For the impersonal passive, however, we would require a second lexical rule. Furthermore, we would have two different lexical items for the passive and the perfect, although the forms are always identical. In the following, we will discuss the basic assumptions that we need for a theory of the passive that can sufficiently explain both personal and impersonal passives and thereby only require one lexical item for the participle form.

- Depictive predicates (Müller 2004d; 2002a; 2008)
- Binding Theory (Grewendorf 1985; Pollard und Sag: 1992; 1994: Chapter 6)

The Case Principle ensures that the subjects of the verbs listed above have to be realized in the nominative and also that objects with structural case are assigned accusative.

With the difference between structural and lexical case, it is possible to formulate a passive-lexical rule that can account for both the personal and the impersonal passive:

(39) Lexical rule for personal and impersonal passive (simplified):

$$\left[\begin{array}{l} \text{PHON } \boxed{1} \\ \text{SYNSEM|LOC|CAT} \left[\begin{array}{l} \text{HEAD } verb \\ \text{SUBCAT } \langle \text{NP}[\text{str}] \rangle \oplus \boxed{2} \end{array} \right] \\ stem \end{array} \right] \mapsto \left[\begin{array}{l} \text{PHON } f(\boxed{1}) \\ \text{SYNSEM|LOC|CAT} \left[\begin{array}{l} \text{HEAD } \left[\begin{array}{l} \text{VFORM } ppp \end{array} \right] \\ \text{SUBCAT } \boxed{2} \end{array} \right] \\ word \end{array} \right]$$

This lexical rule does exactly what we expect it to do from a pretheoretical perspective on the passive: It suppresses the most prominent argument with structural case, that is, the argument that corresponds to the subject in the active clause. After combining it with the passive auxiliary, we arrive at the following SUBCAT lists:

- (40) a. *geschlafen wird* ‘slept is’: SUBCAT $\langle \rangle$
 b. *unterstützt wird* ‘supported is’: SUBCAT $\langle \text{NP}[\text{str}]_k \rangle$
 c. *geholfen wird*: ‘helped is’ SUBCAT $\langle \text{NP}[\text{ldat}]_k \rangle$
 d. *geschenkt wird*: ‘given is’ SUBCAT $\langle \text{NP}[\text{str}]_k, \text{NP}[\text{ldat}]_l \rangle$

In (40), a different NP is in first position. If this NP has structural case, it will receive nominative case. If this is not the case, as in (40c), the case remains as it was, that is, lexically-specified.

We cannot go into the analysis of the perfect here. It should be noted however that the same lexical item for the participle is used for (41).

- (41) a. Er hat den Weltmeister geschlagen.
 he has the world.champion beaten
 ‘He has beaten the world champion.’
 b. Der Weltmeister wurde geschlagen.
 the world.champion was beaten
 ‘The world champion was beaten.’

It is the auxiliary that determines which arguments are realized (Haider 1986a; Müller 2007b: Chapter 17). The lexical rule in (39) licenses a form that can be used both in passive and perfect. Therefore, the *VFORM* value is of *ppp*, which stands for *participle*.

One should note that this analysis of the passive works without movement of constituents. The problems with the GB analysis do not arise here. Reordering of arguments (see Section 9.4) is independent of passivization. The accusative object is not mentioned at all unlike in GPSG, Categorical Grammar or Bresnan's LFG analysis from before the introduction of Lexical Mapping Theory. The passive can be analyzed directly as the suppression of the subject. Everything else follows from interaction with other principles of grammar.

9.3 Verb position

The analysis of verb position that I will present here is based on the GB-analysis. In HPSG, there are a number of different approaches to describing verb position, however in my opinion, the HPSG variant of the GB analysis is the only adequate one (Müller 2005b;c). The analysis of (42) can be summarized as follows: In the verb-initial clauses, there is a trace in verb-final position. There is a special form of the verb in initial position that selects a projection of the verb trace. This special lexical item is licensed by a lexical rule. The connection between the verb and the trace is treated like long-distance dependencies in GPSG via identification of information in the tree or feature structure (structure sharing).

- (42) Kennt_k jeder diesen Mann _{¬k}?
 knows everyone this man
 'Does everyone know this man?'

Figure 9.8 on the next page gives an overview of this: The verb trace in final position behaves just like the verb both syntactically and semantically. The information about the missing word is represented as the value of the feature *DOUBLE SLASH* (abbreviated: *DSL*). This is a head feature and is therefore passed up to the maximal projection (VP). The verb in initial position has a VP in its *SUBCAT* list which is missing a verb (VP//V). This is the same verb that would normally be the input for the lexical rule and would normally occur in final position. In Figure 9.8, there are two maximal projections of the verb: *jeder diesen Mann _{¬k}* with the trace as the head and *kennt jeder diesen Mann _{¬k}* with *kennt* as the head.

This analysis will be explained in more detail in what follows. For the trace in Figure 9.8, one could assume the following lexical entry:

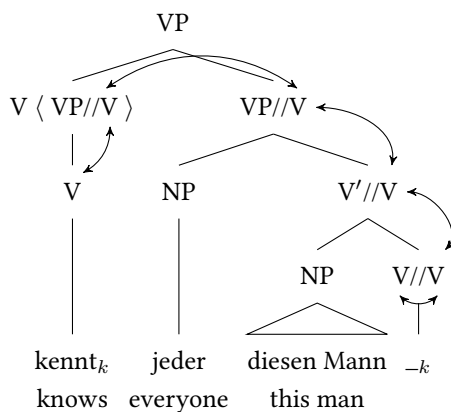


Figure 9.8: Analysis of verb position in HPSG

(43) Verb trace for *kennt* ‘knows’:

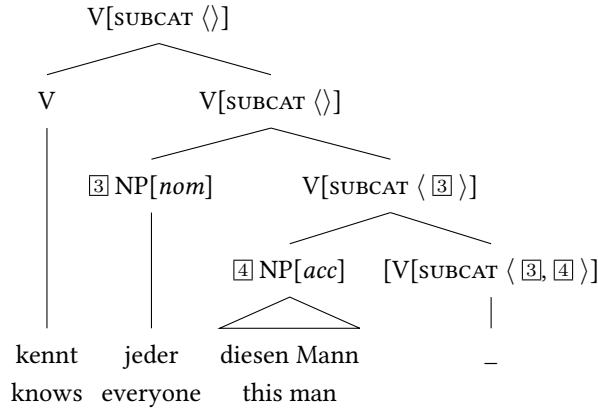
PHON <>	CAT	HEAD	[VFORM <i>fin</i> <i>verb</i>]	SUBCAT < NP[nom] _[1] , NP[acc] _[2] >	IND 3	EVENT 3	EXPERIENCER 1	THEME 2	[<i>kennen</i>]
SYNSEM LOC	CONT	RELS	<	[<i>kennen</i>]	[<i>kennen</i>]	[<i>kennen</i>]	[<i>kennen</i>]	[<i>kennen</i>]	[<i>kennen</i>]

This lexical entry differs from the normal verb *kennt* only in its PHON value. The syntactic aspects of an analysis with this trace are represented in Figure 9.9 on the following page.

The combination of the trace with *diesen Mann* and *jeder* follows the rules and principles that we have encountered thus far. This begs the immediate question as to what licenses the verb *kennt* in Figure 9.9 and what status it has.

If we want to capture the fact that the finite verb in initial position behaves like a complementizer (Höhle 1997), then it makes sense to give head status to *kennt* in Figure 9.9 and have *kennt* select a saturated, verb-final verbal projection. Finite verbs in initial position differ from complementizers in that they require a projection of a verb trace, whereas complementizers need projections of overt verbs:

- (44) a. dass [er das Buch kennt]
 that he the book knows
 ‘that he knows the book’

Figure 9.9: Analysis of *Kennt jeder diesen Mann?* 'Does everyone know this man?'

- b. Kennt [er das Buch _]
 knows he the book
 'Does he know the book?'

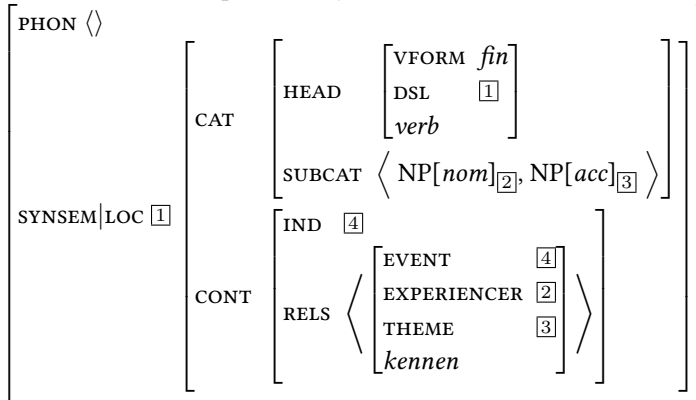
It is normally not the case that *kennen* 'know' selects a complete sentence and nothing else as would be necessary for the analysis of *kennt* as the head in (44b). Furthermore, we must ensure that the verbal projection with which *kennt* is combined contains the verb trace belonging to *kennt*. If it could contain a trace belonging to *gibt* 'gives', for example, we would be able to analyze sentences such as (45b):

- (45) a. Gibt [der Mann der Frau das Buch _{-gibt}].
 gives the man the woman the book
 'Does the man give the woman the book?'
 b. * Kennt [der Mann der Frau das Buch _{-gibt}].
 knows the man the woman the book

In the preceding discussion, the dependency between the fronted verb and the verb trace was expressed by coindexation. In HPSG, identity is always determined by structure sharing. The verb in initial position must therefore require that the trace has exactly those properties of the verb that it would have had, were it in final position. The information that must be shared is therefore all locally relevant syntactic and semantic information, that is, all information under *LOCAL*. Since *PHON* is not part of the *LOCAL* features, it is not shared and this is why the *PHON* values of the trace and verb can differ. Up to now, one crucial detail has been missing in the analysis: structure sharing of the *LOCAL* values of the trace cannot be directly achieved since the verb *kennt* can only select the properties of the projection of the trace and the *SUBCAT* list of the selected projection is an empty list. This leads us to the same problem as with (45b). It must therefore be ensured that all information about the verb trace is available on the highest node of its

projection. This can be achieved by introducing a head feature whose value is identical to the LOCAL value of the trace. This feature is referred to as DSL. DSL stands for *double slash* and is called so because it has a similar function to the SLASH feature, which we will encounter in the following section.¹⁴ (46) shows the modified entry for the verb trace:

(46) Verb trace of *kennt* (preliminary version):



Through sharing of the LOCAL values and the DSL value in (46), the syntactic and semantic information of the verb trace is present at its maximal projection, and the verb in initial position can check whether the projection of the trace is compatible.

The special lexical rule for verb-initial position is licensed by the following lexical rule:¹⁵

¹⁴ The feature DSL was proposed by Jacobson (1987a) as part of Categorical Grammar to describe head movement in English inversions. Borsley (1989) adopted this idea and translated it into HPSG terms thereby showing how head movement in a HPSG variant of the CP/IP system can be modeled using DSL. The introduction of the DSL feature to describe movement processes in HPSG is motivated by the fact that, unlike long-distance dependencies as will be discussed in Section 9.5, this kind of movement is local.

The suggestion to percolate information about the verb trace as part of the head information comes from Oliva (1992).

¹⁵ The lexical rule analysis cannot explain sentences such as (i):

- (i) Karl kennt und liebt diese Schallplatte.
Karl knows and loves this record

This has to do with the fact that the lexical rule cannot be applied to the result of coordination, which constitutes a complex syntactic object. If we apply the lexical rule individually to each verb, then we arrive at variants of the verb which would each select verb traces for *kennen* ‘to know’ and *lieben* ‘to love’. Since the CAT values of the conjuncts are identified with each other in coordinations, coordinative linking of both V1 variants of *kennt* and *liebt* would be ruled out since the DSL value of the selected VPs is contained in the meaning of each of the verbs (Müller 2005b: 13). Instead of a lexical rule, one must assume a unary syntactic rule that applies to the phrase *kennt und liebt* ‘knows and loves’. As we have seen, lexical rules in the HPSG formalization assumed here also correspond to unary rules such that the difference between (47) and a corresponding syntactic rule is mostly a difference in representation.

(47) Lexical rule for verbs in initial position:

$$\left[\begin{array}{c|c} \text{SYNSEM} & \text{LOC } [1] \end{array} \left[\begin{array}{c|c} \text{CAT} & \text{HEAD} \end{array} \left[\begin{array}{c} \text{VFORM } fin \\ \text{INITIAL } - \\ verb \end{array} \right] \right] \right] \mapsto \left[\begin{array}{c|c} \text{SYNSEM} & \text{LOC} & \text{CAT} \end{array} \left[\begin{array}{c|c} \text{HEAD} & \left[\begin{array}{c} \text{VFORM } fin \\ \text{INITIAL } + \\ \text{DSL } none \\ verb \end{array} \right] \\ \text{SUBCAT} & \left\langle \left[\begin{array}{c|c} \text{LOC} & \text{CAT} \end{array} \left[\begin{array}{c} \text{HEAD} \left[\begin{array}{c} \text{DSL } [1] \\ verb \end{array} \right] \\ \text{SUBCAT } \langle \rangle \end{array} \right] \right] \right\rangle \end{array} \right] \right] \right]$$

The verb licensed by this lexical rule selects a maximal projection of the verb trace which has the same local properties as the input verb. This is achieved by the coindexation of the LOCAL values of the input verb and the DSL values of the selected verb projection. Only finite verbs in final position (INITIAL−) can be the input for this rule. The output is a verb in initial-position (INITIAL+).

The corresponding extended analysis is given in Figure 9.10. V1-LR stands for the verb-initial lexical rule.

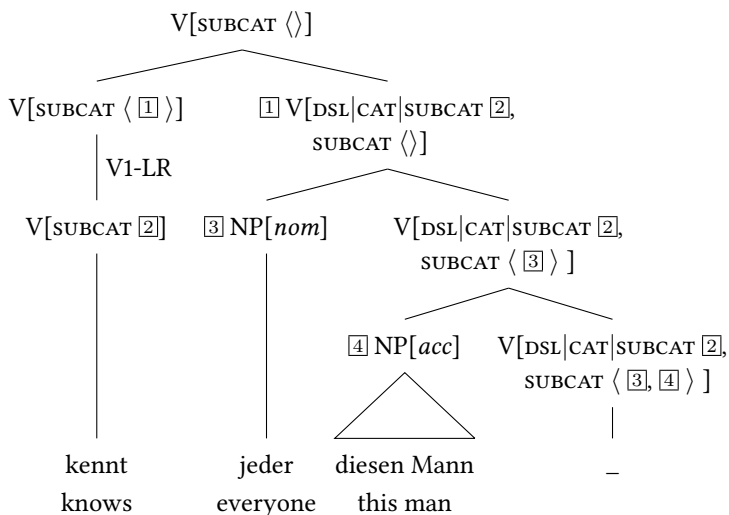


Figure 9.10: Visualization of the analysis of *Kennt jeder diesen Mann?* ‘Does everyone know this man?’

The lexical rule in (47) licenses a verb that selects a VP ($\boxed{1}$ in Figure 9.10). The DSL value of this VP corresponds to the LOC-Wert of the input verb for the lexical rule. Part of the DSL value is also the valency information represented in Figure 9.10 ($\boxed{2}$). Since DSL is a head feature, the DSL value of the VP is identical to that of the verb trace and since the LOC-Wert of the verb trace is identified with the DSL value, the SUBCAT information of the verb *kennen* is also available at the trace. The combination of the trace with its argument proceeds exactly as with an ordinary verb.

It would be unsatisfactory if we had to assume a special trace for every verb. Fortunately, this is not necessary as a general trace as in (48) will suffice for the analysis of sentences with verb movement.

(48) General verb trace following Meurers (2000: 206–208):

$$\left[\begin{array}{l} \text{PHON } \langle \rangle \\ \text{SYNSEM} | \text{LOC } \boxed{1} \left[\text{CAT} | \text{HEAD} | \text{DSL } \boxed{1} \right] \end{array} \right]$$

This may seem surprising at first glance, but if we look closer at the interaction of the lexical rule (47) and the percolation of the DSL feature in the tree, then it becomes clear that the DSL value of the verb projection and therefore the LOCAL value of the verb trace is determined by the LOCAL value of the input verb. In Figure 9.10, *kennt* is the input for the verb movement lexical rule. The relevant structure sharing ensures that, in the analysis of (42), the LOCAL value of the verb trace corresponds exactly to what is given in (46).

The most important points of the analysis of verb position are summarized below:

- A lexical rule licenses a special lexical item for each finite verb.
- This lexical item occupies the initial position and requires as its argument a complete projection of a verb trace.
- The projection of the verb trace must have a DSL value corresponding to the LOCAL value of the input verb of the lexical rule.
- Since DSL is a head feature, the selected DSL value is also present on the trace.
- As the DSL value of the trace is identical to its LOCAL value, the LOCAL value of the trace is identical to the LOCAL value of the input verb in the lexical rule.

After discussing the analysis of verb-first sentences, we will now turn to local reordering.

9.4 Local reordering

There are several possibilities for the analysis of constituent ordering in the middle field: One can assume completely flat structures as in GPSG (Kasper 1994), or instead assume binary branching structures and allow for arguments to be saturated in any order. A

compromise was proposed by Kathol (2001) and Müller (1999a; 2002a; 2004c): Binary branching structures with a special list that contains the arguments and adjuncts belonging to one head. The arguments and adjuncts are allowed to be freely ordered inside such lists. See Reape (1994) for the formal details of these approaches. Both the completely flat analysis and the compromise have proved to be on the wrong track (see Müller (2005b; 2014c) and Müller (2007b: Section 9.5.1)) and as such, I will only discuss the analysis with binary branching structures.

Figure 9.11 shows the analysis of (49a).

- (49) a. [weil] jeder diesen Mann kennt
 because everyone this man knows
 b. [weil] diesen Mann jeder kennt
 because this man everyone knows
 ‘because everyone knows this man’

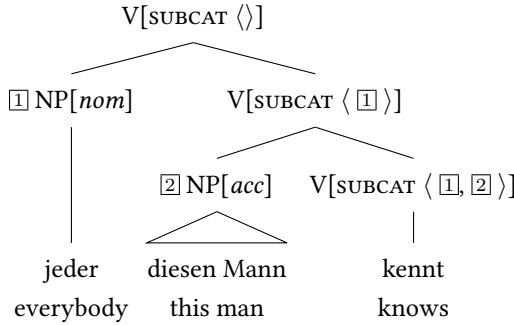


Figure 9.11: Analysis of constituent order in HPSG: normal order

The arguments of the verb are combined with the verb starting with the last, as explained in Section 9.1.2. The analysis of the marked order is shown in Figure 9.12 on the next page. Both trees differ only in the order in which the elements are taken off from the SUBCAT list: In Figure 9.11, the last element of the SUBCAT list is discharged first and in Figure 9.12 the first one is.

The following schema is a revised form of the Head-Argument Schema:

Schema 3 (Head-Argument Schema (binary branching))

head-argument-phrase \Rightarrow

$$\left[\begin{array}{l} \text{SYNSEM|LOC|CAT|SUBCAT } [1] \oplus [3] \\ \text{HEAD-DTR|SYNSEM|LOC|CAT|SUBCAT } [1] \oplus \langle [2] \rangle \oplus [3] \\ \text{NON-HEAD-DTRS } \langle [\text{SYNSEM } [2]] \rangle \end{array} \right]$$

Whereas in the first version of the Head-Argument Schema it was always the last element of the SUBCAT list that was combined with the head, the SUBCAT list is divided into

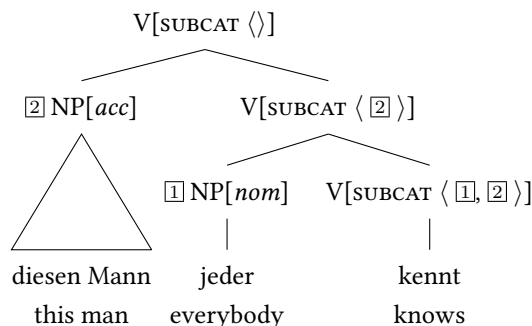


Figure 9.12: Analysis of constituent order in HPSG: marked order

three parts using *append*: A list of arbitrary length ($[1]$), a list consisting of exactly one element ($\langle [2] \rangle$) and a further list of arbitrary length ($[3]$). The lists $[1]$ and $[3]$ are combined and the result is the SUBCAT value of the mother node.

Languages with fixed constituent order (such as English) differ from languages such as German in that they discharge the arguments starting from one side (for more on the subject in English, see Section 9.6), whereas languages with free constituent order can combine arguments with the verb in any order. In languages with fixed constituent order, $[1]$ and $[3]$ are an empty list. Since German structures are not restricted with regard to $[1]$ or $[3]$, that is $[1]$ and $[3]$ can either be an empty list or contain elements, we capture the intuition that there are less restrictions in languages with free constituent order than in languages with fixed order. We can compare this to the Kayneian analysis from Section 4.6.1, where it was assumed that all languages are derived from the base order [specifier [head complement]] (see Figure 4.20 on page 148 for Laenzlinger’s analysis of German as an SVO-language (Laenzlinger 2004)). In these kinds of analyses, languages such as English constitute the most basic case and languages with free ordering require some considerable theoretical effort to get the order right. On the other hand, the analysis proposed here requires more theoretical restrictions if the language has more restrictions on permutations of its constituents. The complexity of the licensed structures does not differ considerably from language to language under an HPSG approach. Languages differ only in the type of branching they have.^{16,17}

The analysis presented here utilizing combination of arguments in any order is similar to that of Fanselow (2001) in the framework of GB/MP as well as the Categorical Grammar

¹⁶ This does not exclude that the structures in question have different properties as far as their processability by humans is concerned. See Gibson (1998); Hawkins (1999) and Chapter 15.

¹⁷ Haider (1997b: 18) has pointed out that VX languages differ from XV languages in branching type, as proposed here. This affects the c-command relations and therefore has implications for Binding Theory in GB/MP. The direction of branching is irrelevant for HPSG analyses as Binding Principles are defined using o-command (Pollard & Sag 1994: Chapter 6) and o-command makes reference to the Obliqueness Hierarchy, that is, the order of elements in the SUBCAT list rather than the order in which these elements are combined with the head.

analyses of Hoffman (1995: Section 3.1) and Steedman & Baldridge (2006). Gunji proposed similar HPSG analyses for Japanese as early as 1986.

9.5 Long-distance dependencies

The analysis of long-distance dependencies utilizes techniques that were originally developed in GPSG: Information about missing constituents is passed up the tree (or feature structure).¹⁸ There is a trace at the position where the fronted element would normally occur. Figure 9.13 shows the analysis of (50).

- (50) [Diesen Mann]_j kennt_i _{-j} jeder _{-i}.
 this man knows everyone
 ‘Everyone knows this man.’

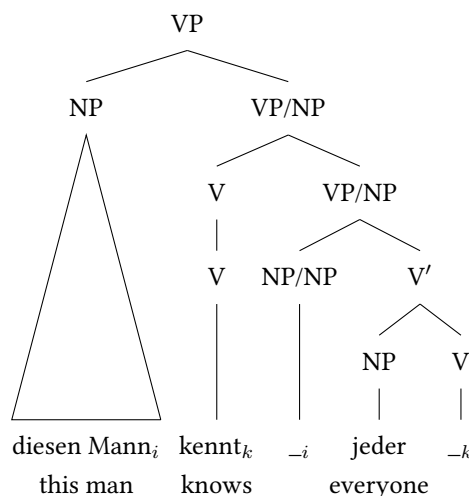


Figure 9.13: Analysis of long-distance dependencies in HPSG

Unlike the verb movement analysis discussed in Section 9.3, constituent movement is non-local, which is why the two movement types are modeled with different features (SLASH vs. DSL). DSL is a head feature and, like all other head features, projects to the highest node of a projection (for more on the Head Feature Principle, see page 270). SLASH, on

¹⁸ Nothing is actually ‘passed up’ in a literal sense in feature structures or trees. For expository purposes, it makes sense to explain the analysis as if the structure were built bottom-up. Linguistic knowledge is independent of the direction of processing. In modern computer implementations, structure building is mostly carried out bottom-up but there were other systems which worked top-down. The only thing that is important in the analysis of nonlocal dependencies is that the information about the missing element on all intermediate nodes is identical to the information in the filler and the gap.

the other hand, is a feature that belongs to the NONLOC features represented under SYNSEM|NONLOC. The value of the NONLOC feature is a structure with the features INHERITED (or INHER for short) and TO-BIND. The value of INHER is a structure containing information about elements involved in a long-distance dependency. (51) gives the structure assumed by Pollard & Sag (1994: 163):

$$(51) \left[\begin{array}{ll} \text{QUE} & \text{list of } n\text{pros} \\ \text{REL} & \text{list of indices} \\ \text{SLASH} & \text{list of local structures} \\ \text{nonloc} & \end{array} \right]$$

QUE is important for the analysis of interrogative clauses as is REL for the analysis of relative clauses. Since these will not feature in this book, they will be omitted in what follows. The value of SLASH is a list of *local*-objects.

As with the analysis of verb movement, it is assumed that there is a trace in the position where the object would normally occur and that this trace shares the properties of that object. The verb can therefore satisfy its valency requirements locally. Information about whether there has been combination with a trace and not with a genuine argument is represented inside the complex sign and passed upward in the tree. The long-distance dependency can then be resolved by an element in the prefield higher in the tree.

Long-distances dependencies are introduced by the trace, which has a feature corresponding to the LOCAL value of the required argument in its SLASH list. (52) shows the description of the trace as is required for the analysis of (50):

(52) Trace of the accusative complement of *kennen* (preliminary):

$$\left[\begin{array}{ll} \text{PHON} & \langle \rangle \\ \text{SYNSEM} & \left[\begin{array}{ll} \text{LOC} & \boxed{1} \\ \text{NONLOC} & \left[\begin{array}{ll} \text{INHER|SLASH} & \langle \boxed{1} \rangle \\ \text{TO-BIND|SLASH} & \langle \rangle \end{array} \right] \end{array} \right] \\ \text{word} & \end{array} \right]$$

Since traces do not have internal structure (no daughters), they are of type *word*. The trace has the same properties as the accusative object. The fact that the accusative object is not present at the position occupied by the trace is represented by the value of SLASH.

The following principle is responsible for ensuring that NONLOC information is passed up the tree.

Principle 6 (Nonlocal Feature Principle)

For each nonlocal feature, the INHERITED value on the mother is a list that contains the INHERITED values on the daughters minus the TO-BIND value on the head daughter.

The Head-Filler Schema licenses the highest node in Figure 9.14 on the facing page:

Schema 4 (Head-Filler Schema)

head-filler-phrase \Rightarrow

$$\left[\begin{array}{l} \text{HEAD-DTR} | \text{SYNSEM} \\ \text{NON-HEAD-DTRS} \left\langle \begin{array}{l} \text{SYNSEM} \left[\begin{array}{l} \text{LOC } \boxed{1} \\ \text{NONLOC} | \text{INHER} | \text{SLASH } \langle \rangle \end{array} \right] \end{array} \right\rangle \end{array} \right. \left[\begin{array}{l} \text{LOC} | \text{CAT} \left[\begin{array}{l} \text{HEAD} \left[\begin{array}{l} \text{VFORM } \textit{fin} \\ \text{INITIAL } + \\ \textit{verb} \end{array} \right] \\ \text{SUBCAT } \langle \rangle \end{array} \right] \\ \text{NONLOC} \left[\begin{array}{l} \text{INHER} | \text{SLASH } \langle \boxed{1} \rangle \\ \text{TO-BIND} | \text{SLASH } \langle \boxed{1} \rangle \end{array} \right] \end{array} \right] \end{array} \right]$$

This schema combines a finite, verb-initial clause (INITIAL+) that has an element in SLASH with a non-head daughter whose LOCAL value is identical to the SLASH element. In this structure, no arguments are saturated. Nothing can be extracted from the filler daughter itself, which is ensured by the specification of the SLASH value of the non-head daughter. Figure 9.14 on the next page shows a more detailed variant of the analysis of fronting to the prefield.

The trace of verb movement for *kennt* is combined with a nominative NP and an extraction trace. The extraction trace stands for the accusative object in the above example. The accusative object is listed in the SUBCAT list of the verb ($\boxed{4}$). Following the mechanism for verb movement, the valency information that was originally contained in the entry for *kennt* ($\langle \boxed{3}, \boxed{4} \rangle$) is present on the verb trace. The combination of the projection of the verb trace with the extraction trace follows in exactly the same way as we saw for non-fronted arguments. The SLASH value of the extraction trace is passed up the tree and bound off by the Head-Filler Schema.

(52) provides the lexical entry for a trace that can function as the accusative object of *kennen*. As with the analysis of verb movement, it is not necessary to have numerous extraction traces with differing properties listed in the lexicon. A more general entry such as the one in (53) will suffice:

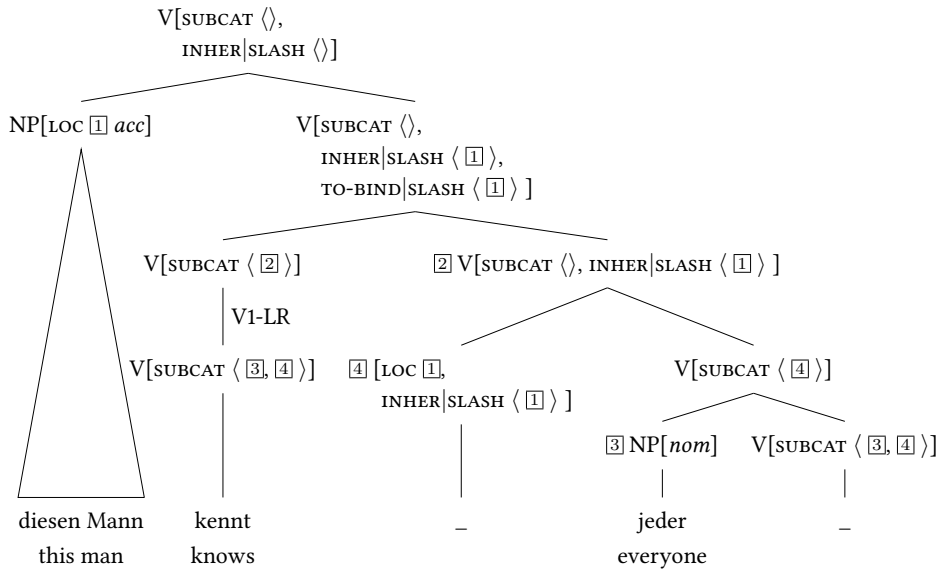
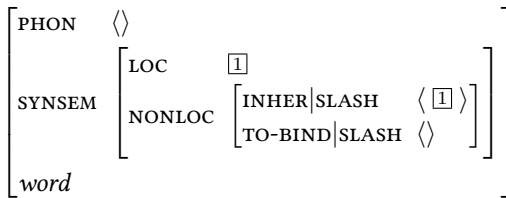


Figure 9.14: Analysis of *Diesen Mann kennt jeder*. 'Everyone knows this man.' combined with the verb movement analysis for verb-initial order

(53) Extraction trace:



This has to do with the fact that the head can satisfactorily determine the LOCAL properties of its arguments and therefore also the local properties of the traces that it combines with. The identification of the object in the SUBCAT list of the head with the SYNSEM value of the trace coupled with the identification of the information in SLASH with information about the fronted element serves to ensure that the only elements that can be realized in the prefield are those that fit the SUBCAT list of the head. The same holds for fronted adjuncts: Since the LOCAL value of the constituent in the prefield is identified with the LOCAL value of the trace via the SLASH feature, there is then sufficient information available about the properties of the trace.

The central points of the preceding analysis can be summarized as follows: Information about the local properties of a trace is contained in the trace itself and then present on all nodes dominating it until one reaches the filler. This analysis can offer an explanation for languages where certain elements show inflection depending on whether they

are part of a constituent that has been introduced by a long-distance dependency. Bouma, Malouf & Sag (2001a) cite Irish, Chamorro, Palauan, Icelandic, Kikuyu, Ewe, Thompson Salish, Moore, French, Spanish, and Yiddish as examples of such languages and provide corresponding references. Since information is passed on step-by-step in HPSG analyses, all nodes intervening in a long-distance dependency can access the elements in that dependency.

9.6 New developments and theoretical variants

In this chapter, we have been treating SUBCAT as the only valency feature. This corresponds to the state of theory in Pollard & Sag (1994: Chapter 1–8). It has turned out to be desirable to assume at least one additional valency feature and a corresponding schema for the combination of constituents. This additional feature is called SPECIFIER (SPR) and is used in grammars of English (Pollard & Sag 1994: Chapter 9) and German (Müller 2007b: Section 9.3) for the combination of a determiner with a noun. It is assumed that the noun selects its determiner. For the noun *Zerstörung* ‘destruction’, we have the following CAT value:

$$(54) \left[\begin{array}{l} \text{HEAD} \quad \left[\begin{array}{l} \text{INITIAL} \quad + \\ \text{noun} \end{array} \right] \\ \text{SPR} \quad \langle \text{DET} \rangle \\ \text{SUBCAT} \quad \langle \text{NP}[\text{GEN}], \text{PP}[\text{durch}] \rangle \end{array} \right]$$

Schema 5 can be used just like the Head-Argument Schema for the combination of noun and determiner.

Schema 5 (Specifier-Head-Schema)

head-specifier-phrase \Rightarrow

$$\left[\begin{array}{l} \text{SYNSEM|LOC|CAT|SPR} \quad [1] \\ \text{HEAD-DTR|SYNSEM|LOC|CAT} \quad \left[\begin{array}{l} \text{SPR} \quad [1] \oplus \langle [2] \rangle \\ \text{SUBCAT} \quad \langle \rangle \end{array} \right] \\ \text{NON-HEAD-DTRS} \quad \langle [\text{SYNSEM} \quad [2]] \rangle \end{array} \right]$$

The analysis of the NP in (55) with the specifier schema is shown in Figure 9.15 on the facing page.

- (55) die Zerstörung der Stadt durch die Soldaten
the destruction of.the city by the soldiers

Following the linearization rules discussed in Section 9.1.3, it is ensured that the noun occurs before the complements as the INITIAL value of the noun is ‘+’. The LP-rule in (56) leads to the determiner being ordered to the left of the noun.

- (56) Specifier < Head

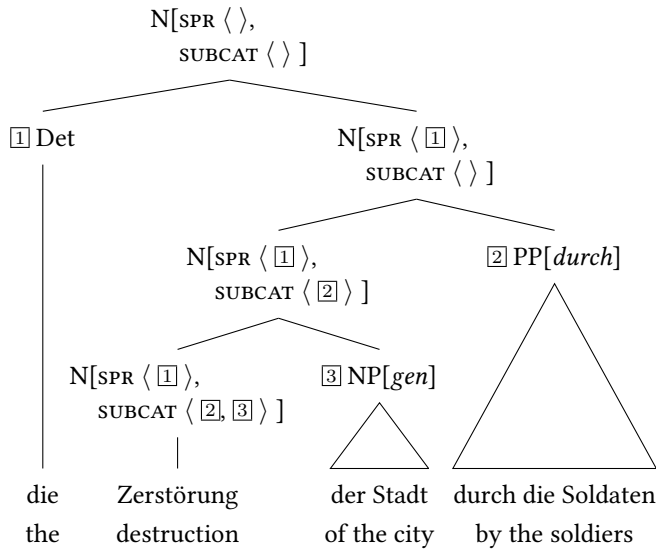


Figure 9.15: NP-Analysis with valency features SPR

In grammars of English, the SPR feature is also used for the selection of the subject of verbs (Sag, Wasow & Bender 2003: Section 4.3). In a sentence such as (57), the verb is first combined with all its complements (the elements in the SUBCAT or COMPS in newer works) and is then combined with the subject in a second step by applying the Schema 5.

(57) Max likes ice cream.

As we have seen in Section 9.4, it made sense to represent subjects and arguments in the same valence list for the analysis of finite sentences. In this way, we can capture the fact that the order in which a verb is combined with its arguments is not fixed.

A further expansion from Pollard & Sag (1994: Chapter 9) is the introduction of an additional list that is called ARG-ST in newer works. ARG-ST stands for Argument Structure. The ARG-ST list corresponds to what we encountered as SUBCAT list in this chapter. It contains the arguments of a head in an order corresponding to the Obliqueness Hierarchy. The elements of the list are linked to argument roles in the semantic content of the head (see Section 9.1.6). Binding Theory operates on the ARG-ST list. This level of representation is probably the same for most languages: In every language there are semantic predicates and semantic arguments. Most languages make use of syntactic categories that play a role in selection, so there is both syntactic and semantic selection.¹⁹ Languages differ with regard to how these arguments are realized. In English, the first

¹⁹ Koenig & Michelson (2012) argue for an analysis of Oneida (a Northern Iroquoian language) that does not include a representation of syntactic valence. If this analysis is correct, syntactic argument structure would not be universal, but would be characteristic for a large number of languages.

element in the valence list is mapped to the SPR list and the remaining arguments in the SUBCAT or COMPS list. In German, the SPR list of verbs remains empty. (58) shows some relevant examples for German and English.

$$\begin{array}{l}
 (58) \quad \text{a.} \quad \left[\begin{array}{c} \text{PHON } \langle \textit{schlag} \rangle \\ \\ \text{SYNSEM|LOC} \end{array} \left[\begin{array}{c} \text{CAT} \\ \\ \text{CONT} \end{array} \left[\begin{array}{c} \left[\begin{array}{c} \text{HEAD } \textit{verb} \\ \text{SPR } \langle \rangle \\ \text{SUBCAT } [1] \\ \text{ARG-ST } [1] \langle \text{NP}[\textit{str}]_{[2]}, \text{NP}[\textit{str}]_{[3]} \rangle \end{array} \right] \\ \left[\begin{array}{c} \text{IND } [4] \textit{event} \\ \text{RELS } \langle \begin{array}{c} \text{EVENT } [4] \\ \text{AGENT } [2] \\ \text{PATIENT } [3] \\ \textit{schlagen} \end{array} \rangle \end{array} \right] \end{array} \right] \right] \right] \\
 \\
 \text{b.} \quad \left[\begin{array}{c} \text{PHON } \langle \textit{beat} \rangle \\ \\ \text{SYNSEM|LOC} \end{array} \left[\begin{array}{c} \text{CAT} \\ \\ \text{CONT} \end{array} \left[\begin{array}{c} \left[\begin{array}{c} \text{HEAD } \textit{verb} \\ \text{SPR } \langle [1] \rangle \\ \text{SUBCAT } [2] \\ \text{ARG-ST } \langle [1] \text{NP}[\textit{str}]_{[3]} \rangle \oplus [2] \langle \text{NP}[\textit{str}]_{[4]} \rangle \end{array} \right] \\ \left[\begin{array}{c} \text{IND } [5] \textit{event} \\ \text{RELS } \langle \begin{array}{c} \text{EVENT } [5] \\ \text{AGENT } [3] \\ \text{PATIENT } [4] \\ \textit{beat} \end{array} \rangle \end{array} \right] \end{array} \right] \right] \right]
 \end{array}$$

One can view the ARG-ST list as the equivalent to Deep Structure in GB theory: Semantic roles are assigned with reference to this list. The difference is that there is no ordered tree that undergoes transformations. The question of whether all languages can be derived from either VO or OV order therefore becomes irrelevant.

9.7 Summary and classification

In HPSG, feature descriptions are used to model all properties of linguistic objects: roots, words, lexical rules and dominance schemata are all described using the same formal tools. Unlike GPSG and LFG, there are no separate phrase structure rules. Thus, although HPSG stands for Head-Driven Phrase Structure Grammar, it is not a phrase structure grammar. In HPSG implementations, a phrase structure component is often used to increase the efficiency of processing. However, this is not part of the theory and not linguistically necessary.

HPSG differs from Categorical Grammar in that it assumes considerably more features and that how these features are grouped plays an important role for the theory.

Long-distance dependencies are not analyzed using functional composition as in Categorical Grammar, but instead as in GPSG by appealing to the percolation of information in the tree. In this way, it is possible to analyze pied-piping constructions such as those discussed in Section 8.6 with just one lexical item per relative pronoun, whose relevant local properties are identical to those of the demonstrative pronoun. The relative clause in (59) would be analyzed as a finite clause from which a PP is extracted:

- (59) der Mann, [_{RS} [_{PP} an den] [_{S/PP} wir gedacht haben]]
 the man on who we thought have
 ‘the man we thought of’

It is required of relative clauses that the first daughter contains a relative pronoun. This can, as shown in the English examples on page 254, be in fact very deeply embedded. Information about the fact that *an den* contains a relative pronoun is situated in the lexical entry for the relative pronoun *den* by specifying the value of `NONLOC|INHER|REL`. The Nonlocal Feature Principle passes this information on upwards so that the information about the relative pronoun is contained in the representation of the phrase *an den*. This information is bound off when the relative clause is put together (Pollard & Sag 1994: Chapter 5; Sag 1997). It is possible to use the same lexical entry for *den* in the analyses of both (59) and (60) as – unlike in Categorical Grammar – the relative pronoun does not have to know anything about the contexts in which it can be used.

- (60) der Mann, [_{RS} [_{NP} den] [_{S/NP} wir kennen]]
 the man that we know
 ‘the man that we know’

Every theory that wants to maintain the analysis sketched here will have to have some mechanism to make information available about the relative pronoun in a complex phrase. If we have such a mechanism in our theory – as is the case in LFG and HPSG – then we can simply use it for the analysis of long-distance dependencies. Theories such as LFG and HPSG are therefore more parsimonious with their descriptive tools than other theories when it comes to the analysis of relative phrases.

In the first decade of HPSG history (Pollard & Sag 1987; 1994; Nerbonne, Netter & Pollard 1994), despite the differences already mentioned here, HPSG was still very similar to Categorical Grammar in that it was a strongly lexicalized theory. The syntactic make-up and semantic content of a phrase was determined by the head (hence the term *head-driven*). In cases where head-driven analyses were not straight-forwardly possible, because no head could be identified in the phrase in question, then it was commonplace to assume empty heads. An example of this is the analysis of relative clauses in Pollard & Sag (1994: Chapter 5). Since an empty head can be assigned any syntactic valency and an arbitrary semantics (for discussion of this point, see Chapter 21), one has not really explained anything as one needs very good reasons for assuming an empty head, for example that this empty position can be realized in other contexts. This is, however, not

the case for empty heads that are only proposed in order to save theory assumptions. Therefore, Sag (1997) developed an analysis of relative clauses without any empty elements. As in the analyses sketched for (59) and (60), the relative phrases are combined directly with the partial clause in order to form the relative clause. For the various observable types of relative clauses in English, Sag proposes difference dominance rules. His analysis constitutes a departure from strong lexicalism: In Pollard & Sag (1994), there are six dominance schemata, whereas there are 23 in Ginzburg & Sag (2000).

It is possible to observe similar tendencies in the proceedings of recent conferences. The proposals range from the elimination of empty elements to radically phrasal analyses (Haugereid 2007; 2009).²⁰

Even when this tendency towards phrasal analyses may result in some problematic analysis, it is still the case that there are areas of grammar where phrasal analyses are required (see Section 23.10). For HPSG, this means that it is no longer entirely head-driven and is therefore neither Head-Driven nor Phrase Structure Grammar.

HPSG makes use of typed feature descriptions to describe linguistic objects. Generalizations can be expressed by means of hierarchies with multiple inheritance. Inheritance also plays an important role in Construction Grammar. In theories such as GPSG, Categorical Grammar and TAG, it does not form part of theoretical explanations. In implementations, macros (abbreviations) are often used for co-occurring feature-value pairs (Dalrymple, Kaplan & King 2004). Depending on the architecture assumed, such macros are not suitable for the description of phrases since, in theories such as GPSG and LFG, phrase structure rules are represented differently from other feature-value pairs (however, see Asudeh, Dalrymple & Toivonen (2008; 2013) for macros and inheritance used for c-structure annotations). Furthermore, there are further differences between types and macros, which are of a more formal nature: In a typed system, it is possible under certain conditions to infer the type of a particular structure from the presence of certain features and of certain values. With macros, this is not the case as they are only abbreviations. The consequences for linguistic analyses made by this differences are, however, minimal.

HPSG differs from GB theory and later variants in that it does not assume transformations. In the 80s, representational variants of GB were proposed, that is, it was assumed that there was no D-Structure from which a S-structure is created by simultaneous marking of the original position of moved elements. Instead, one assumed the S-Structure straight away with traces and the assumption that there were further movements in the mapping of S-Structure to Logical Form was also abandoned (Koster 1978; Haider 1993: Section 1.4; Frey 1993: 14). This view corresponds to the view in HPSG and many of the analyses in one framework can be translated into the other.

In GB theory, the terms subject and object do not play a direct role. Although one can use these terms descriptively, these concepts correspond to the positions in the tree (the subject in specifier position of IP and the object as the complement of the verb). In HPSG, subject and object are also not primitives of the theory. Since valence lists (or ARG-ST lists) are ordered, however, this means that it is possible to derive a link. If there

²⁰ For discussion, see Müller (2007c) and Section 23.3.6.

is an subject, this occurs in the first position of the valence list and objects follow.²¹ For the analysis of (61b) in a transformation-based grammar, the aim is to connect the base order in (61a) and the derived order in (61b). Once one has recreated the base order, then it is clear what is the subject and what is the object. Therefore, transformations applied to the base structure in (61a) have to be reversed.

- (61) a. [weil] jeder diesen Mann kennt
 because everyone this man knows
- b. [weil] diesen Mann jeder kennt
 because this man everyone knows
 ‘because everyone knows this man’

In HPSG and also in other transformation-less models, the aim is to assign arguments in the order in (61b) to descriptions in the valence list. The valence list (or ARG-ST in newer approaches) corresponds in a sense to Deep Structure in GB. The difference is that the head itself is not included in the argument structure, whereas this is the case with D-Structure.

Bender (2008c) has shown how one can analyze phenomena from non-configurational languages such as Wambaya by referring to the argument structure of a head. In Wambaya, words that would normally be counted as constituents in English or German can occur discontinuously, that is an adjective that belongs to a noun phrase and shares the same case, number and gender values with other parts of the noun phrase can occur in another position in the sentence. Nordlinger (1998) has analyzed the relevant data in LFG. In her analysis, the various parts of the constituent refer to the f-structure of the sentence and thus indirectly ensure that all parts of the noun phrase have the same case. Bender adopts a variant of HPSG where valency information is not removed from the valence list after an argument has been combined with its head, but rather this information remains and is passed up towards the maximal projection of the head (Meurers 1999c; Przepiórkowski 1999b; Müller 2007b: Section 17.4). Similar proposals were made in GB by Higginbotham (1985: 560) and Winkler (1997). By projecting the complete valency information, it remains available in the entire sentence and discontinuous constituents can refer to it (e. g. via MOD) and the respective constraints can be formulated.²² In this analysis, the argument structure in HPSG corresponds to f-structure in LFG. The extended head domains of LFG, where multiple heads can share the same f-structure can also be modeled in HPSG. To this end, one can utilize functional composition as it was presented in the chapter on Categorical Grammar (see Chapter 8.5.2). The exact way in which this is translated into HPSG, cannot be explained here due to space restrictions. The reader is referred to the original works by Hinrichs & Nakazawa (1994) and the explanation in Müller (2007b: Chapter 15).

²¹ When forming complex predicates, an object can occur in first position. See Müller (2002a: 157) for the long passive with verbs such as *erlauben* ‘allow’. In general, the following holds: The subject is the first argument with structural case.

²² See Müller (2008) for an analysis of depictive predicates in German and English that make reference to the list of realized or unrealized arguments of a head.

Valency information plays an important role in HPSG. The lexical item of a verb in principle predetermines the set of structures in which the item can occur. Using lexical rules, it is possible to connect one lexical item to other lexical items. These can be used in other sets of structures. So one can see the functionality of lexical rules in establishing a relation between sets of possible structures. Lexical rules correspond to transformations in Transformational Grammar. This point is discussed in more detail in Section 21.5. The effects of lexical rules can also be achieved with empty elements. This will also be the matter of discussion in Section 21.5.

In GPSG, metarules were used to license rules that created additional valence grids for lexical heads. In principle, metarules could also be applied to rules without a lexical head. This is explicitly ruled out by Flickinger (1983) and Gazdar et al. (1985: 59) using a special constraint. Flickinger, Pollard & Wasow (1985: 265) pointed out that this kind of constraint is unnecessary if one uses lexical rules rather than metarules since the former can only be applied to lexical heads.

For a comparison of HPSG and Stabler's Minimalist Grammars, see Section 4.6.4.

Comprehension questions

1. What status do syntactic trees have in HPSG?
2. How does case assignment take place in the analysis of example (62)?

(62) Dem Mann wurde ein Buch geschenkt.
the.DAT man was a.NOM book given
'The man was given a book'

3. What is *linking* and how is it accounted for in HPSG?

Exercises

1. Give a feature description for (63) ignoring *dass*.

(63) [dass] Max lacht
that Max laughs

2. The analysis of the combination of a noun with a modifying adjective in Section 9.1.7 was just a sketch of an analysis. It is, for example, not explained how one can ensure that the adjective and noun agree in case. Consider how it would be possible to expand such an analysis so that the adjective-noun combination in (64a) can be analyzed, but not the one in (64b):

(64) a. eines interessanten Mannes
an.GEN interesting.GEN man.GEN
b. *eines interessanter Mannes
an.GEN interesting.NOM man.GEN

Further reading

Here, the presentation of the individual components of the verb was – as with other theories – kept relatively short. For a more comprehensive introduction to HPSG, including motivation for feature geometry, see Müller (2007b). In particular, the analysis of the passive was sketched in brief here. The entire story including the analysis of unaccusative verbs, adjectival participles, modal infinitives as well as diverse passive variants and the long passive can be found in Müller (2002a: Chapter 3) and Müller (2007b: Chapter 17).

Overviews of HPSG can be found in Levine & Meurers (2006), Przepiórkowski & Kupść (2006) and Müller (2015c).

10 Construction Grammar

Like LFG and HPSG, *Construction Grammar* (CxG) belongs to West Coast linguistics. It has been considerably influenced by Charles Fillmore, Paul Kay and George Lakoff (all three at Berkeley) and Adele Goldberg (completed her PhD in Berkeley and is now in Princeton) (Fillmore 1988; Fillmore, Kay & O'Connor 1988; Kay & Fillmore 1999; Kay 2002; 2005; Goldberg 1995; 2006).

Fillmore, Kay, Jackendoff and others have pointed out the fact that languages consist to a large extent of complex units that cannot straightforwardly be described with the tools that we have seen thus far. In frameworks such as GB, an explicit distinction is made between core grammar and the periphery (Chomsky 1981a: 8), whereby the periphery is mostly disregarded as uninteresting when formulating a theory of Universal Grammar. The criticism leveled at such practices by CxG is justified since what counts as the 'periphery' sometimes seems completely arbitrary (Müller 2014d) and no progress is made by excluding large parts of the language from the theory just because they are irregular to a certain extent.

In Construction Grammar, idiomatic expressions are often discussed with regard to their interaction with regular areas of grammar. Kay & Fillmore (1999) studied the *What's X doing Y?*-construction in their classic essay. (1) contains some examples of this construction:

- (1) a. What is this scratch doing on the table?
- b. What do you think your name is doing in my book?

The examples show that we are clearly not dealing with the normal meaning of the verb *do*. As well as the semantic bleaching here, there are particular morphosyntactic properties that have to be satisfied in this construction. The verb *do* must always be present and also in the form of the present participle. Kay and Fillmore develop an analysis explaining this construction and also capturing some of the similarities between the WXDY-construction and the rest of the grammar.

There are a number of variants of Construction Grammar:

- Berkeley Construction Grammar (Fillmore 1988; Kay & Fillmore 1999; Fried 2015)
- Goldbergian/Lakovian Construction Grammar (Lakoff 1987; Goldberg 1995; 2006)
- Cognitive Grammar (Langacker 1987; 2000; 2008; Dąbrowska 2004)
- Radical Construction Grammar (Croft 2001)
- Embodied Construction Grammar (Bergen & Chang 2005)

- Fluid Construction Grammar (Steels & Beule 2006; Steels 2011)
- Sign-Based Construction Grammar (Sag 2010; 2012)

The aim of Construction Grammar is to describe and theoretically explore language in its entirety. In practice, however, the irregularities in language are often given far more importance than the phenomena described as ‘core grammar’ in GB. Construction Grammar analyses usually analyze phenomena as phrasal patterns. These phrasal patterns are represented in inheritance hierarchies (e.g. Croft 2001; Goldberg 2003b). An example for the assumption of a phrasal construction is Goldberg’s analysis of resultative constructions. Goldberg (1995) and Goldberg & Jackendoff (2004) argue for the construction status of resultative constructions. In their view, there is no head in (2) that determines the number of arguments.

(2) Willy watered the plants flat.

The number of arguments is determined by the construction instead, that is, by a rule or schema saying that the subject, verb, object and a predicative element must occur together and that the entire complex has a particular meaning. This view is fundamentally different from analyses in GB, Categorical Grammar, LFG¹ and HPSG. In the aforementioned theories, it is commonly assumed that arguments are always arguments of lexical heads and not independently licensed by phrasal rules. See Simpson (1983), Neeleman (1994), Wunderlich (1997), Wechsler (1997), and Müller (2002a) for corresponding work in LFG, GB, Wunderlich’s Lexical Decomposition Grammar and HPSG.

Like the theories discussed in Chapters 5–9, CxG is also a non-transformational theory. Furthermore, no empty elements are assumed in most variants of the theory and the assumption of lexical integrity is maintained as in LFG and HPSG. It can be shown that these assumptions are incompatible with phrasal analyses of resultative constructions (see Section 23.2.2 and Müller (2006; 2007c)). This point will not be explained further here. Instead, I will discuss the work of Fillmore and Kay to prepare the reader to be able to read the original articles and subsequent publications. Although the literature on Construction Grammar is now relatively vast there is very little work on the basic formal assumptions or analyses that have been formalized precisely. Examples of more formal works are Kay & Fillmore (1999); Kay (2002), Michaelis & Ruppenhofer (2001), Goldberg (2003b). The book by Jean-Pierre Koenig (1999) (formerly Berkeley) has been heavily influenced by CxG. Fillmore and Kay’s revisions of their earlier works took place in close collaboration with Ivan Sag. The result was a variant of HPSG known as Sign-Based Construction Grammar (SB-CxG) (Sag 2010; 2012). See Section 10.6.2 for further discussion.

John Bryant, Nancy Chang, Eva Mok have developed a system for the implementation of Embodied Construction Grammar². Luc Steels is working on the simulation of language evolution and language acquisition (Steels 2003). In personal communication

¹ See Alsina (1996) and Asudeh, Dalrymple & Toivonen (2008; 2013), however. For more discussion of this point, see Sections 23.1.3 and 23.2.2.

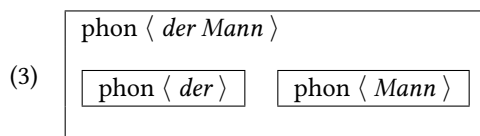
² See <http://www.icsi.berkeley.edu/~jbryant/old-analyzer.html> and Bryant (2003).

(p. M. 2007) he stated that it is a long way to go until robots finally will be able to learn to speak but the current state of the art is already impressive. Steels can use robots that have a visual system (camera and image processing) and use visual information paired with audio information in simulations of language acquisition. The implementation of Fluid Construction Grammar is documented in Steels (2011) and Steels (2012). The second book contains parts about German, in which the implementation of German declarative clauses and *w* interrogative clauses is explained with respect to topological fields (Micelli 2012). The FCG system, various publications and example analyses are available at: <http://www.fcg-net.org/>. Jurafsky (1996) developed a Construction Grammar for English that was paired with a probabilistic component. He showed that many preformance phenomena that are discussed in the literature (see Chapter 15 on the Competence/Preformance Distinction) can be explained with recourse to probabilities of phrasal constructions and valency properties of words. Bannard, Lieven & Tomasello (2009) use a probabilistic context free grammar to model grammatical knowledge of two and three year old children.

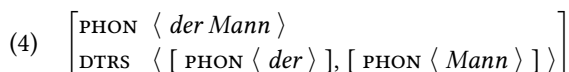
10.1 General remarks on the representational format

In this section I will discuss the mechanisms of Berkeley Construction Grammar (BCG). As I pointed out in Müller (2006) there are fundamental problems with the formalization of BCG. The details will be given in Section 10.6.1. While the framework was developed further by its creators Kay and Fillmore into Sign-Based Construction Grammar (see Section 10.6.2), there are still authors working in that framework. I therefore present the basic mechanisms here to make it possible to understand the original ideas and put them into a broader context.

As we saw in Section 9.1.2, dominance relations in HPSG are treated like all other linguistic objects using feature-value pairs. In general, CxG uses feature-value pairs to describe linguistic objects, but dominance relations are represented by boxes (Kay & Fillmore 1999; Goldberg 2003b):



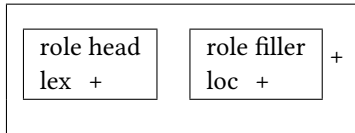
The structure can be written using feature-value pairs as follows:



10.1.1 The head-complement construction

Kay & Fillmore (1999) assume the following construction for the combination of heads with their complements:

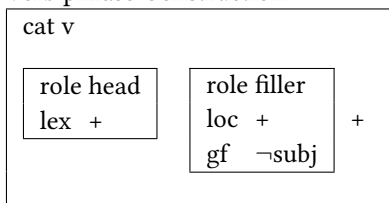
(5) Head plus Complements Construction (HC)



A head is combined with at least one complement (the '+' following the box stands for at least one sign that fits the description in that box). LOC+ means that this element must be realized locally. The value of ROLE tells us something about the role that a particular element plays in a construction. Unfortunately, here the term *filler* is used somewhat differently than in GPSG and HPSG. Fillers are not necessarily elements that stand in a long-distance dependency to a gap. Instead, a *filler* is a term for a constituent that fills the argument slot of a head.

The verb phrase construction is a sub-construction of the head-complement construction:

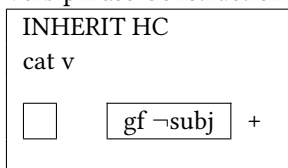
(6) Verb phrase Construction:



The syntactic category of the entire construction is V. Its complements cannot have the grammatical function subject.

The VP construction is a particular type of head-complement construction. The fact that it has much in common with the more general head-complement construction is represented as follows:

(7) Verb phrase Construction with inheritance statement:



This representation differs from the one in HPSG, aside from the box notation, only in the fact that feature descriptions are not typed and as such it must be explicitly stated in the representation from which superordinate construction inheritance takes place. In HPSG, there are type hierarchies as well as schemata, which both say something about the inheritance relation between two types.

10.1.2 Representation of valency information

In Kay and Fillmore, valency information is represented in a set (*VAL*). The Valency Principle states that local filler-daughters have to be identified with an element in the valency set of the mother.³ The Subset Principle states that the set values of the head-daughter are subsets of the corresponding sets of the mother. This is the exact opposite approach to the one taken in Categorical Grammar and HPSG. In HPSG grammars, valence lists at the mother nodes are shorter, whereas in Berkeley CxG at least as many elements are present on the mother node as on the head-daughter.

10.1.3 Semantics

Semantics in CxG are handled exactly as in HPSG: semantic information is contained in the same feature structure as syntactic information. The relation between syntax and semantics is captured by using the same variable in the syntactic and semantic description. (8) contains a feature description for the verb *arrive*:

(8) Lexical entry for *arrive* following Kay & Fillmore (1999: 11):

cat	v
sem	$\left\{ \left[\begin{array}{cc} \text{FRAME} & \text{ARRIVE} \\ \text{ARGS} & \{ A \} \end{array} \right] \right\}$
val	$\{ [\text{SEM} \{ A \}] \}$

Kay & Fillmore (1999: 9) refer to their semantic representations as a notational variant of the Minimal Recursion Semantics of Copestake, Flickinger, Pollard & Sag (2005). In later works, Kay (2005) explicitly uses MRS. As the fundamentals of MRS have already been discussed in Section 9.1.6, I will not repeat them here. For more on MRS, see Section 21.3.

10.1.4 Adjuncts

For the combination of heads and modifiers, Kay and Fillmore assume further phrasal constructions that correspond to the verb phrase constructions discussed above and create a relation between a head and a modifier. Kay and Fillmore assume that adjuncts also contribute something to the *VAL* value of the mother node. In principle, *VAL* is nothing more than the set of all non-head daughters in a tree.

10.2 Passive

The passive has been described in CxG by means of so-called linking constructions, which are combined with lexical entries in inheritance hierarchies. In the base lexicon,

³ Sets in BCG work differently from those used in HPSG. A discussion of this is deferred to Section 10.6.1.

it is only listed which semantic roles a verb fulfills and how these are realized are determined by the exact linking constructions with which the basic lexical entry is combined. Figure 10.1 gives an example of a relevant inheritance hierarchy. There is a linking con-

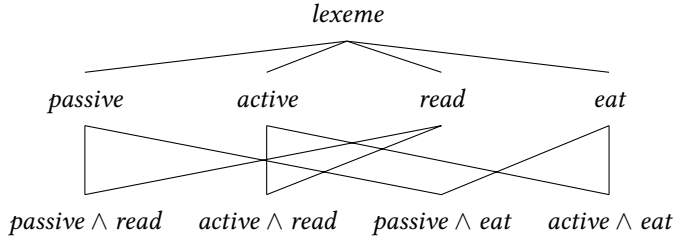


Figure 10.1: Passive and linking constructions

struction for both active and passive as well as lexical entries for *read* and *eat*. There is then a cross-classification resulting in an active variant and a passive variant of each verb.

The idea behind this analysis goes back to works by Fillmore and Kay between 1995 and 1997⁴, but variants of this analysis were first published in Koenig (1999: Chapter 3) and Michaelis & Ruppenhofer (2001: Chapter 4). Parallel proposals have been made in TAG (Candito 1996; Clément & Kinyon 2003: 188; Kallmeyer & Osswald 2012: 171–172) and HPSG (Koenig 1999; Davis & Koenig 2000; Kordoni 2001).

Michaelis & Ruppenhofer (2001: 55–57) provide the following linking constructions:⁵

(9) a. *Transitive Construction:*

$$\left[\begin{array}{l} \text{SYN} \left[\begin{array}{ll} \text{CAT} & v \\ \text{VOICE} & \text{active} \end{array} \right] \\ \text{VAL} \left\{ \left[\begin{array}{l} \text{ROLE} \left[\begin{array}{ll} \text{GF} & \text{obj} \\ \text{DA} & - \end{array} \right] \end{array} \right] \right\} \end{array} \right]$$

b. *the Subject Construction:*

$$\left[\begin{array}{l} \text{SYN} \left[\begin{array}{ll} \text{CAT} & v \end{array} \right] \\ \text{VAL} \left\{ \left[\begin{array}{l} \text{ROLE} \left[\begin{array}{ll} \text{GF} & \text{subj} \end{array} \right] \end{array} \right] \right\} \end{array} \right]$$

⁴ <http://www.icsi.berkeley.edu/~kay/bcg/ConGram.html>. 03.05.2010.

⁵ In the original version of the transitive construction in (9a), the θ value is DA—, however, DA is a feature. I have corrected this in (9a) accordingly.

In the following structures, GF stands for *grammatical function* and DA for *distinguished argument* (designated argument). The designated argument usually corresponds to the subject in an active clause.

c. the *Passive Construction*:

$$\left[\begin{array}{l} \text{SYN} \left[\begin{array}{l} \text{CAT } v \\ \text{FORM } \textit{PastPart} \end{array} \right] \\ \text{VAL} \left\{ \left[\begin{array}{l} \text{ROLE} \left[\begin{array}{l} \text{GF } \textit{obl} \\ \text{DA } + \end{array} \right] \\ \text{SYN } \textit{P[von]/zero} \end{array} \right] \right\} \end{array} \right]$$

The structure in (9a) says that, in the valency set of one of the linguistic objects described by the transitive construction, there has to be an element that has the grammatical function *object* and whose DA value is ‘–’. The DA value of the argument that would be the subject in an active clause is ‘+’ and ‘–’ for all other arguments. The subject construction states that an element of the valency set must have the grammatical function *subject*. In the passive construction, there has to be an element with the grammatical function *oblique* that also has the DA value ‘+’. In the passive construction the element with the DA value ‘+’ is realized either as a *by*-PP or not at all (*zero*).

The interaction of the constructions in (9) will be explained on the basis of the verb *schlagen* ‘to hit’:

(10) Lexical entry for *schlag-* ‘hit’:

$$\left[\begin{array}{l} \text{SYN} \left[\begin{array}{l} \text{CAT } v \end{array} \right] \\ \text{VAL} \left\{ \left[\begin{array}{l} \text{ROLE} \left[\begin{array}{l} \theta \textit{ agent} \\ \text{DA } + \end{array} \right] \right], \left[\begin{array}{l} \text{ROLE} \left[\begin{array}{l} \theta \textit{ patient} \end{array} \right] \end{array} \right] \right\} \end{array} \right]$$

If we combine this lexical entry with the transitive and subject constructions, we arrive at (11a) following Fillmore/Kay/Michaelis and Ruppenhofer, whereas combining it with the subject and passive construction yields (11b):⁶

(11) a. *schlag-* + Subject and Transitive Construction:

$$\left[\begin{array}{l} \text{SYN} \left[\begin{array}{l} \text{CAT } v \\ \text{VOICE } \textit{active} \end{array} \right] \\ \text{VAL} \left\{ \left[\begin{array}{l} \text{ROLE} \left[\begin{array}{l} \theta \textit{ agent} \\ \text{GF } \textit{subj} \\ \text{DA } + \end{array} \right] \right], \left[\begin{array}{l} \text{ROLE} \left[\begin{array}{l} \theta \textit{ patient} \\ \text{GF } \textit{obj} \\ \text{DA } - \end{array} \right] \end{array} \right] \right\} \end{array} \right]$$

⁶ This assumes a particular understanding of set unification. For criticism of this, see below.

b. *schlag-* + Subject and Passive Construction:

$$\left[\begin{array}{c} \text{SYN} \left[\begin{array}{c} \text{CAT } v \\ \text{FORM } \textit{PastPart} \end{array} \right] \\ \text{VAL} \left\{ \left[\begin{array}{c} \text{ROLE} \left[\begin{array}{c} \theta \text{ agent} \\ \text{GF } \textit{obl} \\ \text{DA } + \end{array} \right] \\ \text{SYN } \textit{P[von]/zero} \end{array} \right], \left[\begin{array}{c} \text{ROLE} \left[\begin{array}{c} \theta \text{ patient} \\ \text{GF } \textit{subj} \end{array} \right] \end{array} \right] \right\} \end{array} \right]$$

Using the entries in (11), it is possible to analyze the sentences in (12):

- (12) a. Er schlägt den Weltmeister.
 he beat the world.champion
 ‘He beat the world champion.’
 b. Der Weltmeister wird (von ihm) geschlagen.
 the world.champion is from him beaten
 ‘The world champion was beaten (by him).’

This analysis is formally inconsistent as set unification cannot be formalized in such a way that the aforementioned constructions can be unified (Müller 2006; Müller 2007b: Section 7.5.2). It is, however, possible to fix this analysis by using a HPSG formalism of sets (Pollard & Sag 1987; Pollard & Moshier 1990). The Subject, Transitive and Passive Constructions must then be modified in such a way that they can say something about what an element in VAL looks like, rather than specifying the VAL value of a singleton set.

(13) The *Subject Construction* with Pollard and Moschier’s definition of sets:

$$\left[\begin{array}{c} \text{SYN} | \text{CAT } v \\ \text{VAL } \boxed{1} \end{array} \right] \wedge \left\{ \left[\begin{array}{c} \text{ROLE} \left[\begin{array}{c} \text{GF } \textit{subj} \end{array} \right] \end{array} \right] \right\} \subset \boxed{1}$$

The restriction in (13) states that in the valency set of a head there has to be an element that has the grammatical function *subj*. By these means, it is possible to suppress arguments (by specifying SYN as *zero*). It is not possible to add any additional arguments to the fixed set of arguments of *schlagen* ‘to beat’.⁷ For the analysis of Middle Constructions such as (14), inheritance-based approaches do not work as there is no satisfactory way to list the reflexive pronoun in the valency set:⁸

⁷ One could however assume that the constraint on the main lexical item would be of the kind in (11). One would then require that *schlagen* has at least the two members in its valency set. This would considerably complicate everything and furthermore it would not be clear that the subject referred to in (14) would be one of the arguments that are referred to in the description of the lexical item for *schlagen*.

⁸ One technically possible solution would be the following: one could assume that verbs that occur in middle constructions always have a description of a reflexive pronoun in their valency set. The Transitive Construction would then have to specify the SYN value of the reflexive pronoun as *zero* so that the additional reflexive pronoun is not realized in the Transitive Construction. The middle construction would suppress the subject, but realizes the object and the reflexive.

- (14) Das Buch liest sich gut.
 the book reads REFL good
 ‘The book reads well / is easy to read.’

If we wish to introduce additional arguments, we require auxiliary features. This kind of analysis has been suggested by Koenig (1999). Since there are many argument structure changing processes that interact in various ways and are linked to particular semantic side-effects, it is inevitable that one will end up assuming a large number of syntactic and semantic auxiliary features. The interaction between the various linking constructions becomes so complex that this analysis also becomes cognitively implausible and has to be viewed as technically unusable. For a more detailed discussion of this point, see Müller (2007b: Section 7.5.2).

The following empirical problem is much more serious: Some processes like passivization, impersonalization and causativization can be applied in combination or even allow for iteration, but if the grammatical function of a particular argument is determined once and for all by unification, additional unifications cannot change the initial assignment. We will first look at languages which allow for a combination of passivization and impersonalization, such as Lithuanian (Timberlake 1982: Section 5), Irish (Noonan 1994), and Turkish (Özkaragöz 1986; Knecht 1985: Section 2.3.3). The cited authors discussed their data as instances of double passivization, but it was argued by Blevins (2003) that these and similar examples from other languages are impersonal constructions that can be combined with personal passives. I will use Özkaragöz’s Turkish examples in (15) for illustration (1986: 77):

- (15) a. Bu şato-da boğ-ul-un-ur.
 this château-LOC strangle-PASS-PASS-AOR
 ‘One is strangled (by one) in this château.’
 b. Bu oda-da döv-ül-ün-ür.
 this room-LOC hit-PASS-PASS-AOR
 ‘One is beaten (by one) in this room.’
 c. Harp-te vur-ul-un-ur.
 war-LOC shoot-PASS-PASS-AOR
 ‘One is shot (by one) in war.’

-In, *-n*, and *-Il* are allomorphs of the passive morpheme. According to Özkaragöz the data is best captured by an analysis that assumes that the passive applies to a passivized transitive verb and hence results in an impersonal passive. The cited authors discussed their data as instances of double passivization, but it was argued by Blevins (2003) that these and similar examples from other languages are impersonal constructions that can be combined with personal passives.

This solution cannot be applied to the recursive processes we will encounter in a moment such as causativization in Turkish, unless one wishes to assume infinite valency sets.

Approaches that assume that the personal passive is the unification of some general structure with some passive specific structure will not be able to capture double passivization or passivization + impersonalization, since they committed to a certain structure too early. The problem for nontransformational approaches that state syntactic structure for the passive is that such a structure, once stated, cannot be modified. That is, we said that the underlying object is the subject in the passive sentence. But in order to get the double passivization/passivization + impersonalization, we have to suppress this argument as well. What is needed is some sort of process (or description) that takes a representation and relates it to another representation with a suppressed subject. This representation is related to a third representation which again suppresses the subject resulting in an impersonal sentence. In order to do this one needs different strata as in Relational Grammar (Timberlake 1982; Özkaragöz 1986), metarules (Gazdar, Klein, Pullum & Sag 1985), lexical rules (Dowty, 1978: 412; 2003: Section 3.4; Bresnan 1982b; Pollard & Sag 1987; Blevins 2003; Müller 2003b), transformations (Chomsky 1957), or just a morpheme-based morphological analysis that results in items with different valence properties when the passivization morpheme is combined with a head (Chomsky 1981a).

The second set of problematic data that will be discussed comes from causativization in Turkish (Lewis 1967):

- (16) Öl-dür-t-tür-t-
 'to cause someone to cause someone to kill someone'
 (kill = cause someone to die)

The causative morpheme *-t* is combined three times with the verb (*tür* is an allomorph of the causative morpheme). This argument structure-changing process cannot be modeled in an inheritance hierarchy, since if we were to say that a word can inherit from the causative construction three times, we would still not have anything different to what we would have had the inheritance via the causative construction applied only once. For this kind of phenomena, we would require rules that relate a linguistic object to another, more complex object, that is, lexical rules, unary rules which change the phonology of a linguistic sign or binary rules that combine a particular sign with a derivational morpheme. These rules can semantically embed the original sign (that is, add *cause* to *kill*).

In general derivational morphology cannot be handled by inheritance as was already pointed out by Krieger & Nerbonne (1993) with respect to cases like *preprepreversion*.

If we assume that changes in valency such as passive, causativization and middles should be described with the same means across languages, then evidence from Lithuanian and Turkish form an argument against inheritance-based analyses of the passive (Müller 2006; 2007c; Müller & Wechsler 2014a).

10.3 Verb position

At present, I only know of one article in the framework of CxG that has dealt with the sentence structure in German. This is the article by Vanessa Micelli (2012), where

she describes a computer implementation of a German grammar in Fluid Construction Grammar. This fragment is restricted to declarative V2-clauses and *wh*-questions. In her analysis, the middle field forms a constituent comprising exactly two constituents (the direct and indirect object). The right sentence bracket and the postfield are empty. Long-distance dependencies are not discussed. It is only possible for arguments of the verb in the left sentence bracket to occur in the prefield. Micelli's work is an interesting starting point but one has to wait and see how the analysis will be modified when the grammar fragment is expanded.

In the following, I will not discuss Micelli's analysis further, but instead explore some of the possibilities for analyzing German sentence structure that are at least possible in principle in a CxG framework. Since there are neither empty elements nor transformations, the GB and HPSG analyses as well as their variants in Categorical Grammar are ruled out. The following options remain:

- an analysis similar to LFG with an optional verb
- an entirely flat analysis as proposed in GPSG
- an analysis with binary branching but variable verb position like that of Steedman (2000: 159)

The different variants of CxG make different assumptions about how abstract constructions can be. In Categorical Grammar, we have very general combinatorial rules, which combine possibly complex signs without adding any meaning of their own. (17) shows an example:

- (17) [[[[Gibt] der Mann] der Frau] das Buch]
 give the man the woman the book
 'Does the man give the woman the book?'

If we do not want these kinds of abstract combinatorial rules, then this analysis must be excluded.

The LFG analysis in Section 7.3 could also be unacceptable on a CxG view as it is assumed in this analysis that *der Mann der Frau das Buch* forms a VP although only three NPs have been combined. CxG has nothing like a theory of extended head domains that was presented in Section 7.3.

Thus, both variants with binary-branching structures are ruled out and only the analysis with flat branching structures remains. Sign-based CxG, which can be viewed as a variant of HPSG, as well as Embodied Construction Grammar (Bergen & Chang 2005: 156) allow for a separation of immediate dominance and linear order so that it would be possible to formulate a construction, which would correspond to the dominance rule in (18) for transitive verbs:⁹

- (18) $S \rightarrow V, NP, NP$

⁹ In principle, this is also Micelli's analysis, but she assumed that the middle field forms a separate constituent.

Here, we have the problem that adjuncts in German can occur between any of the arguments. In GPSG, adjuncts are introduced by metarules. In formal variants of CxG, lexical rules are used but not metarules.¹⁰ If one does not wish to expand the formalism to include metarules, then there are three options remaining:

- Adjuncts are introduced in the lexicon (van Noord & Bouma 1994; Bouma, Malouf & Sag 2001a) and treated as arguments in the syntax,
- Constructions always have slots available for an arbitrary number of adjuncts, or
- Constructions can be discontinuous

Kasper (1994) has proposed an analysis of the first type in HPSG: adjuncts and arguments are combined with the head in a flat structure. This corresponds to the dominance rule in (19), where the position of adjuncts is not stated by the dominance rules.

(19) $S \rightarrow V, NP, NP, Adj^*$

If we want to say something about the meaning of the entire construction, then one has to combine the original construction (transitive, in the above example) with the semantics contributed by each of the adjuncts. These calculations are not trivial and require relational constraints (small computer programs), which should be avoided if there are conceptually simpler solutions to describing a particular phenomenon.

The alternative would be to use discontinuous constructions. Analyses with discontinuous constituents have been proposed in both HPSG (Reape 1994) as well as Embodied Construction Grammar (Bergen & Chang 2005). If we apply Bergen and Chang's analysis to German, the italicized words in (20) would be part of a ditransitive construction.

- (20) *Gibt der Mann morgen der Frau unter der Brücke das Geld?*
 gives the man tomorrow the woman under the bridge the money
 'Is the man going to give the woman the money under the bridge tomorrow?'

The construction has been realized discontinuously and the adjuncts are inserted into the gaps. In this kind of approach, one still has to explain how the scope of quantifiers and adjuncts is determined. While this is possible, the solution is not obvious and has not been worked out in the framework of CxG.

10.4 Local reordering

If we assume flat branching structures, then it is possible to use the GPSG analysis for the order of arguments. However, Kay (2002) assumes a phrasal construction for so-called Heavy-NP-Shift in English, which means that there is a new rule for the reordering

¹⁰ Goldberg (2014: 116) mentions metarule-like devices and refers to Cappelle (2006). The difference between metarules and their CxG variant is that in CxG two constructions are related without one construction being basic and the other one derived. Rather there exists a mutual relation between two constructions.

of heavy NPs in English rather than one rule and two different ways to linearize the daughters.

In CxG, it is often argued that the usage contexts of certain orders differ and we therefore must be dealing with a different construction. Accordingly, one would have to assume six constructions for sentences with ditransitive verbs in final position (also see page 183). An alternative would be to assume a similar structure for the ordering variants and make the information-structural properties dependent on the position of constituents (see De Kuthy (2000) for German and Bildhauer (2008) for Spanish).

10.5 Long-distance dependencies

Kay & Fillmore (1999: Section 3.10) discuss long-distance dependencies in their article. Since the number of arguments is not given in the verb phrase construction, it is possible that an argument of the verb is not locally present. Like the LFG and GPSG-analyses in previous chapters, there are no empty elements assumed for the analysis of long-distance dependencies. In the *Left Isolation Construction* that licenses the entire sentence, there is a left daughter and a right daughter. The left daughter corresponds to whatever was extracted from the right daughter. The connection between the fronted element and the position where it is missing is achieved by the operator VAL. VAL provides all elements of the valency set of a linguistic object as well as all elements in the valency set of this elements and so on. It is thereby possible to have unrestricted access to an argument or adjunct daughter of any depth of embedding, and then identify the fronted constituent with an open valency slot. This approach corresponds to the LFG analysis of Kaplan & Zaenen (1989) based on functional uncertainty.

10.6 New developments and theoretical variants

Section 10.6.1 discusses the formal underpinnings of Berkeley Construction Grammar in more detail and shows that they are not suited for what they were intended to do. The section that follows discusses Sign-Based Construction Grammar, which was developed in joint work by Charles Fillmore, Paul Kay and Ivan Sag. It embodies ideas from BCG without having its formal flaws. Section 10.6.3 deals with Embodied Construction Grammar, which is also based on work by Charles Fillmore, Paul Kay and George Lakoff. Section 10.6.4 deals with Fluid Construction Grammar.

10.6.1 Berkeley Construction Grammar

Section 10.2 discussed the valence representation in BCG and linking constructions for active and passive. Kay & Fillmore (1999) represent valency information in sets and we deferred the discussion of the formal properties of sets in BCG to this section. Fillmore and Kay's assumptions regarding set unification differ fundamentally from those that are made in HPSG. Kay and Fillmore assume that the unification of the set { a } with the set

$\{b\}$, where a and b do not unify, results into the union of the two sets, that is $\{a, b\}$. Due to this special understanding of sets it is possible to increase the number of elements in a set by unification. The unification of two sets that contain compatible elements is the disjunction of sets that contain the respective unifications of the compatible elements. This sounds complicated, but we are only interested in a specific case: the unification of an arbitrary set with a singleton set:

$$(21) \quad \{NP[nom], NP[acc]\} \wedge \{NP[nom]\} = \{NP[nom], NP[acc]\}$$

According to Fillmore/Kay the unification of a set with another set that contains a compatible element does not result in an increase of the number of list elements.

(22) illustrates another possible case:

$$(22) \quad \{NP, NP[acc]\} \wedge \{NP[nom]\} = \{NP[nom], NP[acc]\}$$

The case of the first NP in (22) is underspecified with respect to its case. The case of the NP in the second set is specified as nominative. $NP[nom]$ does not unify with $NP[acc]$ but with NP.

This particular conception of unification has consequences. Unification is usually defined as follows:

- (23) The unification of two structures FS_1 and FS_2 is the structure FS_3 , that is subsumed by both FS_1 and FS_2 where there is no other structure that subsumes FS_1 and FS_2 and is subsumed by FS_3 .

A structure FS_1 is said to subsume FS_3 , iff FS_3 contains all feature value pairs and structure sharings from FS_1 . FS_3 may contain additional feature value pairs or structure sharings. The consequence is that the subsumption relations in (24b,c) have to hold, if unification of valence sets works as in (24a):

- (24) Properties of the set unification according to Kay & Fillmore (1999):
- a. $\{NP[nom]\} \wedge \{NP[acc]\} = \{NP[nom], NP[acc]\}$
 - b. $\{NP[nom]\} \succeq \{NP[nom], NP[acc]\}$
 - c. $\{NP[acc]\} \succeq \{NP[nom], NP[acc]\}$

(24b) means that a feature structure with a valency set that contains just one $NP[nom]$ is more general than a feature structure that contains both an $NP[nom]$ and an $NP[acc]$. Therefore the set of transitive verbs is a subset of the intransitive verbs. This is rather unintuitive, but compatible with Fillmore/Kay's system for the licensing of arguments. However, there are problems that result in the interaction of valence specifications with linking constructions.

We have seen the results of combining lexical items with linking constructions, but the question how these are derived has not been addressed so far. Kay (2002) suggests an automatic computation of all compatible combinations of maximally specific constructions. Such a procedure could be used to compute the representations we saw in Section 10.2 and the wellformed sentences in (12) could be analyzed.

However, problems would result for ungrammatical sentences like (25b). *grauen* ('to dread') is a subjectless verb. If one would simply combine all compatible linking constructions with *grauen*, the Kay/Fillmoreian conception of set unification would cause the introduction of a subject into the valency set of *grauen*. (25b) would be licensed by the grammar:

- (25) a. Dem Student graut vor der Prüfung.
 the.DAT student dreads before the exam
 'The student dreads the exam.'
- b. * Ich graue dem Student vor der Prüfung.
 I dread the.DAT student before the exam

One could solve this problem by specifying an element with the grammatical function *subject* in the lexical entry of *grauen*. In addition it would have to be stipulated that this subject can only be realized as an overt or covert expletive (SYN zero). For the covert expletive, this means it neither has a form nor a meaning. Such expletive pronouns without phonological realization are usually frowned upon in Construction Grammar and analyses that can do without such abstract entities are to be preferred.

Kay & Fillmore (1999) represent the semantic contribution of signs as sets as well. This excludes the possibility to prevent the unwanted unification of linking constructions by referring to semantic constraints, since we have the same effect as we have with valency sets: If the semantic descriptions are incompatible, the set is extended. This means that in an automatic unification computation all verbs are compatible with the Transitive Construction in (9a) and this would license analyses for (26) in addition to those of (25b).

- (26) a. * Der Mann schläft das Buch.
 the man sleeps the book
- b. * Der Mann denkt an die Frau das Buch.
 the man thinks at the woman the book

An intransitive verb was unified with the Transitive Construction in the analysis of (26a) and in (26b) a verb that takes a prepositional object was combined with the Transitive Construction. This means that representations like (11) cannot be computed automatically as was intended by Kay (2002). Therefore one would have to specify subconstructions for all argument structure possibilities for every verb (active, passive, middle, ...). This does not capture the fact that speakers can form passives after acquiring new verbs without having to learn about the fact that the newly learned verb forms one.

Michaelis & Ruppenhofer (2001) do not use sets for the representation of semantic information. Therefore they could use constraints regarding the meaning of verbs in the Transitive Construction. To this end one needs to represent semantic relations with feature descriptions as it was done in Section 9.1.6. With such a representation it is possible to talk about two-place relations in an abstract way. See for instance the discussion of (26) on page 275. However, the unification with the Subject Construction cannot be blocked with reference to semantics since there exist so-called verbs that take a subject without assigning a semantic role to it. As is evidenced by subject verb agreement, *du*

‘you’ is the subject in (27), but the subject does not get a semantic role. The referent of *du* is not the one who *seems*.

- (27) Du scheinst gleich einzuschlafen.
 you seem.2SG soon in.to.sleep
 ‘You seem to fall asleep soon.’

This means that one is forced to either assume an empty expletive subject for verbs like *grauen* or to specify explicitly which verbs may inherit from the subject construction and which may not.

In addition to (27), there exist object raising constructions with accusative objects that can be promoted to subject in passives. The subject in the passive construction does not get a semantic role from the finite verb:

- (28) a. Richard lacht ihn an.
 Richard laughs him towards
 ‘Richard laughs at him.’
 b. Richard fischt den Teich leer.
 Richard fishes the pond empty

The object in (28) is a semantic argument of *an* and *leer*, respectively, but not a semantic argument of the verbs *lacht* and *fischt*, respectively. If one wants to explain these active forms and the corresponding passive forms via the linking constructions in (9), one cannot refer to semantic properties of the verb. Therefore one is forced to postulate specific lexical entries for all possible verb forms in active and passive sentences.

10.6.2 Sign-Based Construction Grammar

In more recent work by Fillmore, Kay, Michaelis and Sag, the Fillmore/Kay formalization of the description of valency using the Kay/Fillmore version of sets was abandoned in favor of the HPSG formalization (Kay 2005; Michaelis 2006; Sag 2012). Sign-Based Construction Grammar was developed from the Berkeley variant of CxG. Sign-Based Construction Grammar uses the formal apparatus of HPSG (typed feature structures). Valency and saturation are treated in exactly the same way as in HPSG. Changes in valency are also analyzed as in HPSG using lexical rules. The analysis of long-distance dependencies was adopted from HPSG (or rather GPSG). Minimal Recursion Semantics (MRS) is used for the description of semantic content. The only difference to works in HPSG is the organization of the features into feature structures. A new feature geometry was introduced to rule out constructions that describe daughters of daughters. Such constructions are in contrast to rules in phrase structure grammars, LFG, and GPSG since they have a much larger locality domain. I do not view this introduction of feature geometry as particularly sensible as it can be easily circumvented and serves to complicate the theory. This will be discussed in what follows.

Sag, Wasow & Bender (2003: 475–489) and Sag (2007) suggest using a *MOTHER*-feature in addition to daughter features. The head-complement construction would then have the form in (29):

(29) Head-complement construction following Sag, Wasow & Bender (2003: 481):

$$\text{head-comp-cx} \rightarrow \left[\begin{array}{l} \text{MOTHER} | \text{SYN} | \text{COMPS } \langle \rangle \\ \text{HEAD-DTR } \boxed{1} \left[\begin{array}{l} \text{SYN} | \text{COMPS } \boxed{2} \\ \text{word} \end{array} \right] \\ \text{DTRS } \langle \boxed{1} \rangle \oplus \boxed{2} \end{array} \right]$$

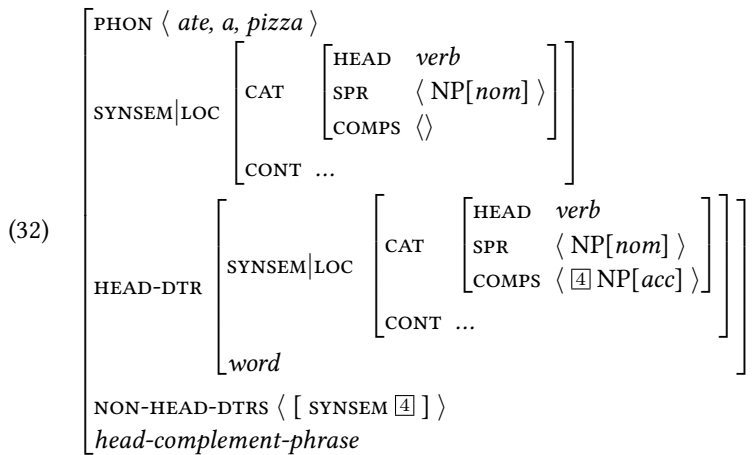
The value of COMPS is then a list of the complements of a head (see Section 9.6). Unlike in HPSG, it is not *synsem* objects that are selected with valence lists, but rather signs. The analysis of the phrase *ate a pizza* takes the form in (30).

$$(30) \left[\begin{array}{l} \text{MOTHER} \left[\begin{array}{l} \text{PHON } \langle \text{ate, a, pizza} \rangle \\ \text{SYN} \left[\begin{array}{l} \text{HEAD } \text{verb} \\ \text{SPR } \langle \text{NP}[\text{nom}] \rangle \\ \text{COMPS } \langle \rangle \end{array} \right] \\ \text{SEM } \dots \\ \text{phrase} \end{array} \right] \\ \text{HEAD-DTR } \boxed{3} \left[\begin{array}{l} \text{SYN} \left[\begin{array}{l} \text{HEAD } \text{verb} \\ \text{SPR } \langle \text{NP}[\text{nom}] \rangle \\ \text{COMPS } \langle \boxed{4} \text{NP}[\text{acc}] \rangle \end{array} \right] \\ \text{SEM } \dots \\ \text{word} \end{array} \right] \\ \text{DTRS } \langle \boxed{3}, \boxed{4} \rangle \\ \text{head-comp-cx} \end{array} \right]$$

The difference between HPSG and Pollard & Sag (1994) is that for Sag, Wasow and Bender, signs do not have daughters and this is why the daughters cannot be selected. As a result, *SYNSEM* feature becomes superfluous (selection of *PHON* is allowed in Sag, Wasow & Bender (2003)). The information about the status of the daughter is only represented in the very outside of the structure. The signed represented under *MOTHER* is of the type *phrase* but does not contain any information about the daughters. The object described in (30) is of course also of another type than the phrasal or lexical signs that can occur as its daughters. We therefore need the following extension so that the grammar will work (Sag et al. 2003: 478):

- (31) Φ is a Well-Formed Structure according to a grammar G if and only if:
1. there is a construction C in G , and
 2. there is a feature structure I that is an instantiation of C , such that Φ is the value of the *MOTHER*-feature of I .

For comparison, a structure is given here with the assumption of the feature geometry that was assumed in Section 9.6:



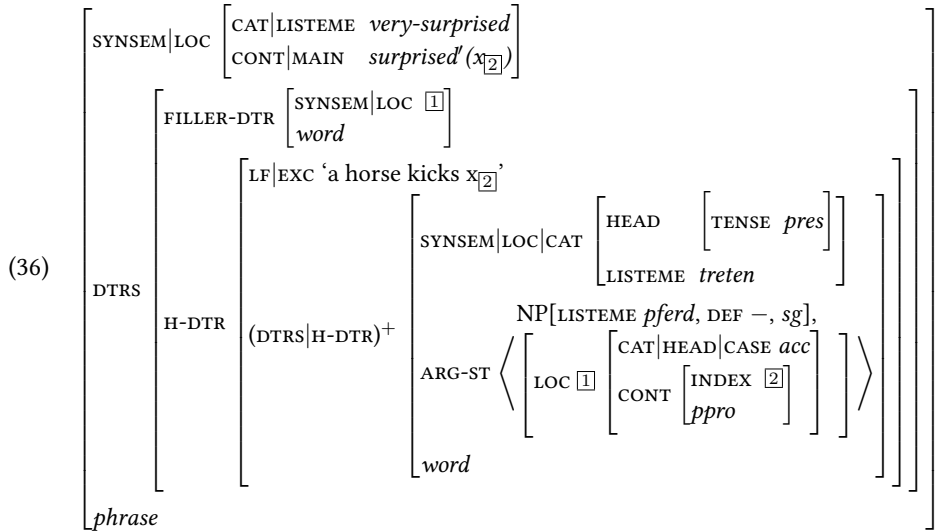
- (34) Ich glaube, mich / #dich tritt ein Pferd.
 I believe me.ACC you.ACC kicks a horse

In order to enforce this co-reference, a restriction has to be able to refer to both the subject of *glauben* ‘believe’ and the object of *treten* ‘kick’ at the same time. In SBCG, there is the possibility to refer to the subject since the relevant information is also available on maximal projections (the value of a special feature (XARG) is identical to the subject of a head). In (33a–c), we are dealing with accusative and dative objects. Instead of only making information about one argument accessible, one could represent the complete argument structure on the maximal projection. This would remove locality of selection, however, for if all heads project their argument structure, then it is possible to determine the properties of arguments of arguments by looking at the elements present in the argument structure. Thus, the argument structure of *wissen* ‘to know’ in (35) would contain the description of a *dass*-clause.

- (35) Peter weiß, dass Klaus kommt.
 Peter knows that Klaus comes
 ‘Peter knows that Klaus is coming.’

Since the description contains the argument structure of *dass*, it is possible to access the argument of *dass*. *wissen* ‘to know’ can therefore access *Klaus kommt*. As such, *wissen* also has access to the argument structure of *kommt* ‘to come’, which is why *Klaus* is also accessible to *wissen*. This should be completely ruled out by a more restrictive feature geometry. Richter and Sailer point out that as well as arguments, adjuncts also have to be accessible in order to be able to capture English idioms such as (33d). In (33d), co-reference has to be established inside an adjunct.

Richter & Sailer (2009: 313) assume a structure for *X-Akk tritt ein Pferd* in (33b) that contains, among others, the constraints in (36). This feature geometry differs somewhat from what was presented in Chapter 9 but that is not of interest here. It is only of importance that the semantic contribution of the entire phrase is *surprised'*($x_{[2]}$). The following is said about the internal structure of the phrase: It consists of a filler-daughter (an extracted element) and also of a head daughter corresponding to a sentence from which something has been extracted. The head daughter means ‘a horse kicks $x_{[2]}$ ’ and has an internal head somewhere whose argument structure list contains a definite NP with the word *Pferd* ‘horse’ as its head. The second element in the argument structure is a pronominal NP in the accusative whose LOCAL value is identical to that of the filler ($[1]$). The entire meaning of this part of the sentence is *surprised'*($x_{[2]}$), whereby $[2]$ is identical to the referential index of the pronoun. In addition to the constraints in (36), there are additional ones that ensure that the partial clause appears with the relevant form of *glauben* or *denken*. The exact details are not that important here. What is important is that one can specify constraints on complex syntactic elements, that is, it must be possible to refer to daughters of daughters. This is possible with classical HPSG feature geometry, but not with the feature geometry in SBCG. For a more general discussion of this, see Chapter 19.



In addition to these empirical problems, there is a conceptual problem with (31): (31) is not part of the formalism of typed feature structures but rather a meta-statement. Therefore, grammars which use (31) cannot be described with the normal formalism. The formalization given in Richter (2004) cannot be directly applied to SBCG, which means that the formal foundations of SBCG still have to be worked out.¹¹ Furthermore, the original problem that (31) was designed to solve is not solved by introducing the new feature geometry and the meta statement. Instead, the problem is just moved to another level since we now need a theory about what is a permissible meta-statement and what is not. As such, a grammarian could add a further clause to the meta statement stating that Φ is only a well-formed structure if it is true of the daughters of a relevant construction C that they are the MOTHER value of a construction C'. It would be possible to formulate constraints in the meta-statement about the construction C' or individual values inside the corresponding feature structures. In this way, locality would have been abandoned since it is possible to refer to daughters of daughters. By assuming (31), the theoretical inventory has been increased without any explanatory gain.

One motivation behind the restrictions on locality was to reduce the computational complexity of the formalism (Ivan Sag, p. c. 2011, See Chapter 18 on computational complexity and generative power). However, the locality restrictions of SBCG can be circumvented easily by structure sharing (Müller 2013b: Section 9.6.1). To see this consider a construction with the following form:

¹¹ A note of caution is necessary since there were misunderstandings in the past regarding the degree of formalizations of SBCG: In comparison to most other theories discussed in this book SBCG is well-formalized. For instance it is easy to come up with a computer implementation of SBCG fragments. I implemented one in the TRALE system myself. The reader is referred to Richter (2004) to get an idea what kind of deeper formalization is talked about here.

$$(37) \left[\begin{array}{c} \left[\begin{array}{cc} \text{PHON} & \text{list of phonemes} \\ \text{SYN} & \text{syntactic information} \\ \text{SEM} & \text{semantic information} \\ \text{NASTY} & \boxed{1} \\ & \text{sign} \end{array} \right] \\ \text{DTRS} & \boxed{1} \text{ list of signs} \end{array} \right]$$

The feature NASTY in the MOTHER sign refers to the value of DTRS and hence all the internal structure of the sign that is licensed by the constructional schema in (37) is available. Of course one could rule out such things by stipulation – if one considered it to be empirically adequate, but then one could as well continue to use the feature geometry of Constructional HPSG and stipulate constraints like “Do not look into the daughters.” An example of such a constraint given in prose is the Locality Principle of Pollard & Sag (1987: 143–144).

Due to the conceptual problems with meta-statements and the relatively simple ways of getting around restrictions, this new organization of features does not bring with it any advantages. Since the grammar becomes more complex (an additional feature, meta-restriction), we should reject this change. However, if we do reject the revised feature geometry, then Sign-Based Construction Grammar and Constructional HPSG are (almost) indistinguishable.

10.6.3 Embodied Construction Grammar

Embodied Construction Grammar was developed by Bergen & Chang (2005) and there are some implementations of fragments of it for German that use this format (Porzel et al. 2006). In the following, I will briefly present the formalism using an example construction. (38) gives the DetNoun construction:¹²

¹² For a similar construction, see Bergen & Chang (2005: 162).

- (38)
- | | |
|-----------------------------|------------|
| Construction DetNoun | |
| subcase of RefExp | |
| constructional | |
| d:Determiner | |
| c:CommonNoun | |
| self.case | ↔ d.case |
| self.number | ↔ c.number |
| d.gender | ↔ c.gender |
| d.case | ↔ c.case |
| d.number | ↔ c.number |
| form | |
| d.f before c.f | |
| meaning | |
| self.m | ↔ c.m |

This representational form is reminiscent of PATR-II grammars (Shieber, Uszkoreit, Pereira, Robinson & Tyson 1983): As in PATR-II, the daughters of a construction are given names. As such, (38) contains the daughters c and d. d is a determiner and c is a count noun. It is possible to refer to the construction itself via the object **self**. Constructions (and also their daughters) are feature-value descriptions. Structure sharing is represented by path equations. For example, $d.\text{gender} \leftrightarrow c.\text{gender}$ states that the value of the gender feature of the determiner is identical to the gender feature of the noun. As well as restrictions for features, there are restrictions on the form. $d.f \text{ before } c.f$ states that the form contribution of the determiner must occur before that of the noun. Bergen & Chang (2005) differentiate between immediate (**meets**) and non-immediate precedence (**before**). Part of the information represented under f is the orthographic form (f.orth). The inheritance relation is given explicitly in the construction as in Kay & Fillmore (1999).

The construction in (38) can be represented in a similar way to the format used in Chapter 6:

- (39)
- | | | | | | |
|--|-----------|---|---|--|---|
| <div style="border-left: 1px solid black; padding-left: 10px;"> F ORTH [1] ⊕ [2]
 CASE [3]
 NUMBER [4]
 M [5] </div> | DTRS < | <div style="border-left: 1px solid black; padding-left: 10px;"> F ORTH [1]
 CASE [3]
 NUMBER [4]
 GENDER [6]
 <i>Determiner</i> </div> | , | <div style="border-left: 1px solid black; padding-left: 10px;"> F ORTH [2]
 CASE [3]
 NUMBER [4]
 GENDER [6]
 M [5]
 <i>CommonNoun</i> </div> | > |
| <i>DetNoun</i> | | | | | |

The structure in (39) corresponds to a construction where the determiner directly precedes the noun because the form contribution of the determiner has been combined with that of the noun. This strict adjacency constraint makes sense as the claim that the determiner must precede the noun is not restrictive enough since sequences such as (40b) would be allowed:

- (40) a. [dass] die Frauen Türen öffnen
 that the women doors open
 ‘that the woman open doors’
 b. * die Türen öffnen
 Frauen

die Türen ‘the doors’ could be analyzed with the DetNoun Construction although another noun phrase intervenes between the determiner and the noun (Müller: 1999a: 424; 1999c). The order in (40b) can be ruled out by linearization constraints or constraints on the continuity of arguments. If we want the construction to demand that the determiner and noun be adjacent, then we would simply use **meets** instead of **before** in the specification of the construction.

This discussion has shown that (39) is more restrictive than (38). There are, however, contexts in which one could imagine using discontinuous constituents such as the deviant ones in (40). For example, discontinuous constituents have been proposed for verbal complexes, particle verbs and certain coordination data (Wells 1947). Examples for analyses with continuous constituents in the framework of HPSG are Reape (1994); Kathol (1995; 2000); Crysmann (2008); Beavers & Sag (2004).¹³ These analyses differ from those previously presented in that they use an additional feature **DOMAIN**, whose value is a list containing the head and the elements dependent on it, instead of daughter features. The elements do not have to necessarily be adjacent in the utterance, that is, discontinuous constituents are permitted. Which elements are entered into this list in which way is governed by the constraints that are part of the linguistic theory. This differs from the simple **before** statement in ECG in that it is much more flexible and in that one can also restrict the area in which a given element can be ordered since elements can be freely ordered inside a domain.

There is a further difference between the representation in (38) and the general HPSG schemata: In the ECG variant, linearization requirements are linked to constructions. In HPSG and GPSG, it is assumed that linearization rules hold generally, that is, if we were to assume the rules in (41), we would not have to state for each rule explicitly that shorter NPs tend to precede longer ones and that animate nouns tend to occur before inanimate ones.

- (41) a. $S \rightarrow NP[nom], NP[acc], V$
 b. $S \rightarrow NP[nom], NP[dat], V$
 c. $S \rightarrow NP[nom], NP[dat], NP[acc], V$

¹³ Crysmann, Beaver and Sag deal with coordination phenomena. For an analysis of coordination in TAG that also makes use of discontinuous constituents, see Sarkar & Joshi (1996) and Section 23.6.2.

d. $S \rightarrow \text{NP}[\text{nom}], \text{NP}[\text{acc}], \text{PP}, \text{V}$

It is possible to capture these generalizations in ECG if one specifies linearization constraints for more general constructions and more specific constructions inherit them from these. As an example, consider the Active-Ditransitive Construction discussed by Bergen & Chang (2005: 170):

- (42) **Construction** Active-Ditransitive
 subcase of Pred-Expr
 constructional
 agent:Ref-Expr
 action:Verb
 recipient:Ref-Expr
 theme:Ref-Expr
 ...
 form
 agent.f **before** action.f
 action.f **meets** recipient.f
 recipient.f **meets** theme.f
 meaning
 ...

These restrictions allow the sentences in (43a,b) and rule out those in (43c):

- (43) a. Mary tossed me a drink.
 b. Mary happily tossed me a drink.
 c. * Mary tossed happily me a drink.

The restriction agent.f **before** action.f forces an order where the subject occurs before the verb but also allows for adverbs to occur between the subject and the verb. The other constraints on form determine the order of the verb and its object: the recipient must be adjacent to the verb and the theme must be adjacent to the recipient. The requirement that an agent in the active occur before the verb is not specific to ditransitive constructions. This restriction could therefore be factored out as follows:

- (44) **Construction** Active-Agent-Verb
 subcase of Pred-Expr
 constructional
 agent:Ref-Expr
 action:Verb
 form
 agent.f **before** action.f

The Active-Ditransitive Construction in (42) would then inherit the relevant information from (44).

In addition to the descriptive means used in (38), there is the evokes operator (Bergen & Chang 2005: 151–152). An interesting example is the representation of the term *hypotenuse*: This concept can only be explained by making reference to a right-angled triangle (Langacker 1987: Chapter 5). Jerome Feldman (slides) gives the following formalization:

- (45)

Schema hypotenuse subcase of line-segment evokes right-triangle as rt constraints self \leftrightarrow rt.long-side

This states that a hypotenuse has a particular length, namely the longest side of a right-angled triangle. The concept of a right-angled triangle is activated by means of the evokes operator. Evokes creates an instance of an object of a certain type (in the example, *rt* of type *right-triangle*). It is then possible to refer to the properties of this object in a schema/construction.

The feature description in (46) is the equivalent example given in the notation from Chapter 6, whereby *hypotenuse* is a sub-type of *line-segment*:

- (46) $\boxed{1} \left[\begin{array}{l} \text{EVOKES} \left\langle \left[\begin{array}{l} \text{LONG-SIDE } \boxed{1} \\ \text{right-triangle} \end{array} \right] \right\rangle \\ \text{hypotenuse} \end{array} \right]$

The value of EVOKES is a list since a schema or construction can evoke more than one concept. The only element in this list in (46) is an object of type *right-triangle*. The value of the feature LONG-SIDE is identified with the entire structure. This essentially means the following: I, as a hypotenuse, am the long side of a right-angled triangle.

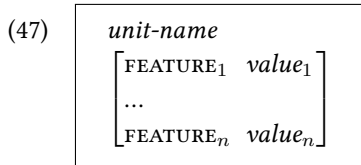
10.6.4 Fluid Construction Grammar

Van Trijp (2013; 2014) claims that SBCG and HPSG are fundamentally different from Fluid Construction Grammar (FCG). He claims that the former approaches are generative ones while the latter is a cognitive-functional one. I think that it is not legitimate to draw these distinctions on the basis of what is done in FCG.¹⁴ I will comment on this at various places in this section. I first deal with the representations that are used in FCG, talk about argument structure constructions, the combination operations fusion and merging that are used in FCG and then compare FCG in detail with SBCG/HPSG.

¹⁴ Steels (2013: 153) emphasizes the point that FCG is a technical tool for implementing constructionist ideas rather than a theoretical framework of its own. However, authors working with the FCG system publish linguistic papers that share a certain formal background and certain linguistic assumptions. So this section addresses some of the key assumptions made and some of the mechanisms used.

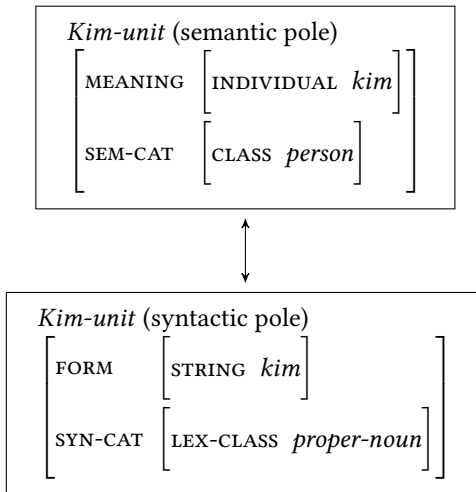
10.6.4.1 General remarks on the representational format

Fluid Construction Grammar (FCG, Steels 2011) is similar to HPSG in that it uses attribute value matrices to represent linguistic objects. However, these AVMs are untyped as in LFG. Since there are no types, there are no inheritance hierarchies, but one can use macros to reach similar effects. Constructions can refer to more general constructions (van Trijp 2013: 105). Every AVM comes with a name and can be depicted as follows:



Linguistic objects have a form and a meaning pole. The two poles could be organized in one feature description by using a SYN and a SEM feature, but in FCG papers the two poles are presented separately and connected via a double arrow. (48) is an example:

(48) *Kim* according to van Trijp (2013: 99):



Depending on the mode in which the lexical items are used, the syntactic pole or the semantic pole is used first. The first processing step is a matching phase in which it is checked whether the semantic pole (for generation) or the syntactic pole (for parsing) matches the structure that was build so far. After this test for unification, the actual unification, which is called merging, is carried out. After this step, the respective other pole (syntax for generation and semantics for parsing) is merged. This is illustrated in Figure 10.2 on the facing page.

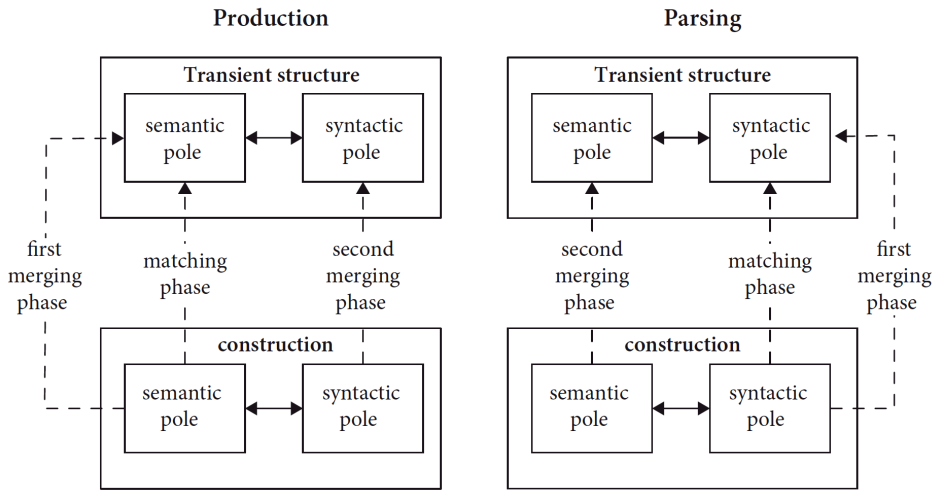


Figure 10.2: Generation and parsing in FCG (van Trijp 2011: 99)

10.6.4.2 Argument Structure Constructions

Fluid Construction Grammar assumes a phrasal approach to argument structure, that is, it is assumed that lexical items enter into phrasal configurations that contribute independent meaning van Trijp (2011). The FCG approach is one version of implementing Goldberg's plugging approach to argument structure constructions (Goldberg 1995). Van Trijp suggests that every lexical item comes with a representation of potential argument roles like Agent, Patient, Recipient, and Goal. Phrasal argument structure constructions are combined with the respective lexical items and realize a subset of the argument roles, that is they assign them to grammatical functions. Figure 10.3 on the next page shows an example: The verb *sent* has the semantic roles Agent, Patient, Recipient, and Goal. Depending on the argument structure construction that is chosen, a subset of these roles is selected for realization.¹⁵

Note that in such an approach it is necessary to have a passive variant of every active construction. For languages that allow for the combination of passive and impersonal constructions, one would be forced to assume a transitive-passive-impersonal construction. As was argued in Müller (2006: Section 2.6) free datives (*commodi/incommodi*) in

¹⁵ It is interesting to note here that van Trijp (2011: 141) actually suggests a lexical account since every lexical item is connected to various phrasal constructions via coapplication links. So every such pair of a lexical item and a phrasal construction corresponds to a lexical item in Lexicalized Tree Adjoining Grammar (LTAG). See also Müller & Wechsler (2014a: 25) on Goldberg's assumption that every lexical item is associated with phrasal constructions.

Note that such coapplication links are needed since without them the approach cannot account for cases in which two or more argument roles can only be realized together but not in isolation or in any other combination with other listed roles.

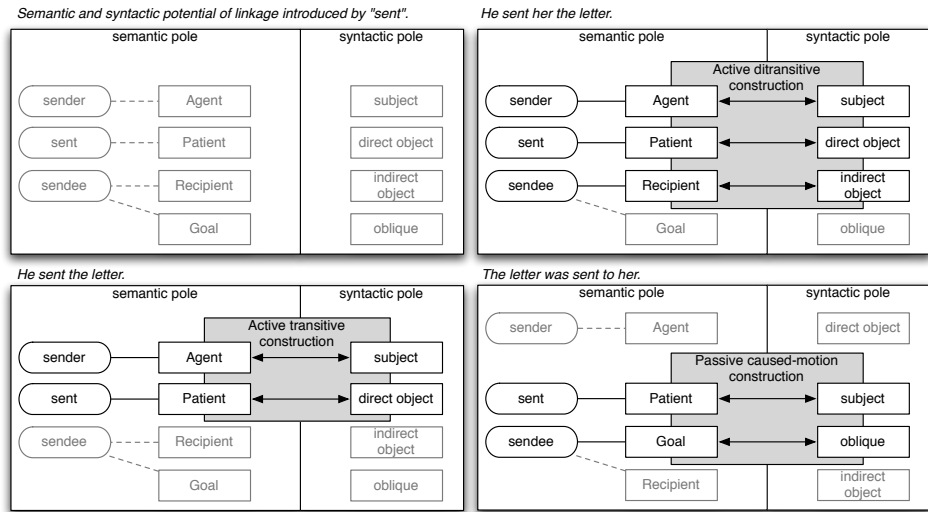


Figure 10.3: Lexical items and phrasal constructions. Figure taken from van Trijp (2011: 122)

German can be added to almost any construction. They interact with the dative passive and hence should be treated as arguments. So, for the resultative construction one would need an active variant, a passive variant, a variant with dative argument, a variant with dative argument and dative passive, and a middle variant. While it is technically possible to list all these patterns and it is imaginable that we store all this information in our brains, the question is whether such listings really reflect our linguistic knowledge. If a new construction comes into existence, let's say an active sentence pattern with a nominative and two datives in German, wouldn't we expect that this pattern can be used in the passive? While proposals that establish relations between active and passive constructions would predict this, alternative proposals that just list the attested possibilities do not.

The issue of how such generalizations should be captured was discussed in connection with the organization of the lexicon in HPSG (Flickinger 1987; Meurers 2001). In the lexical world, one could simply categorize all verbs according to their valence and say that *loves* is a transitive verb and the passive variant *loved* is an intransitive verb. Similarly *gives* would be categorized as a ditransitive verb and *given* as a transitive one. Obviously this misses the point that *loved* and *given* share something: they both are related to their active form in a systematic way. This kind of generalization was called a horizontal generalization as compared to vertical generalizations, which describe classes in an inheritance hierarchy.

The issue is independent of the lexical organization of knowledge, it can be applied to phrasal representations as well. Phrasal constructions can be organized in hierarchies

(vertical), but the relation between certain variants is not covered by this. The analog to the lexical rules in a lexical approach are GPSG-like metarules in a phrasal approach.

10.6.4.3 Fusion, matching and merging

As was pointed out by Dowty (1989: 89–90), it is not sufficient to check for semantic compatibility for the decision whether a verb may enter (or be fused with) a certain construction. The example is the contrast between *dine*, *eat*, and *devour*. While the thing that is eaten may not be realized with *dine*, its realization is optional with *eat* and obligatory with *devour*. So the lexical items have to come with some information about this.

Van Trijp (2011) and Steels & van Trijp (2011) make an interesting suggestion that could help here: Every verb comes with a list of potential roles and argument structure constructions can pick subsets of these roles (see Figure 10.3). This is called *matching*: it is not allowed to introduce new argument roles. This would make it possible to account for *dine*: One could say that there is something that is eaten, but that no Theme role is made available for linking to grammatical functions. To account for the extension of argument roles as it is observed in the caused motion construction Steels & van Trijp (2011) suggest a process called *merging*. Merging is seen as a repair strategy: if an utterance involves an intransitive verb and some other material, the utterance cannot be processed with matching alone. *he sneezed* could be parsed, but *the foam* and *off the cappuccino* would be unintegrated.

(49) He sneezed the foam off the cappuccino.

So, Steels & van Trijp (2011: 319–320) suggest that only if regular constructions cannot apply, merging is allowed. There are two problems with this: First, human language is highly ambiguous and in the case at hand this could result in situations in which there is a reading for an utterance, so that the repair strategy would never kick in. Consider (50):

(50) Er schlug den Mann tot.
 he beat the man dead
 ‘He beat the man to death’ or ‘He beat the dead man.’

(50) has two readings: The resultative reading in which *tot* expresses the result of the beating and another reading in which *tot* is a depictive predicate. So, the problem is that (50) has a reading and therefore repair is not triggered: *schlug* is used with the transitive construction and *tot* is an adjunct (see Winkler (1997)). However, the more likely analysis of (50) is the one with the resultative analysis, in which the valency frame is extended by an oblique element. So this means that one has to allow the application of merging independent of other analyses that might be possible. As Steels and van Trijp note (p. 320), if merging is allowed to apply freely, utterances like (51a) will be allowed and of course (51b) as well.

(51) a. * She sneezed her boyfriend.
 b. * She dined a steak.

In (51a) *sneeze* is used in the ditransitive construction and in (51b) *dined* is used in the transitive construction.

The way out of this dilemma is to establish information in lexical items that specifies in which syntactic environments a verb can be used. This information can be weighted and for instance the probability of *dine* to be used transitively would be extremely low. Steels and van Trijp would connect their lexical items to phrasal constructions via so-called coapplication links and the strength of this link would be very low for *dine* and the transitive construction and reasonably high for *sneeze* and the caused-motion construction. This would explain the phenomena (and in a usage-based way), but it would be a lexical approach, as it is common in CG, HPSG, SBCG.

10.6.4.4 Nonlocal dependencies

Van Trijp (2014) compares the SLASH-based approaches that are used in GPSG, HPSG, and SBCG with the approach that he suggests within the framework of FCG. He claims that there are fundamental differences between SBCG and FCG and assigns SBCG to the class of generative grammars, while placing FCG in the class of cognitive-functional approaches. He claims that his cognitive-functional approach is superior in terms of completeness, explanatory adequacy, and theoretical parsimony (p. 2). What van Trijp (2014) suggests is basically an analysis that was suggested by Reape in the middle of the 90ies in unpublished work. Van Trijp develops a model of grammar that allows for discontinuous constituents and just treats the serialization of the object in sentences like (52) as an alternative linearization option.

- (52) a. This book, I read.
b. What did the boy hit?

Van Trijp's analysis involves several units that do not normally exist in phrase structure grammars, but can be modeled via adjacency constraints or represent relations between items which are part of lexical representations in HPSG/SBCG anyway. An example is the subject-verb anchor that connects the subject and the verb to represent the fact that these two items play an important functional role. Figure 10.4 on the facing page shows the analysis of (53).

- (53) What did the boy hit?

As can be seen in the figure, van Trijp also refers to information structural terms like topic and focus. It should be noted here that the analysis of information structure has quite some history in the framework of HPSG (Engdahl & Vallduví 1996; Kuhn 1995; 1996; Günther et al. 1999; Wilcock 2001; De Kuthy 2002; Paggio 2005; Bildhauer 2008; Bildhauer & Cook 2010). The fact that information structure is not talked about in syntax papers like Sag (2012) does not entail that information structure is ignored or should be ignored in theories like HPSG and SBCG. So much for completeness. The same holds of course for explanatory adequacy. This leaves us with theoretical parsimony, but before I comment on this, I want to discuss van Trijp's analysis in a little bit more detail, to show

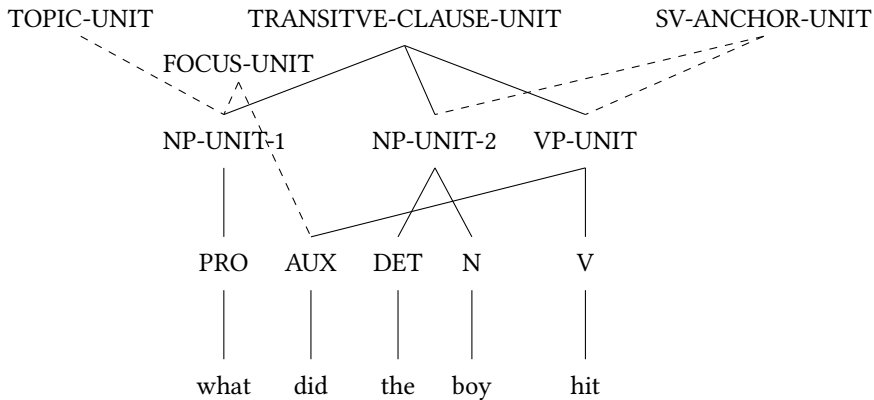


Figure 10.4: The analysis of *What did the boy hit?* according to van Trijp (2014)

that many of his claims are empirically problematic and that his theory therefore cannot be explanatory since empirical correctness is a precondition for explanatory adequacy.

Van Trijp claims that sentences with nonlocal dependency constructions in English start with a topic.¹⁶ Bresnan's sentences in (2) and (3) were discussed on page 220 (Bresnan 2001: 97) and are repeated below for convenience:

- (54) Q: What did you name your cat?
 A: Rosie I named her. (*Rosie* = FOCUS)
- (55) Q: What did you name your pets?
 A: My dog, I named Harold. My cat, I named Rosie. (*my dog, my cat* = TOPIC)

These sentences show that the pre-subject position is not unambiguously a topic or a focus position. So, if this position is to be associated with an information structural function, this association has to be a disjunction.

A further problematic aspect of van Trijp's analysis is that he assumes that the auxiliary *do* is an object marker (p. 10, 22) or a non-subject marker (p. 23). It is true that *do* support is not necessary in subject questions like (56a), but only in (56b), but this does not imply that all items that are followed by *do* are objects.

- (56) a. Who saw the man?
 b. Who did John see?

First, *do* can be used to emphasize the verb:

- (57) Who *did* see the man?

¹⁶ He uses the following definitions for topic and focus: *Topicality is defined in terms of aboutness: the topic of an utterance is what the utterance is 'about'. Focality is defined in terms of salience: focus is used for highlighting the most important information given the current communicative setting.* (van Trijp 2014: 15)

Second all types of other grammatical functions can precede the verb:

- (58) a. Where did you see the man? (adverbial)
 b. How tall is the man? (predicative)
 c. What did John consider Peter? (predicative)
 d. What does this book cost? (adverbial)
 e. About what did you talk? (prepositional object)

And finally, even a subject can appear in front of *do* if it is extracted from another clause:

- (59) Who does he think saw this man? (subject)

There is a further empirical problem: Approaches that assume that a filler is related to its origin can explain scope ambiguities that only arise when an element is extracted. Compare for instance the sentence in (60a) with the sentences in (60b, c): Although the order of *oft* and *nicht* in (60a) and (60c) is the same, (60a) is ambiguous but (60c) is not.

- (60) a. Oft liest er das Buch nicht.
 often reads he the book not
 'It is often that he does not read the book.' or 'It is not the case that he reads the book often.'
 b. dass er das Buch nicht oft liest
 that he the book not often reads
 'It is not the case that he reads the book often.'
 c. dass er das Buch oft nicht liest
 that he the book often not reads
 'It is often that he does not read the book.'

(60a) has the two readings that correspond to (60b) and (60c). A purely linearization-based approach probably has difficulties to explain this. A SLASH-based approach can assume that (60a) has a gap (or some similar means for the introduction of nonlocal dependencies) at the position of *oft* in (60b) or (60c). The gap information is taken into account in the semantic composition at the site of the gap. This automatically accounts for the observed readings.

Another empirical problem that has to be solved is the existence of extraction path marking languages. Bouma, Malouf & Sag (2001a) list a number of languages in which elements vary depending on the existence or absence of a gap in a constituent they attach to. For instance Irish has complementizers that have one form if the clause they attach to has an element extracted and another form if it does not. Van Trijp's answer to that challenge is that all languages are different (p. 22) and that the evidence from one language does not necessarily mean that the analysis for that language is also appropriate for another language. While I agree with this view in principle (see Section 13.1), I do think that extraction is a rather fundamental property of languages and that nonlocal dependencies should be analyzed in parallel for those languages that have it.

Van Trijp points out that SBCG does not have a performance model and contrasts this with FCG. On page 11 he states:

So parsing starts by segmenting the utterance into discrete forms, which are then categorized into words by morphological and lexical constructions, and which can then be grouped together as phrases (see Steels, 2011b, for a detailed account of lexico-phrasal processing in FCG). So the parser will find similar constituents for all four utterances, as shown in examples (21–24). Since auxiliary-*do* in example (24) falls outside the immediate domain of the VP, it is not yet recognized as a member of the VP.

All of these phrases are disconnected, which means that the grammar still has to identify the relations between the phrases. (van Trijp 2014: 7)

Van Trijp provides several tree fragments that contain NPs for subject and object and states that these have to be combined in order to analyze the sentences he discusses. This is empirically inadequate: If FCG does not make the competence/performance distinction, then the way utterances are analyzed should reflect the way humans process language (and this is what is usually claimed about FCG). However, all we know about human language processing points towards an incremental processing, that is, we process information as soon as it is available. We start to process the first word taking into account all of the relevant aspects (phonology, stress, part of speech, semantics, information structure) and come up with an hypothesis about who the utterance could proceed. As soon as we have two words processed (in fact even earlier: integration happens already during the processing words) we integrate the second word into what we know already and continue to follow our hypothesis, or revise it, or simply fail. See Section 15.2 for details on processing and the discussion of experiments that show that processing is incremental. So, we have to say, that van Trijp's analysis fails on empirical grounds: His modeling of performance aspects is not adequate.

The parsing scheme that van Trijp describes is pretty much similar to those of HPSG parsers, but these usually come without any claims about performance. Modeling performance is rather complex since a lot of factors play a role. It is therefore reasonable to separate competence and performance and continue to work the way it is done in HPSG and FCG. This does not mean that performance aspects should not be modeled, in fact psycholinguistic models using HPSG have been developed in the past (Konieczny 1996), but developing both a grammar with large coverage and the performance model that combines with it demands a lot of resources.

I now turn to parsimony: Van Trijp uses a subject-verb anchor construction that combines the subject and the main verb. Because of examples like (61) it must be possible to have discontinuous subject-verb constructions:¹⁷

¹⁷ Unless modals and tense auxiliaries are treated as main verbs (which they should not in English), constructions with modals seem to be another case where the subject and the main verb are not adjacent:

- (i) a. Peter will read the book.
- b. Peter has read the book.

- (61) Peter often reads books.

But if such constructions can be discontinuous one has to make sure that (62b) cannot be an instantiation of the subject-verb construction:

- (62) a. The boy I think left.
b. *I the boy think left.

Here it is required to have some adjacency between the subject and the verb it belongs to modulo some intervening adverbials. This is modeled quite nicely in phrase structure grammars that have a VP node. Whatever the internal structure of such a VP node may be, it has to be adjacent to the subject in sentences like the ones above. The dislocated element has to be adjacent to the complex consisting of subject and VP. This is what the Filler-Head Schema does in HPSG and SBCG. Van Trijp criticizes SBCG for having to stipulate such a schema, but I cannot see how his grammar can be complete without a statement that ensures the right order of elements in sentences with fronted elements.

Van Trijp stated that FCG differs from what he calls generative approaches in that it does not want to characterize only the well-formed utterances of a language. According to him the parsing direction is much more liberal in accepting input than other theories. So it could well be that he is happy to find a structure for (62b). Note though that this is incompatible with other claims of van Trijp: He argued that FCG is superior to other theories in that it comes with a performance model (or rather in not separating competence from performance at all). But then (62a) should be rejected both on competence and performance grounds. It is just unacceptable and speakers reject it for whatever reasons. Any sufficiently worked out theory of language has to account for this.

One of the success stories of non-transformational grammar is the SLASH-based analysis of nonlocal dependencies by Gazdar (1981b). This analysis made it possible for the first time to explain Ross's Across the Board Extraction (Ross 1967). The examples were already discussed on page 196 and are repeated here for convenience:

- (63) a. The kennel which Mary made and Fido sleeps in has been stolen.
(= S/NP & S/NP)
b. The kennel in which Mary keeps drugs and Fido sleeps has been
stolen. (= S/PP & S/PP)
c. * The kennel (in) which Mary made and Fido sleeps has been
stolen. (= S/NP & S/PP)

The generalization is that two (or more) constituents can be coordinated if they have identical syntactic categories and identical SLASH values. This explains why *which* and *in which* in (63a,b) can fill two positions in the respective clauses. Now, theories that do not use a SLASH feature for the percolation of information about missing elements have to find different ways to make sure that all argument slots are filled and that the correct correspondence between extracted elements and the respective argument role is established. Note that this is not straightforward in models like the one that is suggested by van Trijp, for he has to allow that the preposition *in* is combined with some material

to the left of it that is simultaneously also the object of *made*. Usually an NP cannot simply be used by two different heads as their argument. As an example consider (64a):

- (64) a. * John said about the cheese that I like.
 b. John said about the cheese that I like it.

If it would be possible to use material several times, a structure for (64a) would be possible in which *the cheese* is the object of the preposition *about* and of the verb *like*. This sentence, however, is totally out: The pronoun *it* has to be used to fill the object slot.

There is a further problem related to discontinuity. If one does not restrict continuity, then constituent orders like (65b) are admitted by the grammar:

- (65) a. Deshalb klärt, dass Peter kommt, ob Klaus spielt.
 therefore resolves that Peter comes whether Klaus plays
 ‘Therefore that Peter comes resolves whether Klaus plays.’
 b. Deshalb klärt dass ob Peter Klaus kommt spielt.
 therefore resolves that whether Peter Klaus comes plays

The interesting thing about the word salad in (65b) is that the constituent order within the *dass* clause and within the *ob* clause is correct. That is the complementizer precedes the subject, which in turn precedes the verb. The problem is that the constituents of the two clauses are mixed.

In a model that permits discontinuous constituents one cannot require all parts of an argument have to be arranged after all parts that belong to another argument, since the discontinuity is used to account for nonlocal dependencies. So it must be possible to have *Klaus* before other arguments (or parts of other arguments), since *Klaus* can be extracted. An example of mixing parts of phrases is given in (66):

- (66) Dieses Buch hat der Mann mir versprochen, seiner Frau zu geben, der
 this book has the man me promised his wife to give who
 gestern hier aufgetreten ist.
 yesterday here performed is
 ‘The man who performed here yesterday promised me to give this book to his wife.’

We see that material that refers to *der Mann* namely the relative clause *der gestern hier aufgetreten ist* appears to the right. And the object of *geben*, which would normally be part of the phrase *dieses Buch seiner Frau zu geben* appears to the left. So, in general it is possible to mix parts of phrases, but this is possible in a very restricted way only. Some dependencies go to the far left (fronting) and others to the extreme right (extraposition). Extraposition is clause bound, while extraction is not. In approaches like GPSG, HPSG, and SBCG the facts are covered by assuming that constituents for a complete clause are continuous apart from constituents that are fronted or extraposed. The fronted and extraposed constituents are represented in SLASH and EXTRA, respectively, rather than in valence features, so that it is possible to require of constituents that have all their valents saturated to be continuous.

Summing up the discussion of parsimony, it has to be said that van Trijp has to provide the details on how continuity is ensured. The formalization of this is not trivial and only after this is done can the SLASH-based approach be compared with FCG.

In addition to all the points discussed so far there is a logical flaw in van Trijp's argumentation: He states that:

whereas the filler-gap analysis cannot explain why *do*-support does not occur in *wh*-questions where the subject is assigned questioning focus, this follows naturally from the interaction of different linguistic perspectives in this paper's approach. (van Trijp 2014: 22)

The issue here is whether a filler-gap analysis or an analysis with discontinuous constituents is suited better for explaining the data. A correct argumentation against the filler gap analysis would require a proof that information structural or other functional constraints cannot be combined with this analysis. This proof was not provided and in fact I think it cannot be provided since there are approaches that integrate information structure. Simply pointing out that a theory is incomplete does not falsify a theory. This point was already made in my review of Boas (2003) and in a reply to Boas (2014). See Müller (2005a: 655–656), Müller (2007b: Chapter 20), and Müller & Wechsler (2014b: Footnote 15).

The conclusion about the FCG analysis of nonlocal dependencies is that there are some empirical flaws that can be easily fixed or simply be dropped (role of *do* as object marker, claim that the initial position in English fronting construction is the topic), some empirical shortcomings (coordination, admittance of illformed utterances with discontinuous constituents), some empirical problems when the analysis is extended to other languages (scope of adjuncts in German), and the parsimony of the analyses is not really comparable since the restrictions on continuity are not really worked out. If the formalization of restrictions on continuity in FCG are half way near the machinery that is used in HPSG formalizations, the SLASH-based analysis wins.

In any case, I do not see how nonlocal dependencies could be used to drive a wedge between SBCG and FCG. If there are functional considerations that have to be taken into account, they should be modeled in both frameworks. In general FCG should be more restrictive than SBCG since FCG claims to integrate a performance model, so both competence and performance constraints should be operative. I will come back to the competence/performance distinction in the following section, which is a more general comparison of SBCG and FCG.

10.6.4.5 Comparison to Sign-Based Construction Grammar/HPSG

According to van Trijp (2013) there are the differences shown in Table 10.1 on the next page. These tables will be discussed in the following subsections.

Table 10.1: Differences between SBCG and FCG according to van Trijp (2013: 112)

Scientific model	Theoretical physics (abstract calculus)	Evolutionary theory (complex adaptive system)
Linguistic approach	Generative (competence model)	Cognitive-functional (parsing and production)
Formalization	Mathematical (amenable for implementation)	Computational (implemented)
Constructions	Static type constraints	Dynamic mappings
Constructicon	Signature and grammar	Open-ended inventory
Processing	Assumption of processing- independence	Bidirectional processing model

10.6.4.5.1 Competence/performance distinction

As for the linguistic approach, the use of the term *generative* is confusing. What van Trijp means – and also explains in the paper – is the idea that one should separate competence and performance. We will deal with both the generative-enumerative vs. constraint-based view and with the competence/performance distinction in more detail in the Chapters 14 and 15, respectively. Concerning the cognitive-functional approach van Trijp writes:

The goal of a cognitive-functional grammar, on the other hand, is to explain how speakers express their conceptualizations of the world through language (= *production*) and how listeners analyze utterances into meanings (= *parsing*). Cognitive-functional grammars therefore implement both a competence and a processing model. (van Trijp 2013: 90)

It is true that HPSG and SBCG make a competence/performance distinction (Sag & Wasow 2011). HPSG theories are theories about the structure of utterances that are motivated by distributional evidence. These theories do not contain any hypothesis regarding brain activation, planning of utterance, processing of utterances (garden path effects) and similar things. In fact none of the theories that are discussed in this book contains an explicit theory that explains all these things. I think that it is perfectly legitimate to work in this way: It is legitimate to study the structure of words without studying their semantics and pragmatics, it is legitimate to study phonology without caring about syntax, it is legitimate to deal with specific semantic problems without caring about phonology and so on, provided there are ways to integrate the results of such research into a bigger picture. So, it is wrong to develop models like those developed in current versions of Minimalism (called Biolinguistics), where it is assumed that utterances are derived in phases (NPs, CPs, depending on the variant of the theory) and then shipped to the interfaces (spell out and semantic interpretation). This is not what humans do (see Chapter 15). But if we are neutral with respect towards such issues, we are fine. In fact there

is psycholinguistic work that couples HPSG grammars to performance models and similar work exists for TAG. Similarly, Adele Goldberg's book from 1995 does not contain a worked out theory of performance facts. It contains boxes in which grammatical functions that are related to semantic roles. So this basically is a competence theory as well. Of course there are statements about how this is connected to psycholinguistic findings, but this is also true for theories like HPSG and Simpler Syntax that explicitly make the competence/performance distinction.

10.6.4.5.2 Mathematical formalization vs. implementation

The difference between mathematical and computational is a rather strange distinction to make. I think that a formal and precise description is a prerequisite for implementation (see the discussion in Chapter 3.6.2). Apart from this a computer implementation of SBCG is trivial, given the systems that we have for processing HPSG grammars. In order to show this, I want to address one issue that van Trijp discusses. He claims that SBCG cannot be directly implemented. On issues of complexity of constraint solving systems he quotes (Levine & Meurers 2006: Section 4.2.2):

Actual implementations of HPSG typically handle the problem by guiding the linguistic processor using a (rule-based) phrase structure backbone, but the disadvantage of this approach is that the “organization and formulation of the grammar is different from that of the linguistic theory” (Levine & Meurers 2006: Section 4.2.2). (van Trijp 2013)

He concludes:

Applying all these observations to the operationalization of SBCG, we can conclude that an SBCG grammar is certainly amenable for computational implementation because of its formal explicitness. There are at least two computational platforms available, mostly used for implementing HPSG-based grammars, whose basic tenets are compatible with the foundations of SBCG: LKB (Copestake 2002) and TRALE (Richter 2006). However, none of these platforms support a ‘direct’ implementation of an SBCG grammar as a general constraint system, so SBCG’s performance-independence hypothesis remains conjecture until proven otherwise.

There are two issues that should be kept apart here: efficiency and faithfulness to the theory. First, as Levine and Meurers point out there were many constraint solving systems at the beginning of the 90ties. So there are computer systems that can be and have been used to implement and process HPSG grammars. This is very valuable since they can be used for direct verification of specific theoretical proposals. As was discussed by Levine and Meurers, trying to solve constraints without any guidance is not the most efficient way to deal with the parsing/generation problem. Therefore, additional control-structure was added. This control structure is used for instance in a parser to determine the syntactic structure of a phrase and other constraints will apply as soon as there is sufficient information available for them to apply. For instance, the assignment of structural case happens once the arguments of a head are realized. Now, is it bad to have a

phrase structure backbone? One can write down phrase structure grammars that use phrase structure rules that have nothing to do with what HPSG grammars usually do. The systems TRALE (Meurers, Penn & Richter 2002; Penn 2004) and LKB will process them. But one is not forced to do this. For instance the grammars that I developed for the CoreGram project (Müller 2013a; 2015a) are very close to the linguistic theory. To see that this is really the case, let us look at the Head-Argument Schema. The Head-Argument Schema is basically the type *head-argument-phrase* with certain type constraints that are partly inherited from its supertypes. The type with all the constraints was given on page 272 and is repeated here as (67):

$$(67) \text{ (syntactic) constraints on } head\text{-argument-phrase:}$$

$$\left[\begin{array}{l} \text{SYNSEM|LOC|CAT} \left[\begin{array}{l} \text{HEAD} \quad [1] \\ \text{SUBCAT} \quad [2] \end{array} \right] \\ \\ \text{HEAD-DTR|SYNSEM|LOC|CAT} \left[\begin{array}{l} \text{HEAD} \quad [1] \\ \text{SUBCAT} \quad [2] \oplus \langle [3] \rangle \end{array} \right] \\ \\ \text{NON-HEAD-DTRS} \langle [\text{SYNSEM} [3]] \rangle \\ head\text{-argument-phrase} \end{array} \right]$$

This can be translated into a phrase structure grammar in a straight-forward way:

$$(68) \text{ a. } \left[\begin{array}{l} \text{SYNSEM|LOC|CAT} \left[\begin{array}{l} \text{HEAD} \quad [1] \\ \text{SUBCAT} \quad [2] \end{array} \right] \\ \\ \text{HEAD-DTR} [4] | \text{SYNSEM|LOC|CAT} \left[\begin{array}{l} \text{HEAD} \quad [1] \\ \text{SUBCAT} \quad [2] \oplus \langle [3] \rangle \end{array} \right] \\ \\ \text{NON-HEAD-DTRS} \langle [5] [\text{SYNSEM} [3]] \rangle \\ head\text{-argument-phrase} \end{array} \right] \rightarrow [4], [5]$$

$$\text{ b. } \left[\begin{array}{l} \text{SYNSEM|LOC|CAT} \left[\begin{array}{l} \text{HEAD} \quad [1] \\ \text{SUBCAT} \quad [2] \end{array} \right] \\ \\ \text{HEAD-DTR} [4] | \text{SYNSEM|LOC|CAT} \left[\begin{array}{l} \text{HEAD} \quad [1] \\ \text{SUBCAT} \quad [2] \oplus \langle [3] \rangle \end{array} \right] \\ \\ \text{NON-HEAD-DTRS} \langle [5] [\text{SYNSEM} [3]] \rangle \\ head\text{-argument-phrase} \end{array} \right] \rightarrow [5], [4]$$

The left hand side of the rule is the mother node of the tree, that is, the sign that is licensed by the schema provided that the daughters are present. The right hand side in (68a) consists of the head daughter [4] followed by the non-head daughter [5]. We have the opposite order in (68b), that is, the head daughter follows the non-head daughter. The two orders correspond to the two orders that are permitted by LP-rules: The head precedes its argument if it is marked INITIAL+ and it follows it if it is marked INITIAL−.

The following code shows how (69b) is implemented in TRALE:

```
arg_h ## (head_argument_phrase,
```

```

        synsem:loc:cat:head:initial:minus,
        head_dtr:HeadDtr,
        non_head_dtrs:[NonHeadDtr]
    )
    ==>
cat> NonHeadDtr,
cat> HeadDtr.

```

A rule starts with an identifier that is needed for technical reasons like displaying intermediate structures in the parsing process in debugging tools. A description of the mother node follows and after the arrow we find a list of daughters, each introduced by the operator `cat>`.¹⁸ Structure sharing is indicated by values with capital letters. The above TRALE rule is a computer-readable variant of (68b), but includes the explicit specification of the value of INITIAL.

Now, the translation of a parallel schema using a MOTHER feature like (69a) into a phrase structure rule is almost as simple:

$$\begin{array}{lcl}
 (69) & \text{a.} & \left[\begin{array}{l} \text{MOTHER} | \text{SYNSEM} | \text{LOC} | \text{CAT} \left[\begin{array}{l} \text{HEAD} \quad [1] \\ \text{SUBCAT} \quad [2] \end{array} \right] \\ \text{HEAD-DTR} [4] | \text{SYNSEM} | \text{LOC} | \text{CAT} \left[\begin{array}{l} \text{HEAD} \quad [1] \\ \text{SUBCAT} \quad [2] \oplus \langle [3] \rangle \end{array} \right] \\ \text{NON-HEAD-DTRS} \langle [5] [\text{SYNSEM} [3]] \rangle \\ \text{head-argument-cx} \end{array} \right] \\
 & \text{b.} & [6] \rightarrow [4], [5] \text{ where } \left[\begin{array}{l} \text{MOTHER} [6] | \text{SYNSEM} | \text{LOC} | \text{CAT} \left[\begin{array}{l} \text{HEAD} \quad [1] \\ \text{SUBCAT} \quad [2] \end{array} \right] \\ \text{HEAD-DTR} [4] | \text{SYNSEM} | \text{LOC} | \text{CAT} \left[\begin{array}{l} \text{HEAD} \quad [1] \\ \text{SUBCAT} \quad [2] \oplus \langle [3] \rangle \end{array} \right] \\ \text{NON-HEAD-DTRS} \langle [5] [\text{SYNSEM} [3]] \rangle \\ \text{head-argument-cx} \end{array} \right]
 \end{array}$$

(69b) is only one of the two phrase structure rules that correspond to (69a), but since the other one differs from (69b) just in the ordering of [4] and [5] it is not given here.

For grammars in which the order of the elements corresponds to the observable order of the daughters, the connection to phrase structure rules is even simpler:

$$(70) \quad [1] \rightarrow [2] \text{ where } \left[\begin{array}{l} \text{MOTHER} \quad [1] \\ \text{DTRS} \quad [2] \\ \text{construction} \end{array} \right]$$

¹⁸ Other operators are possible in TRALE. For instance, `sem_head` can be used to guide the generator. This is control information that has nothing to do with linguistic theory and not necessarily with the way humans process natural language. There is also a `cats` operator, which precedes lists of daughters. This can be used to implement flat phrase structures.

The value of `DTRS` is a list and hence `⊂` stands for the list of daughters at the right hand side of the phrase structure rule as well. The type *construction* is a supertype of all constructions and hence (70) can be used to analyze all phrases that are licensed by the grammar. In fact (70) is one way to put the meta constraint in (31).

So, this shows that the version of SBCG that has been developed by Sag (2012) has a straightforward implementation in TRALE.¹⁹ The question remains whether *SBCG's performance-independence hypothesis remains conjecture until proven otherwise* as van Trijp sees it. The answer is: It is not a conjecture, since any of the old constraint solving system of the nineties could be used to process SBCG. The question whether this is efficient is an engineering problem that is entirely irrelevant for theoretical linguistics. Theoretical linguistics is concerned with human languages and how it is processed by humans. So whether some processing system that comes without claims about human language processing is efficient or not is absolutely irrelevant. Phrase structure-based backbones therefore are irrelevant as well, provided they refer to the grammar as described in theoretical work.

Now, this begs the question whether there is a contradiction in my claims. On page 324 I pointed out that SBCG is lacking a formalization in Richter's framework. Richter and also Levine and Meurers pointed out that there are problems with certain theoretically possible expressions and it is these expressions that mathematical linguists care about. So the goal is to be sure that any HPSG grammar has a meaning and that it is clear what it is. So this goal is much more foundational than writing a single grammar for a particular fragment of a language. There is no such foundational work for FCG, since FCG is a set of certain implemented grammars.

10.6.4.5.3 Static constraints vs. dynamic mappings and signature and grammar vs. open-endedness

The cool thing about Fluid Construction Grammar is its fluidity, that is there are certain constraints that can be adapted if there is pressure, the inventory of the theory is open-ended, so categories and features can be added if need be.

Again this is not a fundamental difference between HPSG/SBCG and FCG. An HPSG grammar fragment of a specific language is a declarative representation of linguistic knowledge and as such it of course just represents a certain fragment and does not contain any information how this set of constraints evolved or how it is acquired by speakers. For this we need specific theories about language evolution/language change/language acquisition. This is parallel to what we said about the competence/performance distinction, in order to account for language evolution we would have to have several HPSG grammars and say something about how one developed from another. This will involve weighted constraints, it will involve recategorization of linguistic items and lots more.²⁰

¹⁹ A toy fragment of English using a *MOTHER* feature and phrase structure rules with specifications of the kind given above can be downloaded at <http://hpsg.fu-berlin.de/Fragments/SBCG-TRALE/>.

²⁰ There are systems that use weighted constraints. We had a simple version of this in the German HPSG grammar that was developed in *Verbmobil* project (Müller & Kasper 2000) already and there are much more sophisticated, corpus-based systems around today. Usually such weighted constraints are not part

So basically HPSG has to be extended, has to be paired with a model about language evolution in the very same way as FCG is.

10.6.4.5.4 Theoretical physics vs. Darwinian evolutionary theory

Van Trijp compares SBCG and FCG and claims that SBCG follows the model of theoretical physics – like Chomsky does, while FCG adopts a Darwinian model of science – like Croft does, the difference being that SBCG makes certain assumptions that are true of all languages, while FCG does not make any a priori assumptions. The fundamental assumptions that are made in both theories are that the objects that we model are best described by feature value pairs (a triviality). FCG assumes that there is always a syntactic and a semantic pole (fundamental assumption in the system) and researchers working in HPSG/SBCG assume that if languages have certain phenomena they will be analyzed in similar ways. For instance, if a language has nonlocal dependencies, these will be analyzed via the *SLASH* mechanism. However, this does not entail that one believes that grammars of all languages have a *SLASH* feature. And in fact there even may be languages that do not have valence features (Koenig & Michelson 2010), which may be a problem for FCG since it relies on the *SYN*-pole for the matching phase. So, as far as SBCG is concerned, there is considerable freedom to choose features that are relevant in an analysis and of course additional features and types can be assumed in case a language is found that provides evidence for this. The only example of a possibly too strong constraint that van Trijp provides is the locality constraint imposed by the *MOTHER* feature. The idea about this feature is that everything that is of relevance in a more nonlocal context has to be passed up explicitly. This is done for nonlocal dependencies (via *SLASH*) and for instance also for information concerning prepositions (via *PFORM*). Certain verbs require prepositional objects and restrict the form of the preposition. For instance, *wait* has to make sure that its prepositional object has the preposition *for* in it. Since this information is usually available only at the preposition, it has to be passed up to the PP level in order to be directly selectable by the governing verb.

(71) I am waiting for my man.

So, assuming strict locality of selection requires that all phenomena that cannot be treated locally have to be analyzed by passing information up. Assuming strict locality is a design decision that does not have any empirical consequences, as far as it does not rule out any language or construction in principle. It just requires that information has to be passed up that needs to be accessed at higher nodes. As I have shown in Section 10.6.2, the locality constraint is easily circumvented even within SBCG and it makes the analysis of idioms unnecessarily complicated and unintuitive, so I suggest dropping the *MOTHER* feature. But even if *MOTHER* is kept, it is not justified to draw a distinction between SBCG and FCG along the lines that were suggested by van Trijp.

Independent of the *MOTHER* issue, the work done in the CoreGram project (Müller 2013a; 2015a) shows that one can derive generalizations in a bottom-up fashion rather

of theoretical papers, but there are exceptions as for instance the paper by Briscoe and Copestake about lexical rules (Briscoe & Copestake 1999).

than imposing constraints on grammars in a top-down way. The latter paper discusses Croft's methodological considerations and shows how methodological pitfalls are circumvented in the project. HPSG/SBCG research differs from work in Chomskyan frameworks in not trying to show that all languages are like English or Romance or German or whatever, rather languages are treated on their own as it is common in the Construction Grammar community. This does not imply that there is no interest in generalizations and universals or near universals or tendencies, but again the style of working and the rhetoric is usually different. Therefore I think that the purported difference between SBCG and FCG does not exist.

10.6.4.5.5 Permissiveness of the theories

Van Trijp claims that HPSG/SBCG is a “generative grammar” since its aim is to account for and admit only grammatical sentences. FCG on the other hand is more permissive and tries to get the most out of the input even if it is fragmentary or ungrammatical (see also Steels 2013: 166). While it is an engineering decision to be able to parse ungrammatical input – and to be sure there are systems for the robust processing of HPSG grammars (Kiefer, Krieger & Nederhof 2000; Copestake 2007), it is also clear that humans cannot parse everything. There are strong constraints whose violations cause measurable effects in the brain. This is something that a model of language (that includes competence and performance factors or does not make the difference at all) has to explain. The question is what the cause of deviance is: Is it processing complexity? Is it a category mismatch? A clash in information structure? So, if FCG permits structures that are not accepted by human native speakers and that do not make any sense whatsoever, additional constraints have to be added. If they are not added, the respective FCG theory is not an adequate theory of the language under consideration. Again, there is no difference between HPSG/SBCG and FCG.

10.6.4.5.6 A note on engineering

My biggest problem with FCG is that linguistic and engineering aspects are mixed.²¹ Certain bookkeeping features that are needed for technical reasons only appear in linguistic papers, technical assumptions that are made to get a parser running are mixed with linguistic constraints. Bit vector encodings that are used to represent case information are part of papers about interesting case systems. To be sure there is nothing bad about bit vector encodings. They are used in HPSG implementations as well (Reape 1991: 55; Müller 1996c: 269), but this is not mixed into the theoretical papers.

It was a big breakthrough in the 80ties when theoretical linguists and computational linguists worked together and developed declarative formalisms that were independent of specific parsers and processing systems. This made it possible to take over insights

²¹ This is not a problem if all FCG papers are read as papers documenting the FCG-system (see Footnote 14 on page 329), since then it would be necessary to include these technical details. If the FCG papers are to be read as theoretical linguistics papers that document a certain Construction Grammar analysis, the Lisp statements and the implementational details are simply in the way.

from a lot of linguists who were not concerned with the actual implementation but took care of finding linguistic generalizations and specifying constraints. Since this separation is given up in FCG it will remain an engineering project without much appeal to the general linguist.

10.7 Summary and classification

There are currently three formalized variants of Construction Grammar: Sign-Based Construction Grammar, Embodied Construction Grammar, and Fluid Construction Grammar. The first two variants can be viewed as notational variants of (Constructional) HPSG (for SB-CxG with regard to this point, see Sag (2007: 411) and Sag (2010: 486)), or put differently, sister theories of HPSG. This is also true to a large extent for FCG, although van Trijp (2013) spends 25 pages to work out the differences. As I have shown in Section 10.6.4, HPSG and FCG are rather similar and I would say that these theories are sister theories as well.

Due to the origins of all three theories, respective analyses can differ quite considerably: HPSG is a strongly lexicalized theory, where phrasal dominance schemata have only been increasingly more used in the last ten years under the influence of Ivan Sag. The phrasal dominance schemata that Ivan Sag uses in his work are basically refinements of schemata that were present in earlier versions of HPSG. Crucially, all phenomena that interact with valence receive a lexical analysis (Sag et al. 2012: Section 2.3). In CxG, on the other hand, predominantly phrasal analyses are adopted which is due to the influence of Adele Goldberg.

As already emphasized in Chapter 9, these are only tendencies that do not apply to all researchers working in the theories in question.

Exercises

1. Find three examples of utterances whose meaning cannot be derived from the meaning of the individual words. Consider how one could analyze these examples in Categorical Grammar (yes, Categorical Grammar).

Further reading

There are two volumes on Construction Grammar in German: Fischer & Stefanowitsch (2006) and Stefanowitsch & Fischer (2008). Deppermann (2006) discusses Construction Grammar from the point of view of conversational analysis. The 37(3) volume of the *Zeitschrift für germanistische Linguistik* from 2009 was also devoted to Construction Grammar. Goldberg (2003a) and Michaelis (2006) are overview articles in English. Goldberg's books constitute important contributions to Construction Grammar (1995; 2006; 2009). Goldberg (1995) has argued against lexicon-based analyses such as those common in GB, LFG, CG and HPSG. These arguments can be invalidated, however, as

will be shown in Section 23.7.1. Sag (1997), Borsley (2006), Jacobs (2008) and Müller & Lipenkova (2009) give examples of constructions that require a phrasal analysis if one wishes to avoid postulating empty elements. Jackendoff (2008) discusses the noun-preposition-noun construction that can only be properly analyzed as a phrasal construction (see Section 23.10). The discussion on whether argument structure constructions should be analyzed phrasally or lexically (Goldberg 1995; 2006; Müller 2006) culminated in a series of papers (Goldberg 2013a) and a target article by Müller & Wechsler (2014a) with several responses in the same volume.

Tomasello's publications to language acquisition (Tomasello 2000; 2003; 2005; 2006c) constitute a Construction Grammar alternative to the Principle & Parameters theory of acquisition as it does not have many of the problems that P&P analyses have (for more on language acquisition, see Section 16). For more on language acquisition and Construction Grammar, see Behrens (2009).

Dąbrowska (2004) looks at psycholinguistic constraints for possible grammatical theories.

11 Dependency Grammar

Dependency Grammar¹ is the second oldest framework described in this book. It was developed by the French linguist Lucien Tesnière (1893–1954). His foundational work *Éléments de syntaxe structurale* ‘Elements of structural syntax’ was basically finished in 1938 only three years after Ajdukiewicz’ paper on Categorical Grammar (1935), but the publication was delayed until 1959, five years after his death. Since valency is central in Dependency Grammar, it is sometimes also referred to as Valency Grammar. Tesnière’s ideas are wide-spread nowadays. The conceptions of valency and dependency are present in almost all of the current theories (Ágel & Fischer 2010: 262–263, 284).

Dependency Grammar is most popular in central Europe and especially so in Germany (Engel 1996: 56–57). Ágel & Fischer (2010: 250) identified a possible reason for this: Tesnière’s original work was not available in English until very recently (Tesnière 2015), but there is a translation to German for more than 35 years now (Tesnière 1980). Since dependency grammar focuses on dependency relations rather than linearization of constituents, it is often felt to be more appropriate for languages with freer constituent order, which is one reason for its popularity among researchers working on Slavic languages: The New Prague School (Sgall, Hajičová, Panevova) developed dependency grammar further beginning in the 1960s (Hajičová & Sgall 2003). Igor A. Mel’čuk, A. K. Žolkovskij worked since the 1960s in the Soviet Union on a model called Meaning–text theory, which was also used in Machine Translation projects (Mel’čuk 1964; 1981; 1988; Kahane 2003). Mel’čuk left the Soviet Union towards Canada in the 1970s and now works in Montréal.

Dependency Grammar is very wide-spread in Germany and among Germanists worldwide. It is used very successfully for teaching German as a foreign language (Helbig & Buscha 1969; 1998). Helbig and Buscha, who worked in Leipzig, East Germany, started to compile valency dictionaries (Helbig & Schenkel 1969) and later researchers working at the Institut für Deutsche Sprache (Institute for German Language) in Mannheim started similar lexicographic projects (Schumacher et al. 2004).

The following list provides a probably incomplete list of linguists who are/were based in Germany: Vilmos Ágel (2000), Kassel; Klaus Baumgärtner (1965; 1970), Leipzig later Stuttgart; Ulrich Engel (1977; 2014), IDS Mannheim; Hans-Werner Eroms (1985; 1987; 2000), Passau; Heinz Happ, Tübingen; Peter Hellwig (1978; 2003), Heidelberg; Jürgen Heringer (1996), Augsburg; Jürgen Kunze (1975), Berlin; Henning Lobin (1993), Gießen; Klaus Schubert (1987), Hildesheim; Heinz Josef Weber (1997), Trier; Klaus Welke (1988;

¹ This chapter is a rough draft. Comments on this chapter are especially welcome. References are missing. If you have pointers to relevant literature, please let me know. I am especially looking for implementations of parsers/generators and grammars.

2011), Humboldt University Berlin.

Although there was work done in many countries and continuously over the decades since 1959, a periodical international conference was established as late as 2011.^{2,3}

Dependency Grammar was used in computational projects from early on. Meřčuk worked on Machine Translation in the Soviet Union (Meřčuk 1964) and David G. Hays worked on Machine Translation in the United States (Hays & Ziehe 1960). Jürgen Kunze, based in East Berlin at the German Academy of Sciences, where he had a chair for computational linguistics, also started to work on machine translation in the 1960s. A book that describes the formal background of the linguistic work was published as Kunze (1975). Various researchers worked in the Collaborative Research Center 100 *Electronic linguistic research* (SFB 100, Elektronische Sprachforschung) from 1973–1986 in Saarbrücken. The main topic of this SFB was machine translation as well. There were projects on Russian to German, French to German, English to German, and Esperanto to German translation. For work from Saarbrücken in this context see Klein (1971), Rothkegel (1976), and Weissgerber (1983). Muraki et al. (1985) used Dependency Grammar in a project that analyzed Japanese and generated English. Richard Hudson started to work in a dependency grammar-based framework called Word Grammar in the 1980s (Hudson 1984; 2007) and Sleator and Temperly have been working on Link Grammar since the 1990s (?).

In recent years Dependency Grammar got more and more popular among computational linguists. The reason for this is that there are many annotated corpora (tree banks) that contain dependency information.⁴ Statistical parsers are trained on such tree banks (Yamada & Matsumoto 2003; Attardi 2006; Nivre 2003). Many of the parsers work for multiple languages since the general approach is language independent. It is easier to annotate dependencies consistently since there are fewer possibilities to do so. While syntacticians working in constituency-based models may assume binary branching or flat models, high or low attachment of adjuncts, empty elements or no empty elements and fiercely fight about this, it is fairly clear what the dependencies in an utterance are. Therefore it is easy to annotate consistently and train statistical parsers on such annotated data.

Apart from statistical modeling there are also so-called deep processing systems, that is, systems that rely on a hand-crafted, linguistically motivated grammar. I already mentioned Meřčuk's work in the context of machine translation, Hays & Ziehe (1960) had a parser for Russian, Starosta & Nomura (1986) developed a parser that was used with an English grammar, Jäppinen, Lehtola & Valkonen (1986) developed a parser that was demoed with Finnish, Hellwig (1986; 2003; 2006) implemented grammars of German in the framework of Dependency Unification Grammar, Hudson (1989) developed a Word Grammar for English, Covington (1990) developed a parser for Russian and Latin, which can parse discontinuous constituents, and Menzel (1998) did robust parsing of a dependency grammar of German. Other work on computational parsing to be mentioned is

² <http://depling.org/>. 10.04.2015.

³ A conference on Meaning Text Theory is taking place biannually since 2003.

⁴ According to Kay (2000) the first treebank whatsoever was developed by Hays and did annotate dependencies.

Kettunen (1986); Lehtola (1986); Menzel & Schröder (1998b). The following is a list of languages for which Dependency Grammar fragments exist:

- English (Muraki et al. 1985; Starosta & Nomura 1986; Hudson 1989; ?)
- Finnish (Nelimarkka, Jäppinen & Lehtola 1984; Jäppinen, Lehtola & Valkonen 1986)
- German (Hellwig 1986; 2003; 2006; Menzel & Schröder 1998a; Gerdes & Kahane 2001)
- Japanese (Muraki, Ichiyama & Fukumochi 1985)
- Latin (Covington 1990)
- Old Icelandic (Maas 1977)
- Russian (Hays & Ziehe 1960; Mel'čuk 1964; Covington 1990)

11.1 General remarks on the representational format

11.1.1 Valency information, nucleus and satellites

The central concept of Dependency Grammar is valency (see Section 1.6). The central metaphor for this is the formation of stable molecules, which is explained in chemistry with reference to layers of electrons. A difference between chemical compounds and linguistic structures is that the compounding is not directed, that is, it would not make sense to claim that oxygen is more important than hydrogen in forming water. In contrast to this the verb is more important than the nominal phrases it combines with to form a complete clause. The verb determines the form of its dependents.

One way to depict dependencies is shown in Figure 11.1. The highest node is the verb

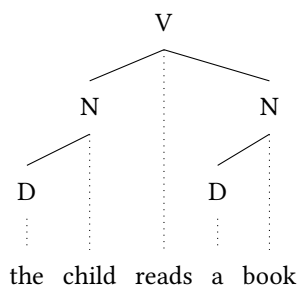


Figure 11.1: Analysis of *The child reads a book*.

reads. Its valency is a nominative NP (the subject) and an accusative NP (an object). This is depicted by the dependency links between the node representing the verb and the

nodes representing the respective nouns. The nouns themselves require a determiner, which again is shown by the dependency links to *the* and *a* respectively. Note that the analysis presented here corresponds to the NP analysis that is assumed for instance in HPSG, that is, the noun selects its specifier (see Section 9.6). It should be noted though that the discussion whether an NP or a DP analysis is appropriate was also led within the Dependency Grammar framework (Hudson 1984: 90; Van Langendonck 1994; Hudson 2004). See Engel (1977) for an analysis with the N as head and Welke (2011: 31) for an analysis with the determiner as head.

The verb is the head of the clause and the nouns are called *dependents*. Alternative terms for head and dependent are *nucleus* and *satellite*, respectively.

An alternative way to depict the dependencies is provided in Figure 11.2. This graph

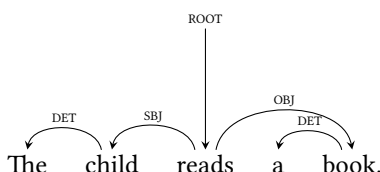


Figure 11.2: Alternative presentation of the analysis of *The child reads a book*.

displays the grammatical functions rather than part of speech information, but apart from this it is equivalent to the representation in Figure 11.1. What is the highest node in Figure 11.1 is labeled with the ROOT arrow. Downward links are indicated by the direction of the arrows.

A third form of representing the same dependencies provided in Figure 11.3 has the tree format again. This tree results if we pull the root node in Figure 11.2 upwards. Since we

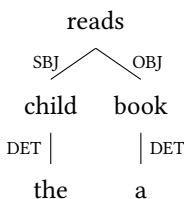


Figure 11.3: Alternative presentation of the analysis of *The child reads a book*.

have a clear visualization of the dependency relation that represents the nucleus above the dependents, we do not need to use arrows to encode this information. Of course part of speech information can be added to the Figures 11.2 and 11.3, grammatical function labels could be added to Figure 11.1, and word order can be added to Figure 11.3.

The above figures depict the dependency relation that holds between a head and the respective dependents. This can be written down more formally as an n-ary rule that is

similar to phrase structure rules that were discussed in Chapter 2 (Baumgärtner 1970: 61; Heringer 1996: Section 4.1). For instance Baumgärtner suggests the rule in (1):

$$(1) \quad \chi \rightarrow \varphi_1 \dots \varphi_i * \varphi_{i+2} \dots \varphi_n, \text{ where } 0 < i \leq n$$

The asterisk in (1) corresponds to the word of the category χ . In our example, χ would be V, the position of the ‘*’ would be taken by *reads*, and φ_1 and φ_3 would be N. Together with the rule in (2b) for the determiner-noun combination, the rule in (2a) would license the dependency tree in Figure 11.1.

- (2) a. $V \rightarrow N * N$
 b. $N \rightarrow D *$

Alternatively several binary rules can be assumed that combine a head with its subject, direct object, or indirect object (Kahane 2009). Dependency rules will be discussed in more detail in Section 11.7.2, where dependency grammars are compared with phrase structure grammars.

11.1.2 Adjuncts

Another metaphor that was used by Tesnière is the drama metaphor. The core participants of an event are the *actants* and apart from this there is the background, the stage, the general setting. The actants are the arguments in other theories and the stage-describing entities are called *circumstants*. These circumstants are modifiers and usually analyzed as adjuncts in the other theories described in this book. As far as the representation of dependencies is concerned, there is not much of a difference between arguments and adjuncts in Dependency Grammar: Figure 11.4 shows the analysis of (3):

- (3) The child often reads the book slowly.

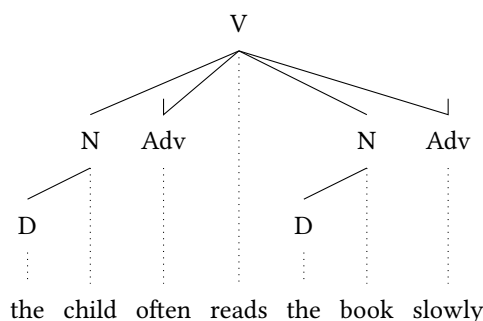


Figure 11.4: Analysis of *The child often reads the book slowly.*

The dependency annotation uses a technical device suggested by Engel (1977) to depict different dependency relations: Adjuncts are marked with an additional line upwards

from the adjunct node (see also Eroms (2000)). An alternative way to specify the argument/adjunct or rather the actant/circumstant distinction is of course an explicit specification of the status as argument or adjunct. So like the grammatical functions in the previous figures one can use explicit labels to this extent. German grammars and valency dictionaries often use the labels E and A for *Ergänzung* and *Angabe*, respectively.

11.1.3 Linearization

So far we have seen dependency graphs that had connections to words that were linearized in a certain order. However, in principle the order of the dependents is not determined by the dependency and therefore a Dependency Grammar has to contain additional statements that take care of the proper linearization of linguistic objects (stems, morphemes, words). Engel (2014: 50) assumes the dependency graph in Figure 11.5 for the sentences in (4).⁵

- (4) a. Gestern war ich bei Tom.
yesterday was I with Tom
'I was with Tom yesterday.'
- b. Ich war gestern bei Tom.
I was yesterday with Tom
- c. Bei Tom war ich gestern.
with Tom was I yesterday
- d. Ich war bei Tom gestern.
I was with Tom yesterday

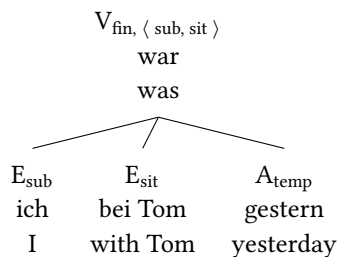


Figure 11.5: Dependency graph for several orders of *ich*, *war*, *bei Tom*, and *gestern* 'I was with Tom yesterday.' according to Engel (2014: 50)

According to Engel (2014: 50), the correct order is enforced by surface syntactic rules as for instance the rule that states that there is always exactly one element in the Vorfeld in declarative main clauses, the finite verb is in second position.⁶ Furthermore, there

⁵ Engel uses E_{sub} for the subject and E_{acc}, E_{dat}, and E_{gen} for the objects with respective cases.

⁶ "Die korrekte Stellung ergibt sich dann zum Teil aus oberflächensyntaktischen Regeln (zum Beispiel: im Vorfeld des Konstativsatzes steht immer genau ein Element; das finite Verb steht an zweiter Stelle [...])"

are linearization rules that concern pragmatic properties like given before new information, and weak pronouns have to be placed into the Vorfeld or at the beginning of the Mittelfeld. This conception of linear order is problematic both for empirical and conceptual reasons and we will turn to it again in Section 11.7.1. It should be noted here that approaches that deal with dependency alone admit discontinuous realizations of heads and their dependents. Without any further constraints Dependency Grammars would have a problem that was already discussed on page 327 in Section 10.6.3 on Embodied Construction Grammar: One argument could interrupt another argument as in Figure 11.6. In order to exclude such linearizations in languages in which they are im-

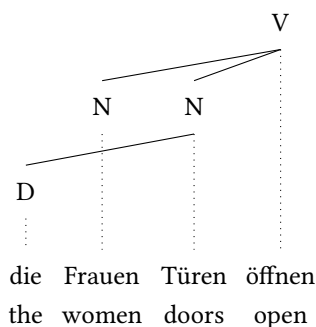


Figure 11.6: Unwanted analysis of *dass die Frauen Türen öffnen* ‘that the women open doors’

possible, it is sometimes assumed that analyses have to be projective, that is crossing branches like those in Figure 11.6 are not allowed. This basically reintroduces the concept of constituency into the framework, since this means that all dependents of a head have to be realized close to the head unless special mechanisms for liberation are used (see for instance Section 11.5 on nonlocal dependencies). Some authors explicitly use a phrase structure component to be able to formulate restrictions on serializations of constituents (Gerdes & Kahane 2001; Hellwig 2003).

11.1.4 Semantics

Tesnière already distinguished the participants of a verb in a way that was later common in theories of semantic roles. He suggested that the first actant is the agent, the second one a patient and the third a benefactive (Tesnière 2015: Chapter 106). Given that Dependency Grammar is a lexical framework, all lexical approaches to argument linking can be taken over. However, argument linking and semantic role assignment is just a small part of the problem that has to be solved when natural language expressions have to be assigned a semantics. Issues regarding the scope of adjuncts and quantifiers have to be solved and it is clear that dependency graphs that represent dependencies without taking into account linear order are not sufficient. An unordered dependency graph assigns

grammatical functions to a dependent of a head and hence is similar in many respects to an LFG f-structure.⁷ For a sentence like (25a) on page 226, repeated here as (5), one gets the f-structure in (25b) on page 226. This f-structure contains a subject (*David*), an object (*a sandwich*), and an adjunct set with two elements (*at noon* and *yesterday*).

- (5) David devoured a sandwich at noon yesterday.

This is exactly what is encoded in an unordered dependency graph. Because of this parallel it comes as no surprise that Bröker (2003: 308) suggested to use glue semantics (Dalrymple, Lamping & Saraswat 1993; Dalrymple 2001: Chapter 8) for Dependency Grammar as well. Glue semantics was introduced in Section 7.1.5 already.

11.2 Passive

Dependency Grammar is a lexical theory and valency is the key concept. Hence it is not surprising that the analysis of the passive is a lexical one. That is, it is assumed that there is a passive participle that has a different valency requirement than the active verb.

Our standard example in (6) is analyzed as shown in Figure 11.7.

- (6) [dass] der Weltmeister geschlagen wird
 that the world.champion beaten is
 ‘that the world champion is (being) beaten’

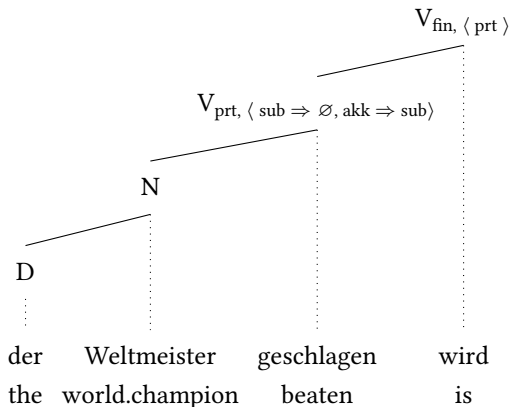


Figure 11.7: Analysis of [dass] *der Weltmeister geschlagen wird* ‘that the world champion is (being) beaten’ parallel to the analyses provided by Engel (2014: 53–54)

This figure is an intuitive depiction of what is going on in passive. A formalization of this would probably amount into a lexical rule for the personal passive. See Hellwig

⁷ Tim Osborne (p. c. 2015) reminds me that this is not true in all cases: For instance non predicative prepositions are not reflected in f-structures, but of course they are present in dependency graphs.

(2003: 629–630) for an explicit suggestion of a lexical rule for the analysis of the passive in English.

Note that *der Weltmeister* ‘the world champion’ is not an argument of the passive auxiliary *wird* ‘is’ in Engel’s analysis. This means that subject verb agreement cannot be determined locally and some elaborated mechanism has to be developed for ensuring agreement.⁸ Groß & Osborne (2009) and Eroms (2000: Section 5.3) assume that subjects depend on auxiliaries rather than on the main verb. This requires some argument transfer as it is common in Categorical Grammar (see Section 8.5.2) and HPSG (Hinrichs & Nakazawa 1994). The adapted analysis that treats the subject of the participle as a subject of the auxiliary is given in Figure 11.8.

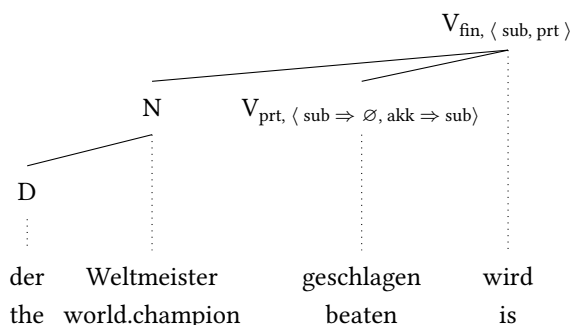


Figure 11.8: Analysis of [*dass*] *der Weltmeister geschlagen wird* ‘that the world champion is (being) beaten’ with the subject as dependent of the auxiliary

⁸ This problem would get even more pressing for cases of the so-called remote passive:

- (i) weil der Wagen zu reparieren versucht wurde
 because the car to repair tried was
 ‘because it was tried to repair the car’

Here the object of *zu reparieren*, which is the object of a verb that is two levels embedded, agrees with the auxiliary *wurde* ‘was’. However, the question how to analyze these remote passives is open in Engel’s system anyway and the solution of this problem probably involves the mechanism that was applied in HPSG: The arguments of *zu reparieren* are raised to the governing verb *versucht*, passive applies to this verb and turns the object into a subject which is then raised by the auxiliary, which explains the agreement between the underlying object of *zu reparieren* ‘to repair’ and *wurde* ‘was’. Hudson (1997), working in the framework of Word Grammar, suggests an analysis of verbal complementation in German that involves what he calls *generalized raising*, that is, he assumes that both subjects and complements may be raised to the governing head. Note that such an analysis involving generalized raising would make an analysis of sentences like (i) straight forward, since the object would depend on the same head as the subject, namely on *hat* ‘has’ and hence can be ordered before the subject.

- (ii) Gestern hat sich der Spieler verletzt.
 yesterday has self the player injured
 ‘The player injured himself yesterday.’

For a discussion of Groß & Osborne’s account of (ii) see page 566.

11.3 Verb position

Dependency grammar does not focus on linearization aspects. The dependency relations between a verb and its arguments are basically the same in verb-initial and verb-final sentences. If we compare the dependency graphs of the sentences in (7), given in Figure 11.9 and 11.10 we see that only the position of the verb is different, but the dependency relation is the same, as it should be.⁹

- (7) a. [dass] jeder diesen Mann kennt
 that everybody this man knows
 ‘that everybody knows this man’
 b. Kennt jeder diesen Mann?
 knows everybody this man
 ‘Does everybody know this man?’

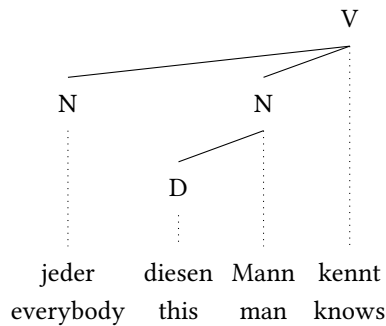


Figure 11.9: Analysis of [dass] *jeder diesen Mann kennt* ‘that everybody knows this man’

The correct linearization of the verb with respect to its arguments and adjuncts is ensured by linearization constraints that refer to the respective topological fields. See Section 11.1.3 and Section 11.7.1 for further details on linearization.

⁹ Eroms (2000) uses the part of speech *Pron* for pronouns like *jeder* ‘everybody’. If part of speech information plays a role in selection, this makes necessary a disjunctive specification of all valency frames of heads that govern nominal expressions, since they can either combine with an NP with internal structure or with a pronoun. By assigning pronouns the category *N* such a disjunctive specification is avoided. A pronoun differs from a noun in its valency (it is fully saturated, while a noun needs a determiner), but not in its part of speech. Eroms & Heringer (2003: 259) use the symbol *N_pro* for pronouns. If the *pro*-part is to be understood as a special property of items with the part of speech *N*, this is compatible with what I have said above: Heads could then select for *Ns*. If *N_pro* and *N* are assumed to be distinct, atomic symbols, the problem remains.

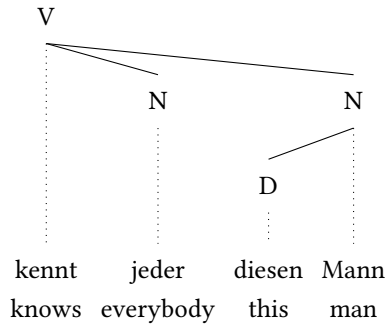


Figure 11.10: Analysis of *Kennt jeder diesen Mann?* ‘Does everybody know this man?’

11.4 Local reordering

The situation regarding local reordering is the same, the dependency relations of the sentence in (8b) is shown in Figure 11.11. The analysis of the sentence with normal order in (8a) was given in Figure 11.9 already.

- (8) a. [dass] jeder diesen Mann kennt
 that everybody this man knows
 ‘that everybody knows this man’
 b. [dass] diesen Mann jeder kennt
 that this man everybody knows
 ‘that everybody knows this man’

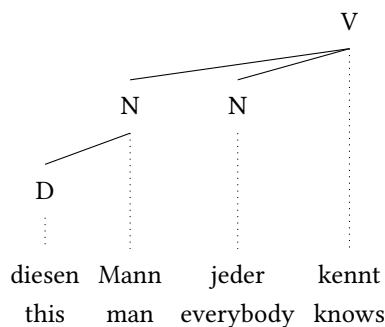


Figure 11.11: Analysis of [dass] *diesen Mann jeder kennt* ‘that everybody knows this man’

11.5 Long-distance dependencies

There are several possibilities to analyze nonlocal dependencies in Dependency Grammar. The easiest is the one we already saw in the previous sections. One just focuses on the dependency relations and assumes that the order with the verb in second position is just one of the possible linearization variants. Figure 11.12 shows the analysis of (9):

- (9) [Diesen Mann] kennt jeder.
 this man knows everybody
 ‘Everyone knows this man.’

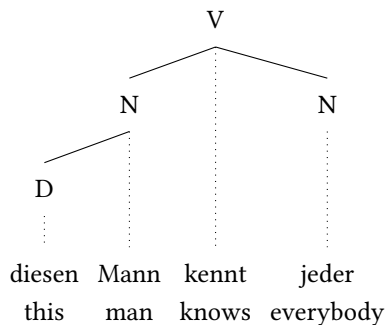


Figure 11.12: Analysis of *Diesen Mann kennt jeder*. ‘This man, everybody knows.’ without special treatment of fronting

Now, this is the simplest case, so let us look at the example in (10), which really involves a *nonlocal* dependency:

- (10) Wen_i glaubst du, daß ich _i gesehen habe.¹⁰
 who believe you that I seen have
 ‘Who do you think I saw?’

The dependency relations are depicted in Figure 11.13 on the next page. This graph differs from most graphs we have seen before by not being projective, which means that there are crossing lines: The connection between V_{prt} and the N for *wen* ‘who’ crosses the lines connecting *glaubst* ‘believes’ and *du* ‘you’ with their heads. Depending on the version of Dependency Grammar that is assumed this is seen as a problem or it is not. Let us explore the two options: If discontinuity of the type shown in Figure 11.13 is allowed for, there has to be something in the grammar that excludes discontinuities that are ungrammatical. For instance, an analysis of (11) as in Figure 11.14 on page 364 should be excluded.

¹⁰ Scherpenisse (1986: 84).



(11) *Wen glaubst ich du, dass gesehen habe?
 who believes I.NOM you.NOM that seen has
 Intended: 'Who do you think I saw?'

An alternative to assuming discontinuous constituents is to assume additional mechanisms that promote the dependency of an embedded head to a higher head in the structure. Such an analysis was suggested by Kunze (1968), Hudson (1997; 2000), Kahane (1997), Kahane et al. (1998), and Groß & Osborne (2009). Kahane, Nasr & Rambow (1998) call the respective operation *lifting* and Groß & Osborne call it *rising*. In what follows I use the analysis by Groß & Osborne (2009) as an example for such analyses. Groß & Osborne depict the reorganized dependencies with a dashed line as in Figure 11.15 on the next page.^{11, 12} The origin of the dependency (V_{prt}) is marked with a *g* and the dependent

¹² Note that Groß & Osborne (2009) do not assume a uniform analysis of simple and complex V2 sentences. That is, for cases that can be explained as local reordering they assume an analysis without rising. Their

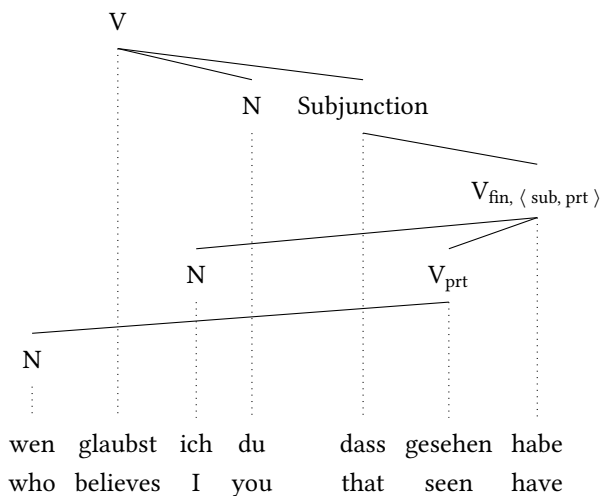


Figure 11.14: Unwanted dependency graph of **Wen glaubst ich du, dass gesehen habe?* ‘Who do you think I saw?’

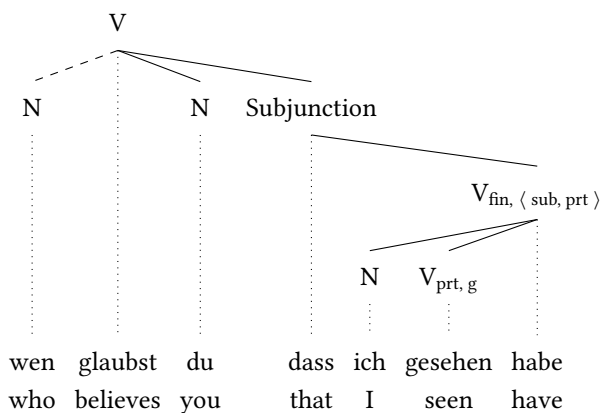


Figure 11.15: Projective analysis of *Wen glaubst du, dass ich gesehen habe?* ‘Who do you think I saw?’ involving rising

is connected to the node to which it has risen (the top-most V) by a dashed line.¹³ Instead of realizing the accusative dependent of *gesehen* ‘seen’ locally, the information about the missing element is transferred to a higher node and realized there.

The constituent that rose cannot be attached in an arbitrary position of the tree. It has to go to a place higher in the tree. Groß & Osborne (2009: 80) use the concepts of *chain* (called *catena* in later work (Osborne, Putnam & Groß 2012)) to formulate the respective constraints.

Def. 1 (Catena) *A word or a combination of words that is top-down (or bottom-up) continuous.*

The head of a given chain is defined as the word that immediately dominates that chain and the governor of a chain is the word that licenses the elements of that chain. Usually the head and the governor are identical, but if dislocation is involved, the head is a node that dominates the governor. For instance in Figure 11.15, *glaubt* ‘believe’ is the head, but *gesehen* ‘seen’ is the governor.

Principle 7 (Rising Principle) *The head of a given chain must either be that chain’s governor or dominate that chain’s governor.*

The Rising Principle ensures that “movement”¹⁴ is upwards, that is, a constituent may not attach to a node to its left or right that is deeper in the tree or to a node that is higher in the tree but not adjacent to a node in a chain of governors. Groß & Osborne (2009: 54) illustrate this with Figure 11.16 on the following page. The node D is governed by C and hence can rise to all nodes dominating this node. So D may attach to E, but crucially not to A, B or F or G.

11.6 New developments and theoretical variants

This section mainly deals with Tesnière’s variant of Dependency Grammar. Section 11.6.1 deals with Tesnière’s part of speech system and Section 11.6.2 describes the modes of combinations of linguistics objects that Tesnière assumed.

analysis of (9) is the one that is depicted in Figure 11.12. This leads to problems that are discussed in Section 11.7.1.

¹³ Hudson (1997; 2000) makes the dependency between the verb that selects an extracted element and the extracted element explicit in his diagrams. He states that the use of multiple dependencies in Word Grammar corresponds to structure sharing in HPSG (Hudson 1997: 15).

¹⁴ Groß & Osborne (2009: 54) explicitly state that there is no literal movement involved: “Our dependency-based grammar is decidedly non-derivational. We do not assume that the risen chain ever appears as a dependent of its governor at some stage of a putative derivation below or beyond the surface. But rather the notion of rising is understood figuratively. The terms *rising* and *risen* are convenient metaphors for denoting a constellation in which a given chain has attached to a word that is not its governor.”

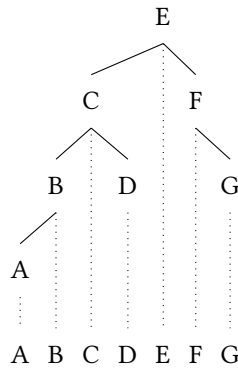


Figure 11.16: Abstract dependency tree used by Groß & Osborne (2009: 54) to illustrate the options for rising

11.6.1 Tesnière’s part of speech classification

As mentioned in the introduction, Tesnière is the founder of Dependency Grammar (Tesnière 1959; 1980; 2015). There are many versions of Dependency Grammar today and most of them use the part of speech labels that are used in other theories as well (N, P, A, V, Adv, Conj, ...). Tesnière had a system of four major categories: noun, verb, adjective, and adverb. The labels for these categories were derived from the endings that are used in Esperanto, that is, they are O, I, A, and E, respectively. These categories were defined semantically as specified in Table 11.1.¹⁵ Tesnière assumed these categories to be

Table 11.1: Semantically motivated part of speech classification by Tesnière

	substance	process
concrete	noun	verb
abstract	adjective	adverb

universal and suggested that there are limits in which way these categories may depend on others.

According to Tesnière nouns and adverbs may depend on verbs, adjectives may depend on nouns, and adverbs may depend on adjectives or adverbs. This situation is depicted in the general dependency graph in Figure 11.17 on the next page. The ‘*’ means

¹⁵ As Weber (1997: 77) points out this categorization is not without problems: In what sense is *Angst* ‘fear’ a substance? Why should *glauben* ‘believe’ be a concrete process? See also Klein (1971: Section 3.4) for the discussion of *schlagen* ‘to beat’ and *Schlag* ‘the beat’ and similar cases. Even if one assumes that *Schlag* is derived from the concrete process *schlag-* by a transfer into the category O, the assumption that such Os stand for concrete substances is questionable.

that there can be arbitrarily many dependencies between Es. Of course it is easy to find

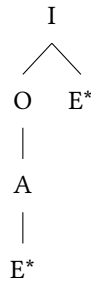


Figure 11.17: Universal configuration for dependencies according to Tesnière
(I = verb, O = noun, A = adjective, E = adverb)

examples in which adjectives depend on verbs and sentences (verbs) depend on nouns. Such cases are handled via so-called *transfers* in Tesnière's system. Furthermore, conjunctions, determiners, and prepositions are missing from this set of categories. For the combination of these elements with their dependents Tesnière used special combinatoric relations: junction and transfer. We will deal with these in the following subsection.

11.6.2 Connexion, junction, and transfer

Tesnière (1959) suggested three basic relations between nodes: connexion, junction, and transfer. Connexion is the simple relation between a head and its dependents that we already covered in the previous sections. Junction is a special relation that plays a role in the analysis of coordination and transfer is a tool that allows one to change the category of a lexical item or a phrase.

11.6.2.1 Junction

Figure 11.18 illustrates the junction relation: The two conjuncts *John* and *Mary* are connected with the conjunction *and*. It is interesting to note that both of the conjuncts are

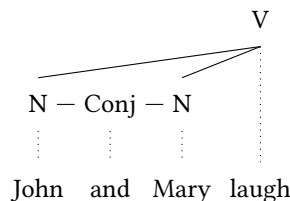


Figure 11.18: Analysis of coordination using the special relation *junction*

connected to the head *laugh*.

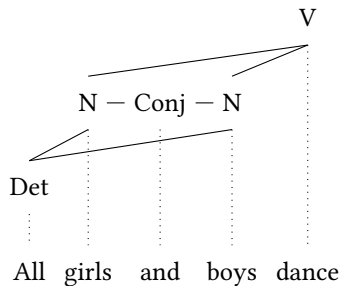


Figure 11.19: Analysis of coordination using the special relation *junction*

An alternative to such a special treatment of coordination would be to treat the conjunction as the head and the conjuncts as its dependents (I did not use Tesnière’s category labels here to spare the reader the work of translating I to V and N to O). The only problem of such a proposal would be the category of the conjunction. It cannot be Conj since the governing verb does not select a Conj but an N. The trick that could be applied here is basically the same trick as in Categorical Grammar (see Section 23.6.2): The category of the conjunction in Categorical Grammar is $(X\backslash X)/X$. We have a functor that takes two arguments of the same category and the result of the combination is an object that has the same category as the two arguments. Translating this approach to Dependency Grammar one would get an analysis as the one depicted in Figure 11.20 rather than the ones in Figure 11.18 and Figure 11.19. The figure for *all girls and boys*

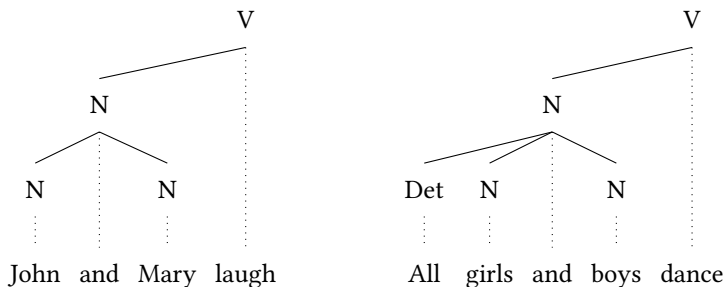


Figure 11.20: Analysis of coordination without *junction* and the conjunction as head

looks rather strange since both the determiner and the two conjuncts depend on the conjunction, but since the two Ns are selecting a Det, the same is true for the result of the coordination. In Categorical Grammar notation the category of the conjunction would be $((NP\backslash Det)(NP\backslash Det))/(NP\backslash Det)$ since X is instantiated by the nouns which would have the category $(NP\backslash Det)$ in an analysis in which the noun is the head and the determiner

is the dependent.

Note that both approaches have to come up with an explanation of subject–verb agreement. Tesnière’s original analysis assumes two dependencies between the verb and the individual conjuncts.¹⁶ Since the conjuncts are singular and the verb is plural, agreement cannot be modeled in tandem with dependency relations in this approach. If the second analysis finds ways of specifying the agreement properties of the coordination in the conjunction, the agreement facts can be accounted for without problems.

The alternative to a headed approach that is depicted in Figure 11.20 is an unheaded one. Several authors working in phrase structure-based frameworks suggested analyses of coordination without a head and such analyses are also assumed in Dependency Grammar (Hudson 1988; Kahane 1997). Hudson (1988) and others who make similar assumptions assume a phrase structure component for coordination, that is, the two nouns and the conjunction are combined to form a larger object that has properties that do not correspond to a single word among the combined words.

Similarly, the junction-based analysis of coordination poses problems for the interpretation of the representations. If semantic role assignment happens in parallel to dependency relations, there would be a problem with graphs like the one in Figure 11.18, since the semantic role of *laugh* cannot be filled by *John* and *Mary* simultaneously. Rather it is filled by one entity, namely the one that refers to the set that contains John and Mary. This semantic representation would belong to the phrase *John and Mary* and the natural candidate for being the top-most entity in this coordination is the *and*, since it embeds the meaning of *John* and the meaning of *Mary*: *and*'(*John*', *Mary*').

Such junctions are also assumed for the coordination of verbs. This however is not without problems, since adjuncts can have scope over the conjunct that is closest to them or over the whole coordination. An example is the following sentence from Levine (2003: 217):

- (12) Robin came in, found a chair, sat down, and whipped off her logging boots in exactly thirty seconds flat.

The adjunct *in exactly thirty seconds flat* can refer either to *whipped off her logging boots* as in (13a) or scope over all three conjuncts together as in (13b):

- (13) a. Robin came in, found a chair, sat down, and [[pulled off her logging boots] in exactly thirty seconds flat].
 b. Robin [[came in, found a chair, sat down, and pulled off her logging boots] in exactly thirty seconds flat].
 c. Robin came in in exactly thirty seconds flat and Robin found a chair in exactly thirty seconds flat and Robin pulled off her logging boots in exactly thirty seconds flat.

¹⁶ Eroms (2000: 467) notes the agreement problem and describes the facts. In his analysis he connects the first conjunct to the governing head, although it seems to be more appropriate to assume an internally structured coordination structure and then connect the highest conjunction.

However, there is no reading in which the adjunct scopes over each event individually. In order to express such a meaning one would have to utter (13c). The Tesnièrean analysis in Figure 11.21 corresponds to (13c), while an analysis that treats the conjunction as the head as in Figure 11.22 corresponds to (13b).

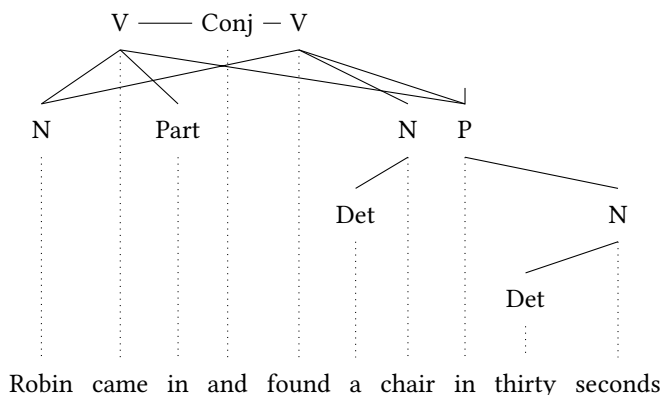


Figure 11.21: Analysis of verb coordination involving the junction relation

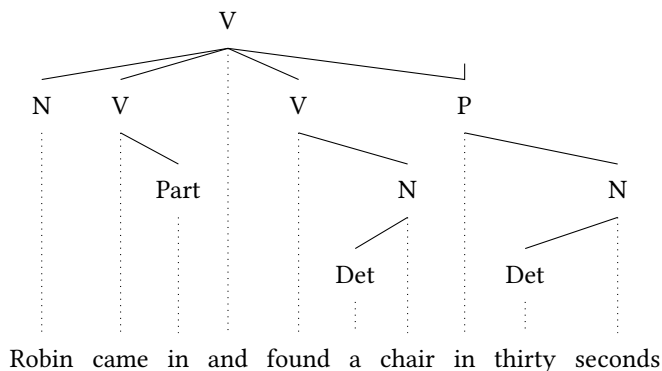


Figure 11.22: Analysis of verb coordination involving the connexion relation

Levine (2003: 217) discusses these sentences in connection to the HPSG analysis of extraction by Bouma, Malouf & Sag (2001a). Bouma, Malouf, and Sag suggest an analysis in which adjuncts are introduced lexically as dependents of a certain head. Since adjuncts are introduced lexically, the coordination structures basically have the same structure as the ones that are assumed in a Tesnièrean analysis. It may be possible to come up with a way to get the semantic composition right even though the syntax does not correspond to the semantic dependencies (see Chaves (2009) for suggestions), but it is clear that it

is simpler to derive the semantics from a syntactic structure that corresponds to what is going on in semantics.

11.6.2.2 Transfer

Transfers are used in Tesnière's system for the combination of words or phrases with a head of one of the major categories (for instance nouns) with words in minor categories (for instance prepositions). In addition transfers can transfer a word or phrase into another category without any other word participating.

Figure 11.23 shows an example of a transfer. The preposition *in* causes a category

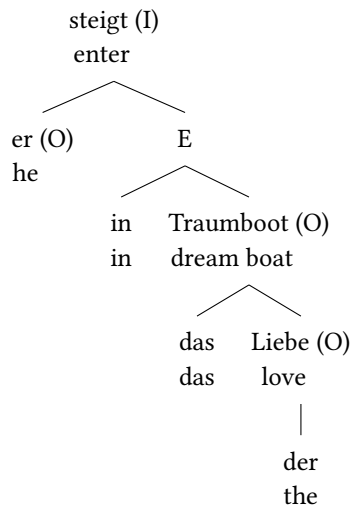


Figure 11.23: Transfer with an example adapted from Weber (1997: 83)

change: While *Traumboot* is an O (noun), the combination of the preposition and the noun is an E. The example shows that Tesnière used the grammatical category to encode grammatical functions. In theories like HPSG there is a clear distinction: There is part of speech information on the one hand and the function of elements as modifiers and predicates on the other hand. The modifier function is encoded by the selectional feature *MOD*, which is independent of the part of speech. Hence it is possible to have modifying and non-modifying adjectives, modifying and non-modifying prepositional phrases, modifying and non-modifying noun phrases and so on. For the example at hand one would assume a preposition with directional semantics that selects for an NP. The preposition is the head of a PP with a filled *MOD* value.

Another area in which transfer is used is morphology. For instance, the derivation of *frappant* by suffixation of *-ant* to the verb stem *frapp* is shown in Figure 11.24 on the following page. Such transfers can be subsumed under the general connexion relation,

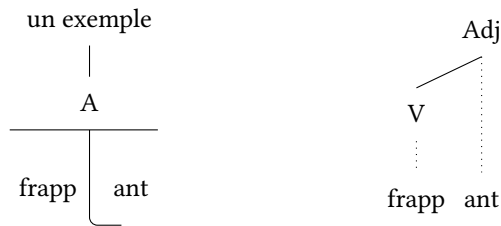


Figure 11.24: Transfer in morphology and its reconceptualization as normal dependency

if the affix is treated as the head. Morphologists working in realizational morphology and construction morphology argue against such morpheme-based analyses since they involve a lot of empty elements for conversions as for instance the conversion of the verb *play* into the noun *play* (see Figure 11.25). Hence, in theories like HPSG lexical rules are



Figure 11.25: Conversion as transfer from I (verb) to O (substantive) and as dependency with an empty element of the category N as head

assumed for derivations and conversions. The HPSG lexical rules are basically equivalent to unary branching rules (see the discussion of (37) on page 281). The affixes are integrated into the lexical rules or into realization functions that specify the morphological form of the item that is licensed by the lexical rule.

Concluding it can be said that transfers correspond to

- binary-branching phrase structure rules, if a word or phrase is combined with another word,
- unary phrase structure rules or binary branching phrase structure rules together with an empty head, if a phrase is converted to another category without any additional element present.
- a (unary) lexical rule, if a word or stem is mapped to a word or a stem.

For further discussion of the relation between Tesnière's transfer rules and constituency rules see Osborne & Kahane (2015: Section 4.9.1–4.9.2). Osborne & Kahane point out that transfer rules can be used to model exocentric constructions, that is, constructions

in which there is no single part that could be identified as the head. For more on headless constructions see Section 11.7.2.4.

11.6.3 Scope

As Osborne & Kahane (2015: lix) point out, Tesnière uses so-called polygraphs to represent scopal relations. So, since *that you saw yesterday* in (14) refers to *red cars* rather than *cars* alone, this is represented by a line that starts at the connection between *red* and *cars* rather than on one of the individual elements (Tesnière 2015: 150, Stemma 149).

(14) red cars that you saw yesterday

Tesnière's analysis is depicted in the left representation in Figure 11.26. It is interesting to see that this representation corresponds to the phrase structure tree at the right in Figure 11.26. The combination B between *red* and *cars* corresponds to the B node in the

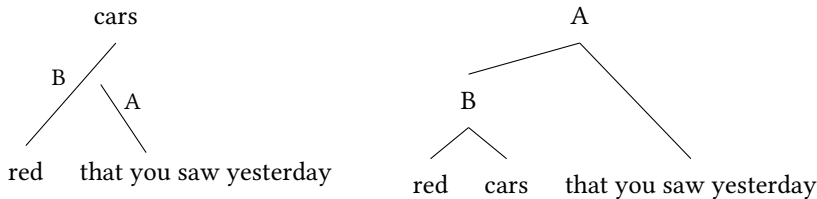


Figure 11.26: Tesnière's way to represent scope and the comparison with phrase structure-based analyses by Osborne & Kahane (2015: lix)

right-hand figure and the combination A of *red cars* and *that you saw yesterday* corresponds to A the A node. So, what is made explicit and is assigned a name in phrase structure grammars remains nameless in Tesnière's analysis, but due to the assumption of polygraphs, it is possible to refer to the combinations.

11.7 Summary and classification

Proponents of Dependency Grammar emphasize the point that Dependency Grammar is much simpler than phrase structure grammars, since there are fewer nodes and the general concept is more easy to grasp. This is indeed true: Dependency Grammar is well-suited for teaching grammar in introductory classes. However, as Sternefeld & Richter (2012: 285) pointed out in a rather general discussion, a simple syntax has the price of a complex semantics and vice versa. So, in addition to the dependency structure that is described in Dependency Syntax, one needs other levels. One level is the level of semantics and another one is linearization. As far as linearization is concerned, Dependency Grammar has two options: Assuming continuous constituents, that is, projective structures (Section 11.7.1), or allowing for discontinuous constituents (Section 11.7.2.2).

These options will be discussed in the following subsections. Section 11.7.2 compares dependency grammars with phrase structure grammars and shows that projective Dependency Grammars can be translated into phrase structure grammars. It also shows how HPSG can handle discontinuous constituents.

11.7.1 Linearization

We have seen several approaches to linearization in this chapter. Many just assume a dependency graph and some linearization according to the topological fields model. As was argued in Section 11.5 allowing discontinuous serialization of a head and its dependents opens Pandora's box. I discussed a suggestion by Kunze (1968), Hudson (1997; 2000), Kahane, Nasr & Rambow (1998), and Groß & Osborne (2009), who allow dependents of a head to rise to a dominating head for those cases in which a discontinuity would arise otherwise. However, there seems to be a reason to assume that fronting should be treated by special mechanisms even in cases that allow for continuous serialization. For instance, the ambiguity or lack of ambiguity of the examples in (15) cannot be explained in a straightforward way:

- (15) a. Oft liest er das Buch nicht.
 often reads he the book not
 'It is often that he does not read the book.' or 'It is not the case that he reads the book often.'
- b. dass er das Buch nicht oft liest
 that he the book not often reads
 'It is not the case that he reads the book often.'
- c. dass er das Buch oft nicht liest
 that he the book often not reads
 'It is often that he does not read the book.'

The point about the three examples is that only (15a) is ambiguous. Even though (15c) has the same order as far as *oft* 'often' and *nicht* 'not' are concerned, the sentence is not ambiguous. So it is the fronting of an adjunct that is the cause for the ambiguity. The dependency graph for (15a) is shown in Figure 11.27 on the next page. Of course the dependencies for (15b) and (15c) do not differ, so the graphs would be the same only differing in serialization. Therefore the scope differences could not be derived from the dependencies and complicated statements like (16) would be necessary:

- (16) If a dependent is linearized in the *Vorfeld* it can both scope over and under all other adjuncts of the head it is a dependent of.

Eroms (1985: 320) suggests an analysis of negation in which the negation is treated as the head, that is, the sentence in (17) has the structure in Figure 11.28 on the facing page.¹⁷ This analysis is equivalent to analyses in the Minimalist Program that assume a NegP

¹⁷ But see Eroms (2000: Section 11.2.3).

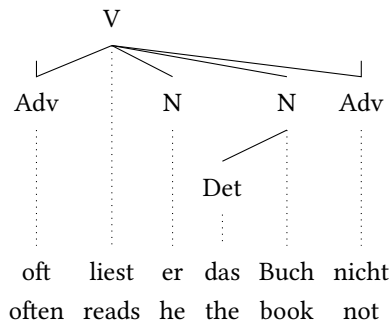


Figure 11.27: Dependency graph for *Oft liest er das Buch nicht*. 'He does not read the book often.'

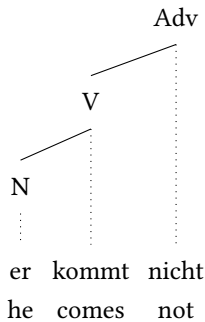


Figure 11.28: Analysis of negation according to Eroms (1985: 320)

and it has the same problem: The category of the whole object is Adv, but it should be V. This is a problem since higher predicates may select for a V rather than an Adv. See for instance the analysis of embedded sentences like (18) below.

The same is true for constituent negation or for other scope bearing elements. For instance the analysis of (17) would have to be the one in Figure 11.29 on the next page.

- (17) der angebliche Mörder
the alleged murderer

This structure would have the additional problem of being non-projective. Eroms does treat the determiner differently from what is assumed here, so this type of non-projectivity may not be a problem for him. However, the head analysis of negation would result in non-projectivity in so-called coherent constructions in German. The following sentence has two readings: In the first reading the negations scopes over *singen* 'sing' and in the second one over *singen darf* 'sing may'.

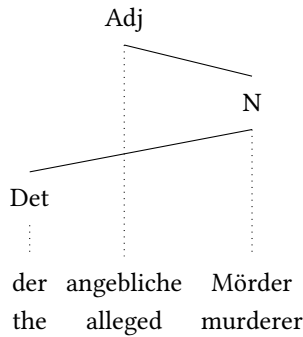


Figure 11.29: Analysis that would result if one considered all scope-bearing adjuncts to be heads

- (18) dass er nicht singen darf
 that he not sing may
 ‘that he may not sing’

The reading in which *nicht* scopes over the whole verbal complex would result in the non-projective structure that is given in Figure 11.30. Eroms also considers an analysis

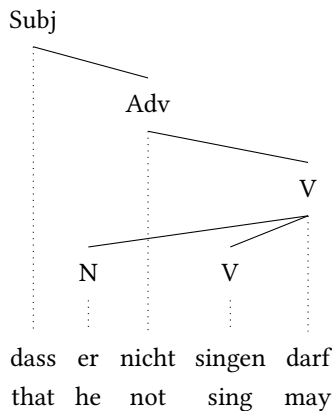


Figure 11.30: Analysis that results if one assumes the negation to be a head

in which the negation is a word part (‘Wortteiläquivalent’), but this does not help here since first the negation and the verb are not adjacent in V2 contexts like (15a) and even in verb final contexts like (18) Eroms would have to assume that the object to which the negation attaches is the whole verbal complex *singen darf*, that is, a complex object consisting of two words.

So, this leaves us with the analysis provided in Figure 11.27 and hence with a problem since we have one structure with two possible adjunct realizations that correspond to different readings, which is not predicted by an analysis that treats the two possible linearizations simply as alternative orderings.

Thomas Groß (p. c. 2013) suggested an analysis in which *oft* does not depend on the verb but on the negation. This corresponds to constituent negation in phrase structure approaches. The dependency graph is shown at the left-hand side in Figure 11.31. The

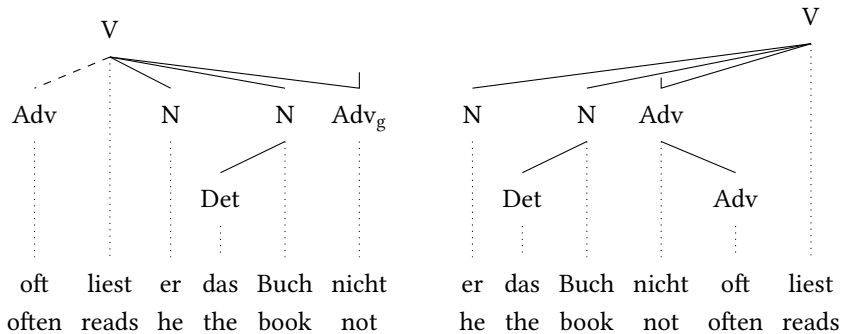


Figure 11.31: Dependency graph for *Oft liest er das Buch nicht.* 'He does not read the book often.' according to Groß and verb-final variant

figure at the right-hand side shows the graph for the corresponding verb-final sentence. The reading that corresponds to constituent negation can be illustrated with contrastive expressions. While in (19a) it is just the *oft* 'often' that is negated, it is *oft gelesen* 'often read' that is in the scope of negation in (19b).

- (19) a. Er hat das Buch nicht oft gelesen, sondern selten.
 he has the book not often read but seldom
 'He did not read the book often, but seldom.'
- b. Er hat das Buch nicht oft gelesen, sondern selten gekauft.
 he has the book not often read but seldom bought
 'He did not read the book often but rather bought it seldom.'

These two readings correspond to the two phrase structure trees in Figure 11.32 on the following page. Note that in an HPSG analysis the adverb *oft* would be the head of the phrase *nicht oft* 'not often'. This is different from the Dependency Grammar analysis suggested by Groß. Furthermore, the Dependency Grammar analysis has two structures: one flat one with all adverbs depending on the same verb and one in which *oft* depends on the negation. The phrase structure-based analysis has three structures: one with the order *oft* before *nicht*, one with the order *nicht* before *oft* and the one with direct combination of *nicht* and *oft*. The point about the example in (15a) is that one of the first two structures is missing in the Dependency Grammar representations. This probably

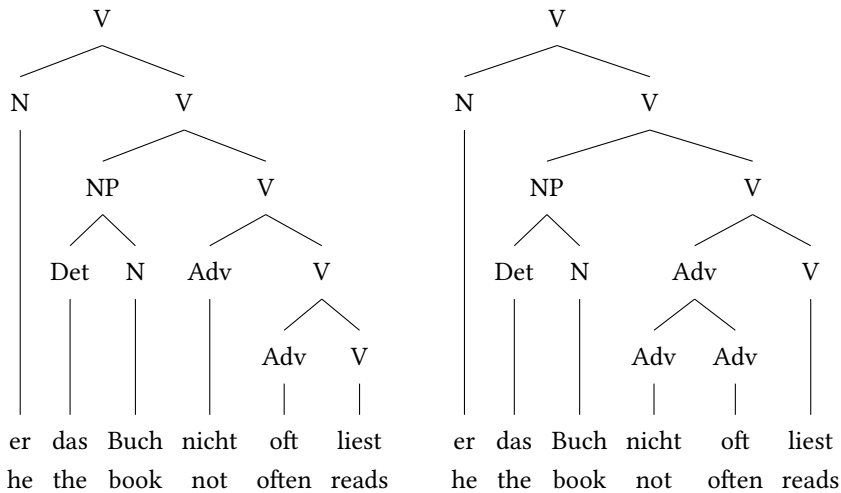


Figure 11.32: Possible syntactic analyses for *er das Buch nicht oft liest* 'He does not read the book often.'

does not make it impossible to derive the semantics, but it is more difficult than it is in constituent-based approaches.

Note also that models that directly relate dependency graphs to topological fields will not be able to account for sentences like (20).

- (20) Dem Saft eine kräftige Farbe geben Blutorangen.¹⁸
 the juice a strong color give blood.oranges
 'Blood oranges give the juice a strong color.'

The dependency graph of this sentence is given in Figure 11.33 on the next page.

Such apparent multiple frontings are not restricted to NPs. Various types of dependents can be placed in the *Vorfeld*. An extensive data discussion is provided in (Müller 2003a). Additional data has been collected in a research project on multiple frontings and information structure (Bildhauer 2011). Every theory that is based on dependencies alone and that does not allow for empty elements is forced to give up the restriction that is commonly assumed in the analysis of V2 languages, namely that the verb is in second position. In comparison, analyses like GB and those HPSG variants that assume an empty verbal head can assume that a projection of such a verbal head occupies the *Vorfeld*. This explains why the material in the *Vorfeld* behaves like verbal projections containing a visible verb: Such *Vorfelds* are internally structured topologically, they may have a filled *Nachfeld* and even a particle that fills the right sentence bracket. See Müller (2005c; 2015b) for further data, discussion, and a detailed analysis. The equivalent of the

¹⁸ Bildhauer & Cook (2010) found this example in the *Deutsches Referenzkorpus* (DeReKo), hosted at Institut für Deutsche Sprache, Mannheim: <http://www.ids-mannheim.de/kl/projekte/korpora>

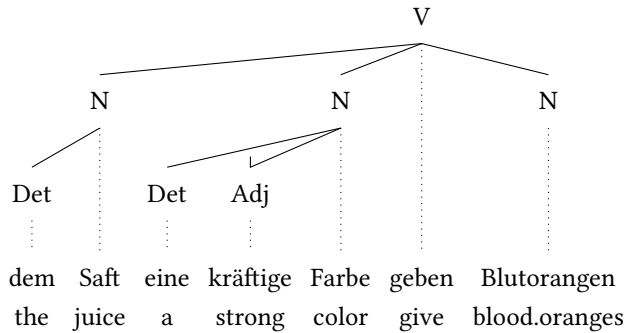


Figure 11.33: Dependency graph for *Dem Saft eine kräftige Farbe geben Blutorangen*. ‘Blood oranges give the juice a strong color.’

analysis in Gross & Osborne’s framework (2009) would be something like the graph that is shown in Figure 11.34, but note that Groß & Osborne (2009: 73) explicitly reject empty elements and in any case an empty element that is stipulated just to get the multiple fronting cases right would be entirely ad hoc.¹⁹ It is important to note that the issue is

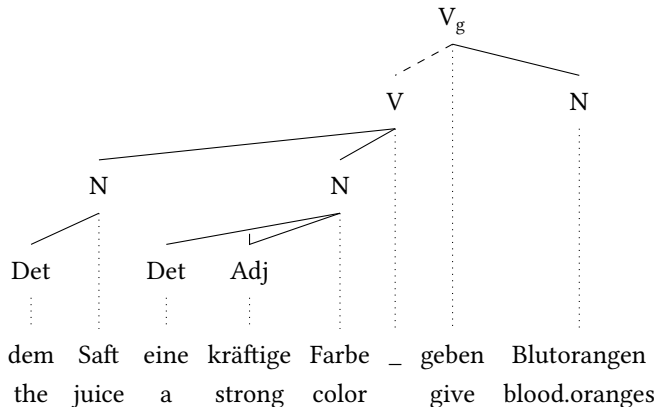


Figure 11.34: Dependency graph for *Dem Saft eine kräftige Farbe geben Blutorangen*. ‘Blood oranges give the juice a strong color.’ with an empty verbal head for the *Vorfeld*

not solved by simply dropping the V2 constraint and allowing dependents of the finite

¹⁹ I stipulated such an empty element in a linearization-based variant of HPSG allowing for discontinuous constituents (Müller 2002b), but later modified this analysis so that only continuous constituents are allowed and verb position is treated as head-movement and multiple frontings involve the same empty verbal head as is used in the verb movement analysis (Müller 2005c; 2015b).

verb to be realized to its left, since the fronted constituents do not necessarily depend on the finite verb as the examples in (21) show:

- (21) a. [Gezielt] [Mitglieder] [im Seniorenbereich] wollen die Kendoka
targeted members in.the senior.citizens.sector want.to the Kendoka
allerdings nicht werben.²⁰
however not recruit
'However, the Kendoka do not intend to target the senior citizens sector
with their member recruitment strategy.'
- b. [Kurz] [die Bestzeit] hatte der Berliner Andreas Klöden [...] gehalten.²¹
briefly the best.time had the Berliner Andreas Klöden [...] held
'Andreas Klöden from Berlin had briefly held the record time.'

And although such structures are marked such multiple frontings can even cross clause boundaries:

- (22) a. Der Maria einen Ring glaube ich nicht, daß er je schenken wird.²²
the Maria a ring believes I not that he ever give will
'I dont think that he would ever give Maria a ring.'
- b. (Ich glaube) Kindern Bonbons gibt man besser nicht.²³
I think children candy gives one better not
'I think it's better not to give candy to children.'

If such dependencies are permitted it is really difficult to constrain them. The details cannot be discussed here but the reader is referred to (Müller 2005c; 2015b).

Note also that Engel's statement regarding the linear order in German sentences (2014: 50) referring to one element in front of the finite verb (see footnote 6) is very imprecise. One can only guess what is intended by the word *element*. One interpretation is that it is a continuous constituent in the classical sense of constituency-based grammars. An alternative would be that there is a continuous realization of a head and some but not necessarily all of its dependents. This alternative would allow an analysis of extraposition with discontinuous constituents of (23) as it is depicted in Figure 11.35 on the next page.

- (23) Ein junger Kerl stand da, mit langen blonden Haaren, die sein Gesicht
a young guy stood there with long blond hair that his face
einrahmten, [...] ²⁴
framed
'A young guy was standing there with long blond hair that framed his face'

²⁰ taz, 07.07.1999, p. 18. Quoted from Müller (2002b).

²¹ Märkische Oderzeitung, 28./29.07.2001, p. 28

²² Fanselow (1993: 67)

²³ (G. Müller, 1998: 260)

²⁴ Charles Bukowski, *Der Mann mit der Ledertasche*. München: Deutscher Taschenbuch Verlag, 1994, p. 201, translation by Hans Hermann.

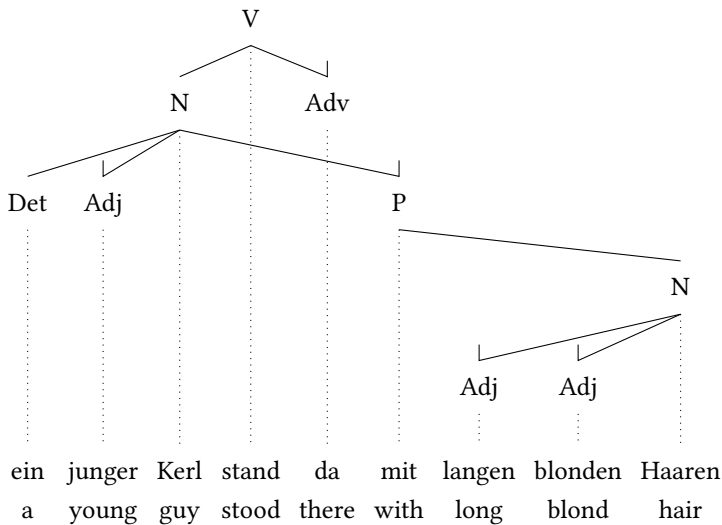


Figure 11.35: Dependency graph for *Ein junger Kerl stand da, mit langen blonden Haaren.* ‘A young guy was standing there with long blond hair.’ with a discontinuous constituent in the *Vorfeld*

A formalization of such an analysis is not trivial, since it has to be made precise what can be realized discontinuously and which parts of a dependency must be realized continuously. Kathol & Pollard (1995) developed such an analysis of extraposition in the framework of HPSG. See also Müller (1999a: Section 13.3). I discuss the basic mechanisms for such linearization analyses in HPSG in the following section.

11.7.2 Dependency Grammar vs. Phrase Structure Grammar

This section deals with the relation between Dependency Grammars and Phrase Structure Grammars. I first show that projective Dependency Grammars can be translated into Phrase Structure Grammars (Section 11.7.2.1) and then deal with non-projective DGs and show how they can be captured in linearization-based HPSG (Section 11.7.2.2). Section 11.7.2.3 argues for the additional nodes that are assumed in phrase structure-based theories and Section 11.7.2.4 discusses headless constructions, which pose a problem for all Dependency Grammar accounts.

11.7.2.1 Translating projective Dependency Grammars into Phrase Structure Grammars

As noted by Gaifman (1965), Covington (1990: 234), Oliva (2003) and Hellwig (2006: 1093), certain projective headed phrase structure grammars can be turned into Dependency

Grammars by moving the head one level up to replace the dominating node. So in an NP structure, the N is shifted into the position of the NP and all other connections remain the same. Figure 11.36 illustrates. Of course this procedure cannot be applied to all phrase



Figure 11.36: Analysis of *a book* in a phrase structure and a Dependency Grammar analysis

structure grammars directly, since some involve more elaborate structure. For instance, the rule $S \rightarrow NP, VP$ cannot be translated into a dependency rule, since NP and VP are both complex categories.

In what follows, I want to show how the dependency graph in Figure 11.1 on page 353 can be recast in headed phrase structure rules that license a similar tree, namely the one in Figure 11.37. I did not use the labels NP and VP to keep the two figures maximally

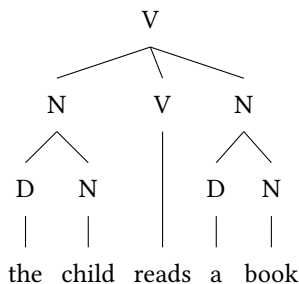


Figure 11.37: Analysis of *The child reads a book*. in a phrase structure with flat rules

similar. The P part of NP and VP refers to the saturation of a projection and is often ignored in figures. See for instance Chapter 9 on HPSG. The grammar that licenses the tree is given in (24), again ignoring valence information.

- (24) $N \rightarrow DN$ $N \rightarrow \text{child}$ $D \rightarrow \text{the}$ $D \rightarrow a$
 $V \rightarrow NVN$ $N \rightarrow \text{book}$ $V \rightarrow \text{reads}$

If one replaces the N and V in the right-hand site of the two left-most rules in (24) by the respective lexical items and removes the rules that license the words, one gets the following lexicalized variant of the grammar:

- (25) $N \rightarrow D \text{ book}$ $D \rightarrow \text{the}$
 $N \rightarrow D \text{ child}$ $D \rightarrow a$
 $V \rightarrow N \text{ reads } N$

Lexicalized means that every partial tree licensed by a grammar rule contains a lexical element. The grammar in (25) licenses exactly the tree in Figure 11.1.²⁵

One important difference between classical phrase structure grammars and Dependency Grammars is that the phrase structure rules impose a certain order on the daughters. That is the V rule in (25) implies that the first nominal projection, the verb, and the second nominal projection have to appear in the order stated in the rule. Of course this ordering constraint can be relaxed as it is done in GPSG. This would basically permit any order of the daughters at the right hand side of rules. This leaves us with the integration of adjuncts. Since adjuncts depend on the head as well (see Figure 11.4 on page 355), a rule could be assumed that allows arbitrarily many adjuncts in addition to the arguments. So the V rule in (25) would be changed to the one in (26):

- (26) $V \rightarrow N \text{ reads } N \text{ Adv}^*$

See page 188 for a similar rule in GPSG and see Kasper (1994) for an HPSG analysis of German that assumes totally flat structures and integrates an arbitrary number of adjuncts.

Such generalized phrase structures would give us the equivalent of projective Dependency Grammars.²⁶ However, as we have seen, some researchers allow for crossing edges, that is, for discontinuous constituents. In what follows I show how such Dependency Grammars can be formalized in HPSG.

11.7.2.2 Non-projective Dependency Grammars and phrase structure grammars with discontinuous constituents

The equivalent to non-projective dependency graphs are discontinuous constituents in phrase structure grammar. The technique that is used to model discontinuous constituents in frameworks like HPSG goes back to Mike Reape's work on German (1991; 1992; 1994). Reape uses a list called *DOMAIN* in order to represent the daughters of a sign

²⁵ Baumgärtner (1970: 57) and Heringer (1996: 37) suggest a general rule format for dependency rules that has a special marker ('*' and '~', respectively) in place of the lexical words in (25). Heringer's rules have the form in (26):

(i) $X[Y1, Y2, \sim, Y3]$

X is the category of the head, Y1, Y2, and Y3 are dependents of the head and '~' is the position into which the head is inserted.

²⁶ Sylvain Kahane (p. c. 2015) states that binarity is important for Dependency Grammars, since there is one rule for the subject, one for the object and so on (as for instance in Kahane 2009, which is an implementation of Dependency Grammar in the HPSG formalism). However, I do not see any reason not to allow for flat structures. For instance, Ginzburg & Sag (2000: 364) assumed a flat rule for subject auxiliary inversion in HPSG. In such a flat rule the specifier/subject and the other complements are combined with the verb in one go. This would also work for more than two valence features that correspond to grammatical functions like subject, direct object, indirect object. See also Footnote 25 on flat rules.

in the order in which they appear at the surface of an utterance. (27) shows an example in which the **DOM** value of a headed-phrase is computed from the **DOM** value of the head and the list of non-head daughters.

$$(27) \text{ headed-phrase} \Rightarrow \left[\begin{array}{l|l} \text{HEAD-DTR} & \text{DOM } \boxed{1} \\ \text{NON-HEAD-DTRS} & \boxed{2} \\ \text{DOM} & \boxed{1} \bigcirc \boxed{2} \end{array} \right]$$

The symbol ‘ \bigcirc ’ stands for the *shuffle* relation. *shuffle* relates three lists A, B and C, iff C contains all elements from A and B and the order of the elements in A and the order of the elements of B is preserved in C. (28) shows the combination of two sets with two elements each:

$$(28) \langle a, b \rangle \bigcirc \langle c, d \rangle = \langle a, b, c, d \rangle \vee \\ \langle a, c, b, d \rangle \vee \\ \langle a, c, d, b \rangle \vee \\ \langle c, a, b, d \rangle \vee \\ \langle c, a, d, b \rangle \vee \\ \langle c, d, a, b \rangle$$

The result is a disjunction of six lists. *a* is ordered before *b* and *c* before *d* in all of these lists, since this is also the case in the two lists $\langle a, b \rangle$ and $\langle c, d \rangle$ that have been combined. But apart from this *b* can be placed between or after *c* and *d*. Every word comes with a domain value that is a list the contains the word itself:

(29) Domain contribution of single words, here *gibt* ‘to give’:

$$\boxed{1} \left[\begin{array}{l} \text{PHON} \quad \langle \textit{gibt} \rangle \\ \text{SYNSEM} \quad \dots \\ \text{DOM} \quad \langle \boxed{1} \rangle \end{array} \right]$$

The description in (29) may seem strange at first glance, since it is cyclic, but it can be understood as a statement saying that *gibt* contributes itself to the items that occur in linearization domains.

The constraint in (30) is responsible for the determination of the **PHON** values of phrases:

$$(30) \text{ phrase} \Rightarrow \left[\begin{array}{l} \text{PHON} \quad \boxed{1} \oplus \dots \oplus \boxed{n} \\ \text{DOM} \quad \left\langle \left[\begin{array}{l} \text{PHON} \quad \boxed{1} \\ \text{sign} \end{array} \right], \dots, \left[\begin{array}{l} \text{PHON} \quad \boxed{n} \\ \text{sign} \end{array} \right] \right\rangle \end{array} \right]$$

It states that the **PHON** value of a sign is the concatenation of the **PHON** values of its **DOMAIN** elements. Since the order of the **DOMAIN** elements corresponds to their surface order, this is the obvious way to determine the **PHON** value of the whole linguistic object.

Figure 11.38 shows how this machinery can be used to license binary branching structures with discontinuous constituents. Words or word sequences that are separated by

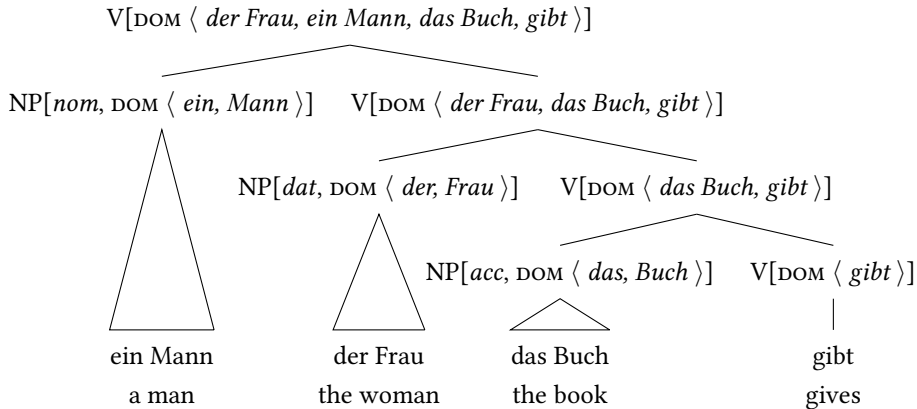


Figure 11.38: Analysis of *dass der Frau ein Mann das Buch gibt* ‘that a man gives the woman the book’ with binary branching structures and discontinuous constituents

commas stand for separate domain objects, that is, < das, Buch > contains the two objects *das* and *Buch* and < das Buch, gibt > contains the two objects *das Buch* and *gibt*. The important point to notice here is that the arguments are combined with the head in the order accusative, dative, nominative although the elements in the constituent order domain are realized in the order dative, nominative, accusative rather than nominative, dative, accusative as one would expect. This is possible since the formulation of the computation of the DOM value using the shuffle operator allows for discontinuous constituents. The node for *der Frau das Buch gibt* ‘the woman the book gives’ is discontinuous: *ein Mann* ‘a man’ is inserted into the domain between *der Frau* ‘the woman’ and *das Buch* ‘the book’. This is more obvious in Figure 11.39 on the next page, which has a serialization of NPs that corresponds to their order.

Such binary branching structures were assumed for the analysis of German by Kathol (1995; 2000) and Müller (1995; 1996c; 1999a; 2002a), but as we have seen throughout this chapter, Dependency Grammar assumes flat representations (but see Footnote 26 on page 383). Schema 6 licenses structures in which all arguments of a head are realized in one go.²⁷

²⁷ I assume here that all arguments are contained in the SUBCAT list of a lexical head but nothing hinges on that. One could also assume several valence features and nevertheless get a flat structure. For instance, Borsley (1989: 339) suggests a schema for auxiliary inversion in English and verb-initial sentences in Welsh that refers to both the valence feature for subjects and for complements and realizes all elements in a flat structure.

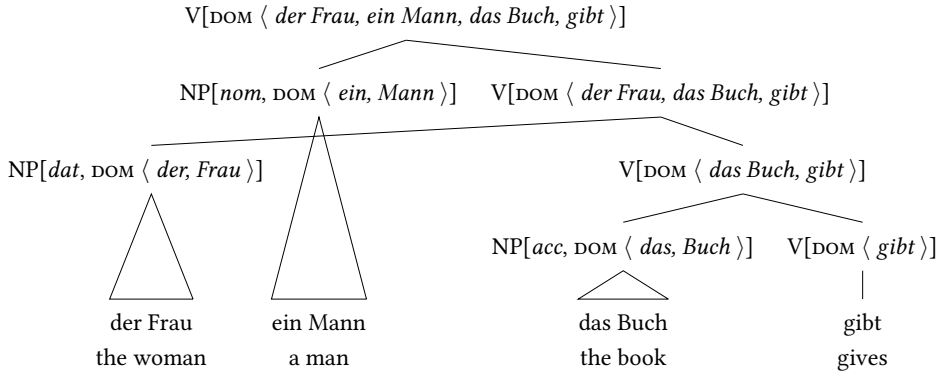


Figure 11.39: Analysis of *dass der Frau ein Mann das Buch gibt* ‘that a man gives the woman the book’ with binary branching structures and discontinuous constituents showing the discontinuity

Schema 6 (Head-Argument Schema (flat structure))

head-argument-phrase \Rightarrow

$$\left[\begin{array}{l} \text{SYNSEM|LOC|CAT|SUBCAT } \langle \rangle \\ \text{HEAD-DTR|SYNSEM|LOC|CAT|SUBCAT } \boxed{1} \\ \text{NON-HEAD-DTRS } \boxed{1} \end{array} \right]$$

To keep the presentation simple, I assume that the SUBCAT list contains descriptions of complete signs. Therefore the whole list can be identified with the list of non-head daughters.²⁸ The computation of the DOM value can be constrained in the following way:

$$(31) \text{ headed-phrase} \Rightarrow \left[\begin{array}{l} \text{HEAD-DTR} \quad \boxed{1} \\ \text{NON-HEAD-DTRS} \quad \langle \boxed{2}, \dots, \boxed{n} \rangle \\ \text{DOM} \quad \langle \boxed{1} \rangle \circ \langle \boxed{2} \rangle \circ \dots \circ \langle \boxed{n} \rangle \end{array} \right]$$

This constraint says that the value of DOM is a list that is the result of shuffling singleton lists that contain all the daughters as elements. The result of such a shuffle operation is a disjunction of all possible permutations of the daughters. This seems to be overkill for something that GPSG already gained by abstracting away from the order of the elements on the right hand side of a phrase structure rule, but note that this machinery can be used to reach even freer orders: By referring to the DOM values of the daughters rather than the daughters themselves we can insert individual words into the DOM list.

$$(32) \text{ headed-phrase} \Rightarrow \left[\begin{array}{l} \text{HEAD-DTR|DOM} \quad \boxed{1} \\ \text{NON-HEAD-DTRS} \quad \langle [\text{DOM } \boxed{2}] \dots [\text{DOM } \boxed{n}] \rangle \\ \text{DOM} \quad \langle \boxed{1} \rangle \circ \langle \boxed{2} \rangle \circ \dots \circ \langle \boxed{n} \rangle \end{array} \right]$$

²⁸ Without this assumption one would need a relational constraint that maps a list with descriptions of type *synsem* onto a list with descriptions of type *sign*.

Using this constraint we get DOM values that basically contain all the words in an utterance in any permutation. What we have is a pure Dependency Grammar without any constraints on projectivity. With such a grammar we could analyze the non-projecting structure of Figure 11.6 on page 357 and much more. The analysis in terms of domain union is shown in Figure 11.40. It is clear that such discontinuity is unwanted and hence

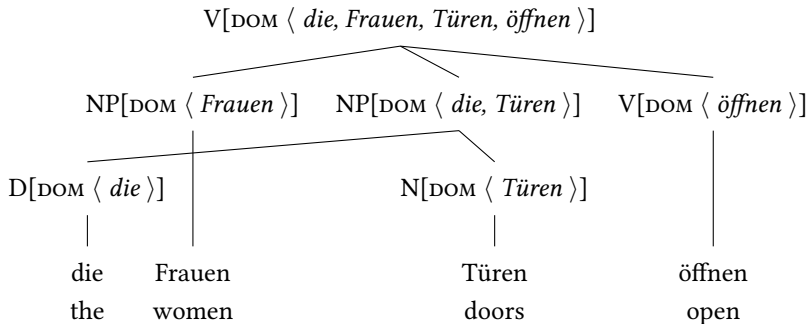


Figure 11.40: Unwanted analysis of *dass die Frauen Türen öffnen* ‘that the women open doors’ using Reape-style constituent order domains

one has to have restrictions that enforce continuity. One possible restriction is to request projectivity and hence equivalence to phrase structure grammars in the sense that was discussed above.

There is some dispute going on about the question whether constituency/dependency is primary/necessary to analyze natural language: While Hudson (1980), and Engel (1996) claim that dependency is sufficient, a claim that is shared by dependency grammarians (according to Engel (1996)), Leiss (2003) claims that it is not. In order to settle the issue, let us have a look at some examples:

- (33) Dass Peter kommt, klärt nicht, ob Klaus spielt.
 that Peter comes resolves not whether Klaus plays
 ‘That Peter comes does not resolve the question whether Klaus plays.’

If we know the meaning of the utterance, we can assign a dependency graph to it. Let us assume that the meaning is something like (34):

- (34) $\neg \text{resolve}'(\text{that}'(\text{come}'(\text{Peter}')), \text{whether}'(\text{play}'(\text{Klaus}'))))$

With this semantic information we can of course construct a dependency graph for (33). The reason is that the dependency relation is reflected one by one in the semantic representation in (34). The respective graph is given in Figure 11.41 on the following page. But note that this does not hold in the general case. Take for instance the example in (35):

- (35) Dass Peter kommt, klärt nicht, ob Klaus kommt.
 that Peter comes resolves not whether Klaus plays
 ‘That Peter comes does not resolve the question whether Klaus comes.’

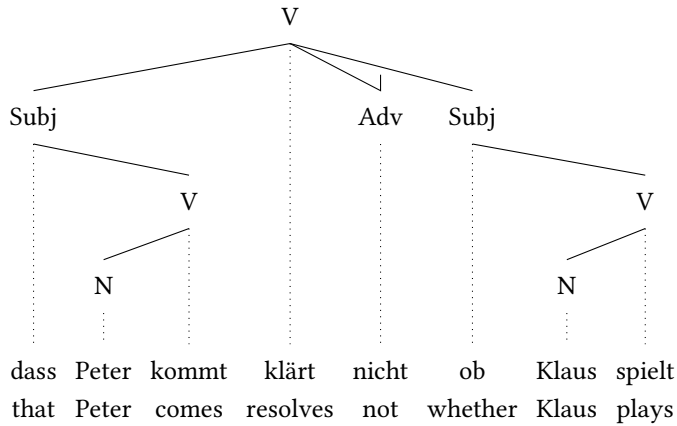


Figure 11.41: The dependency graph of *Dass Peter kommt, klärt nicht, ob Klaus spielt*. ‘That Peter comes does not resolve the question whether Klaus plays.’ can be derived from the semantic representation.

Here the words *dass* and *kommt* appear two times. Without any notion of constituency or restrictions regarding adjacency, linear order and continuity, we cannot assign a dependency graph unambiguously. For instance, the graph in Figure 11.42 is perfectly compatible with the meaning that this sentence has: *dass* dominates *kommt* and *kommt* dominates *Peter*, while *ob* dominates *kommt* and *kommt* dominates *Klaus*. I used the wrong

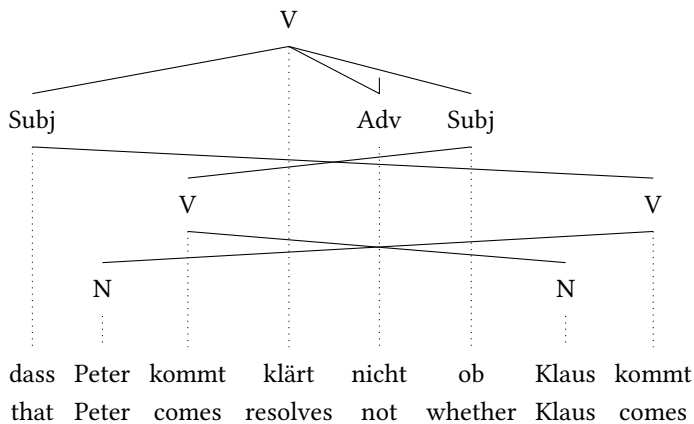


Figure 11.42: The dependency graph of *Dass Peter kommt, klärt nicht, ob Klaus kommt*. ‘That Peter comes does not resolve the question whether Klaus comes.’ is not unambiguously determined by semantics.

kommt in the dependency chains, but this is an issue of linearization and independent of dependency. If one takes linearization information into account, the dependency graph in Figure 11.42 is ruled out since *ob* ‘whether’ does not precede its verbal dependent *kommt* ‘comes’. But this explanation does not work for the example in Figure 11.6 on page 357. Here, all dependents are linearized correctly, it is just the discontinuity of *die* and *Türen* that is inappropriate. If it is required that *die* and *Türen* are continuous, we basically have constituents back in.

Similarly, nonprojective analyses without any constraints regarding continuity would permit the word salad in (36b):

- (36) a. Deshalb klärt, dass Peter kommt, ob Klaus spielt.
 therefore resolves that Peter comes whether Klaus plays
 b. Deshalb klärt dass ob Peter Klaus kommt spielt.
 therefore resolves that whether Peter Klaus comes plays

(36b) is a variant of (36a) in which the elements of the two clausal arguments are in correct order with respect to each other, but both clauses are discontinuous in such a way that the elements of each clause alternate. The dependency graph is shown in Figure 11.43. As was explained in Section 10.6.4.4 on the analysis of nonlocal dependencies

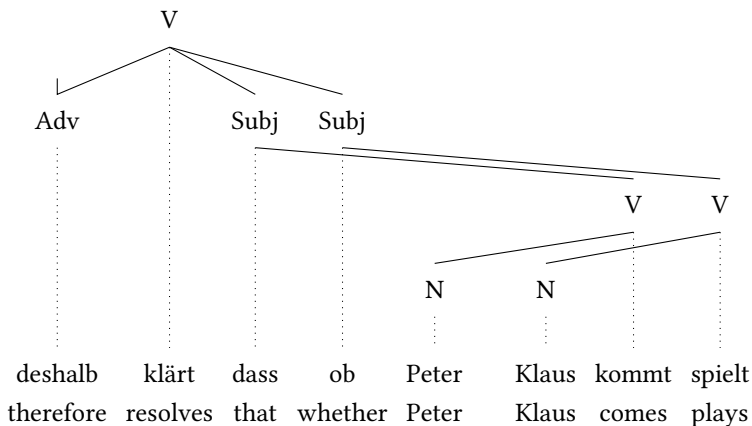


Figure 11.43: The dependency graph of the word salad *Deshalb klärt dass ob Peter Klaus kommt spielt*. ‘Therefore resolves that whether Peter Klaus comes plays’ which is admitted by non-projective dependency grammars that do not restrict discontinuity

in Fluid Construction Grammar, a grammar of languages like English and German has to constrain the clauses in such a way that they are continuous with the exception of extractions to the left. A similar statement can be found in Hudson (1980: 192). Hudson also states that an item can be fronted in English provided all of its dependents are

fronted with it (p.184). This “item with all its dependents” is the constituent in constituent-based grammars. The difference is that this object is not given an explicit name and is not assumed to be a separate entity containing the head and its dependents in most Dependency Grammars. See however Hellwig (2003) for an explicit proposal that assumes that there is a linguistic object that represents the whole constituent rather than just the lexical head.

Summing up what has been covered in this section so far, I have shown what a phrase structure grammar that corresponds to a certain dependency grammar looks like. I also showed how discontinuous constituents can be allowed for. However, there are issues that remained unaddressed so far: Not all properties that a certain phrase has are identical to its lexical head and the differences have to be represented somewhere. I will discuss this in the following subsection.

11.7.2.3 Features that are not identical between heads and projections

As Oliva (2003) pointed out the equivalence of Dependency Grammar and HPSG only holds up as far as HEAD values are concerned. That is, the node labels in dependency graphs correspond to the HEAD values in an HPSG. There are, however, additional features like CONT for the semantics and SLASH for nonlocal dependencies. These values usually differ between a lexical head and its phrasal projections. For illustration let us have a look at the phrase *a book*. The semantics of the lexical material and the complete phrase is given in (37) (for lambda expressions see Section 2.3):

- (37) a. $a: \lambda P \lambda Q \exists x (P(x) \wedge Q(x))$
 b. $book: \lambda y (book'(y))$
 c. $a\ book: \lambda Q \exists x (book'(x) \wedge Q(x))$

Now, the problem for the Dependency Grammar notation is that there is no NP node that could be associated with the semantics of *a book* (see Figure 11.36 on page 382), the only thing present in the tree is a node for the lexical N, that is, the node for *book*.²⁹ This is not a big problem, however: The lexical properties can be represented as part of the highest node as the value of a separate feature. The N node in a dependency graph would then have a CONT value that corresponds to the semantic contribution of the complete phrase and a LEX-CONT value that corresponds to the contribution of the lexical head of the phrase. So for *a book* we would get the following representation:

- (38)
$$\left[\begin{array}{ll} \text{CONT} & \lambda Q \exists x (book'(x) \wedge Q(x)) \\ \text{LEXICAL-CONT} & \lambda y (book'(y)) \end{array} \right]$$

²⁹ Hudson (2003: 391–392) is explicit about this: “In dependency analysis, the dependents modify the head word’s meaning, so the latter carries the meaning of the whole phrase. For example, in *long books about linguistics*, the word *books* means ‘long books about linguistics’ thanks to the modifying effect of the dependents.”

An alternative is to assume different representational levels as in Meaning Text Theory (?). In fact the CONT value in HPSG is also a different representational level. However, this representational level is in sync with the other structure that is build.

With this kind of representation one could maintain analyses in which the semantic contribution of a head together with its dependents is a function of the semantic contribution of the parts.

Now, there are probably further features in which lexical heads differ from their projections. One such feature would be SLASH, which is used for nonlocal dependencies in HPSG and could be used to establish the relation between risen element and the head in an approach à la Groß & Osborne (2009). Of course we can apply the same trick again. We would then have a feature LEXICAL-SLASH. But this could be improved and the features of the lexical item could be grouped under one path. The general skeleton would then be (39):

$$(39) \left[\begin{array}{c} \text{CONT} \\ \text{SLASH} \\ \text{LEXICAL} \left[\begin{array}{c} \text{CONT} \\ \text{SLASH} \end{array} \right] \end{array} \right]$$

But if we rename LEXICAL to HEAD-DTR, we basically get the HPSG representation.

Hellwig (2003: 602) states that his special version of Dependency Grammar, which he calls Dependency Unification Grammar, assumes that governing heads select complete nodes with all their daughters. These nodes may differ in their properties from the head (p. 604). They are in fact constituents. So this very explicit and formalized variant of Dependency Grammar is really close to HPSG, as Hellwig states himself (p. 603).

So, the conclusion of this section is that the only principled difference between phrase structure grammars and Dependency Grammar is the question of how much intermediate structure is assumed: Is there a VP without the subject? Are there intermediate nodes for adjunct attachment? It is difficult to decide these questions in the absence of fully worked out proposals that include semantic representations.

11.7.2.4 Non-headed constructions

Hudson (1980: Section 4.E) discusses headless constructions like those in (40):

- (40) a. the rich
b. the biggest
c. the longer the stem
d. (with) his hat over his eyes

He argues that the terms *adjective* and *noun* should be accompanied by the term *substantive*, which subsumes both terms. Then he suggests that *if a rule needs to cover the constructions traditionally referred to as noun-phrases, with or without heads, it just refers to 'nouns', and this will automatically allow the constructions to have either substantives or adjectives as heads.* (p. 195) The question that has to be asked here however is what the internal dependency structure of substantive phrases like *the rich* would be. The only way to connect the items seems to be to assume that the determiner is dependent on the

adjective. But this would allow for two structures of phrases like *the rich man*: one in which the determiner depends on the adjective and one in which it depends on the noun. So underspecification of part of speech does not seem to solve the problem. Of course all problems with non-headed constructions can be solved by assuming empty elements (see Section 2.4.1 for the assumption of an empty head in a phrase structure grammar for noun phrases). This has been done in HPSG in the analysis of relative clauses (Pollard & Sag 1994: Chapter 5). English and German relative clauses consist of a phrase that contains a relative word and a sentence in which the relative phrase is missing. Pollard and Sag assumed an empty relativizer that selects for the relative phrase and the clause with a gap (Pollard & Sag 1994: 216–217). Similar analyses can be found in Dependency Grammar (Eroms 2000: 291).³⁰ Now, the alternative to empty elements are phrasal constructions (see Chapter 21 on empty elements in general and Subsection 23.10.3 on relative clauses in particular). Sag (1997) working on relative clauses in English suggested a phrasal analysis of relative clauses in which the relative phrase and the clause from which it is extracted forms a new phrase. A similar analysis was assumed by Müller (1996c) and is documented in Müller (1999a: Chapter 10). As was discussed in Section 8.6 it is not plausible to assume the relative pronoun or some other element in the relative phrase to be the head of the entire relative clause, neither it is plausible to assume the verb to be the head of the entire relative clause (pace Sag), since relative clauses modify \bar{N} s, something that projections of (finite) verbs usually do not do. So assuming an empty head or a phrasal schema seems to be the only option.

Chapter 23 is devoted to the discussion of the question whether certain phenomena should be analyzed as involving phrase structural configurations or whether lexical analyses are better suited in general or for modeling some phenomena. I argue there that all phenomena that interact with valency should be treated lexically. But there are other phenomena as well and Dependency Grammar is forced to assume lexical analyses for all linguistic phenomena. There always has to be some element on which others depend. It has been argued by Jackendoff (2008) that it does not make sense to assume that one of the elements in N-P-N constructions like those in (41) is the head.

- (41) a. day by day, paragraph by paragraph, country by country
b. dollar for dollar, student for student, point for point

³⁰ The Dependency Grammar representations usually have a *d*-element as the head of the relative clause. However, since the relative pronoun is also present in the clause and since the *d*- is not pronounced twice, assuming an additional *d*-head is basically assuming an empty head.

Another option is to assume that words may have multiple functions: So, a relative pronoun may be both a head and a dependent simultaneously (Tesnière 2015: Chapter 246, §8–11; Osborne & Kahane 2015: xlvii; Kahane 2009: 129–130). At least the analysis of Kahane is an instance of the Categorical Grammar analysis that was discussed in Section 8.6 and it suffers from the same problems: If the relative pronoun is a head that selects for a clause that is missing the relative pronoun, it is not easy to see how this analysis extends to cases of pied-piping like (i) in which the extracted element is a complete phrase containing the relative pronoun rather than just the pronoun itself.

(i) die Frau, von deren Schwester ich ein Bild gesehen habe
the woman of whose sister I a picture seen have
'the woman of whose sister I saw a picture'

- c. face to face, bumper to bumper
- d. term paper after term paper, picture after picture
- e. book upon book, argument upon argument

Of course there is a way to model all the phenomena that would be modeled by a phrasal construction in frameworks like GPSG, CxG, HPSG, or Simpler Syntax: an empty head. Figure 11.44 shows the analysis of *student after student*. The lexical item for the empty N

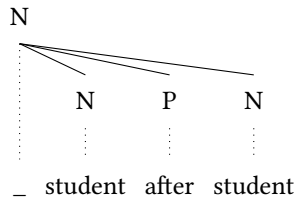


Figure 11.44: Dependency Grammar analysis of the N-P-N Construction with empty head

would be very special, since there are no similar non-empty lexical nouns, that is, there is no noun that selects for two bare Ns and a P.

One way out of these problems would of course to assume that there are special combinatorial mechanisms that assign a new category to one or several elements. This would basically be an unheaded phrase structure rule and this is what Tesnière suggested: transfer rules (see Section 11.6.2.2). But this is of course an extension of pure Dependency Grammar towards a mixed model.

Exercises

Provide the dependency graphs for the following three sentences:

- (42) a. Ich habe einen Mann getroffen, der blonde Haare hat.
 I have a man met who blond hair has
 ‘I met a man who has blond hair.’
- b. Einen Mann getroffen, der blonde Haare hat, habe ich noch nie.
 a man met who blond hair has have I yet never
 ‘I never met a man who has blond hair.’
- c. Dass er morgen kommen wird, freut uns.
 that he tomorrow come will pleases us
 ‘That he will come tomorrow pleases us.’

You may use non-projective dependencies. For the analysis of relative clauses authors usually assume an abstract entity that functions as a dependent of the modified noun and as a head of the verb in the relative clause.

Further reading

In the section on further reading in Chapter 3 I referred to the book called *Syntaktische Analyseperspektiven* ‘Syntactic perspectives on analyses’. The chapters in this book have been written by proponents of various theories and all analyze the same newspaper article. The book also contains a chapter by Engel (2014) assuming his version of Dependency Grammar, namely *Dependent Verb Grammar*.

Ágel, Eichinger, Eroms, Hellwig, Heringer & Lobin (2003; 2006) published a handbook on dependency and valency that discusses all aspects related to Dependency Grammar in any imaginable way. Many of the papers have been cited in this chapter. Especially relevant in the context of this book are those papers that compare Dependency Grammar with other theories: Lobin (2003) compares Dependency Grammar and Categorical Grammar, Oliva (2003) deals with the representation of valency and dependency in HPSG and Bangalore, Joshi & Rambow (2003) describe how valency and dependency is covered in TAG. Hellwig (2006) compares rule-based grammars with Dependency Grammars with special consideration of parsing by computer programs.

Osborne & Groß (2012) compare Dependency Grammar with Construction Grammar and Osborne, Putnam & Groß (2011) argue that certain variants of Minimalism are in fact reinventions of dependency-based analyses.

The original work on Dependency Grammar by Tesnière (1959) is also available in parts in German (Tesnière 1980) and in full in English (Tesnière 2015).

12 Tree Adjoining Grammar

Tree Adjoining Grammar (TAG) was developed by Aravind Joshi at the University of Pennsylvania in the USA (Joshi, Levy & Takahashi 1975). Several important dissertations in TAG have been supervised by Aravind Joshi and Anthony Kroch at the University of Pennsylvania (e.g. Rambow (1994)). Other research centers with a focus on TAG are Paris 7 (Anne Abeillé), Columbia University in the USA (Owen Rambow) and Düsseldorf, Germany (Laura Kallmeyer).

Rambow (1994) and Gerdes (2002b) are more detailed studies of German.¹

TAG and its variants with relevant extensions are of interest because it is assumed that this grammatical formalism can – with regard to its expressive power – relatively accurately represent what humans do when they produce or comprehend natural language. The expressive power of Generalized Phrase Structure Grammar was deliberately constrained so that it corresponds to context-free phrase structure grammars (Type-2 languages) and it could in fact be demonstrated that this is too little (Shieber 1985; Culy 1985).² Grammatical theories such as HPSG and CxG can generate/describe so-called type-0 languages and are thereby far above the level of complexity presently assumed for natural languages. The assumption is that this complexity lies somewhere between context-free and context-sensitive (Type 1) languages. This class is thus referred to as *mildly context sensitive*. Certain TAG-variants are outside of this language class and it is assumed that they can produce exactly those structures that occur in natural languages. For more on the discussion of complexity, see Section 12.6.3 and Chapter 18.

There exist various systems for the processing of TAG grammars (Doran, Hockey, Sarkar, Srinivas & Xia 2000; Parmentier, Kallmeyer, Maier, Lichte & Dellert 2008; Kallmeyer, Lichte, Maier, Parmentier, Dellert & Evang 2008). Smaller and larger TAG fragments have been developed for the following languages:

- Arabic (Fraj, Zribi & Ahmed 2008),
- German (Rambow 1994; Gerdes 2002a; Kallmeyer & Yoon 2004; Lichte 2007),
- English (XTAG Research Group 2001; Frank 2002; Kroch & Joshi 1987),
- French (Abeillé 1988; Candito 1996; 1998; 1999; Crabbé 2005),
- Italian (Candito 1998; 1999),

¹ Since my knowledge of French leaves something to be desired, I just refer to the literature in French here without being able to comment on the content.

² See Pullum (1986) for a historical overview of the complexity debate and G. Müller (2011) for argumentation for the non-context-free nature of German, which follows parallel to Culy with regard to the NPN construction (see Section 23.10.4).

- Korean (Han, Yoon, Kim & Palmer 2000; Kallmeyer & Yoon 2004),
- Vietnamese (Le, Nguyen & Roussanaly 2008)

Candito (1996) developed a system for the representation of meta grammars, which allows the uniform specification of crosslinguistic generalizations. This system was used by some of the projects mentioned above for the derivation of grammars for specific languages. For instance Kinyon, Rambow, Scheffler, Yoon & Joshi (2006) derive verb second languages from the meta grammar. Among those grammars for verb second languages is a grammar of Yiddish for which there was no TAG grammar until 2006.

Resnik (1992) combines TAG with a statistics component.

12.1 General remarks on representational format

12.1.1 Representation of valency information

Figure 12.1 shows so-called elementary trees. These are present in the lexicon and can be combined to create larger trees: Nodes for the insertion of arguments are specially

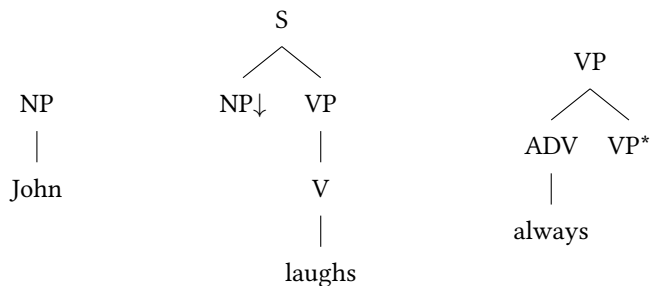


Figure 12.1: Elementary trees

marked (NP↓ in the tree for *laughs*). Nodes for the insertion of adjuncts into a tree are also marked (VP* in the tree for *always*). Grammars where elementary trees always contain at least one word are referred to as *Lexicalized Tree Adjoining Grammar* (LTAG).

12.1.2 Substitution

Figure 12.2 on the facing page shows the substitution of nodes: Other subtrees have to be inserted at substitution nodes such as the NP node in the tree for *laughs*. The tree for *John* is inserted there in the example derivation.

12.1.3 Adjunction

Figure 12.3 on the next page shows an example of how the adjunction tree for *always* can be used.

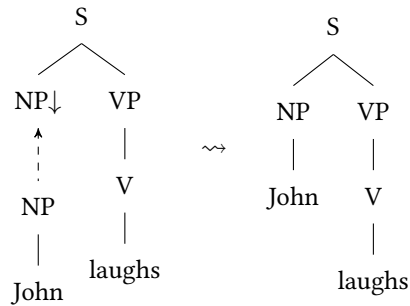


Figure 12.2: Substitution

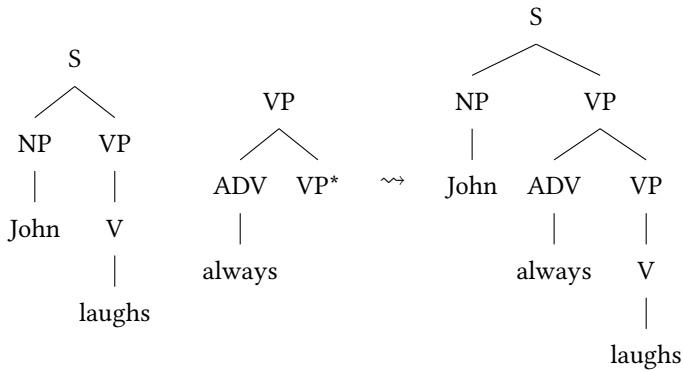


Figure 12.3: Adjunction

Adjunction trees can be inserted into other trees. During insertion, a node in a tree corresponding to the ‘*’ in the adjunction tree is split up and the adjunction tree is entered in its place.

TAG differs considerably from the simple phrase structure grammars we encountered in Chapter 2 in that the trees extend over a larger domain: For example, there is an NP node in the tree for *laughs* that is not a sister of the verb. In a phrase structure grammar (and of course in GB and GPSG since these theories are more or less directly built on phrase structure grammars) it is only ever possible to describe subtrees one level deep. For the tree for *laughs* the relevant rules would be those in (1):

- (1) $S \rightarrow NP VP$
 $VP \rightarrow V$
 $V \rightarrow laughs$

In this context, it is common to speak of *locality domains*. The extension of the locality domain is of particular importance for the analysis of idioms (see Chapter 19).

TAG differs from other grammatical theories in that it is possible for structures to be broken up again. In this way, it is possible to use adjunction to insert any amount of material into a given tree and thereby cause originally adjacent constituents to end up being arbitrarily far away from each other in the final tree. As we will see in Section 12.5, this property is important for the analysis of long-distance dependencies without movement.

12.1.4 Semantics

There are different approaches to the syntax-semantics interface in TAG. One possibility is to assign a semantic representation to every node in the tree. The alternative is to assign each elementary tree exactly one semantic representation. The semantics construction does not make reference to syntactic structure but rather the way the structure is combined. This kind of approach has been proposed by Candito & Kahane (1998) and then by Kallmeyer & Joshi (2003) who build on it. The basic mechanisms will be briefly presented in what follows.

In the literature on TAG, a distinction is made between derived trees and derivation trees. Derived trees correspond to constituent structure (the trees for *John laughs* and *John always laughs* in Figures 12.2 and 12.3). The derivation tree contains the derivational history, that is, information about how the elementary trees were combined. The elements in a derivation tree represent predicate-argument dependencies, which is why it is possible to derive a semantic derivation tree from them. This will be shown on the basis of the sentence in (2):

(2) Max likes Anouk.

The elementary tree for (2) and the derived tree are given in Figure 12.4. The nodes in

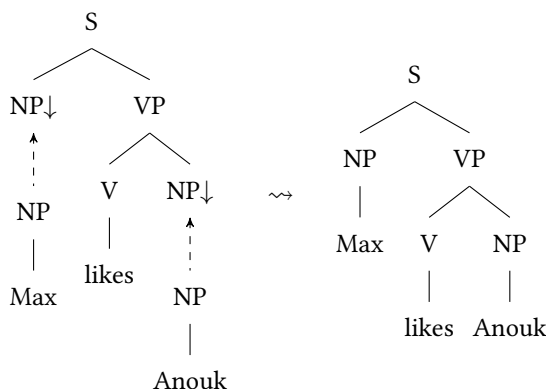


Figure 12.4: Elementary trees and derived tree for *Max likes Anouk*.

trees are numbered from top to bottom and from left to right. The result of this numbering of nodes for *likes* is shown in Figure 12.5 on the next page. The top-most node in

the tree for *likes* is S and has the position 0. Beneath S, there is an NP and a VP node. These nodes are again numbered starting at 0. NP has the position 0 and VP the position 1. The VP node has in turn two daughters: V and the object NP. V receives number 0 and the object NP 1. This makes it possible to combine these numbers and then it is possible to unambiguously access individual elements in the tree. The position for the subject NP is 00 since this is a daughter of S and occurs in first position. The object NP has the numeric sequence 011 since it is below S (0), in the VP (the second daughter of S = 1) and occurs in second position (the second daughter of VP = 1).

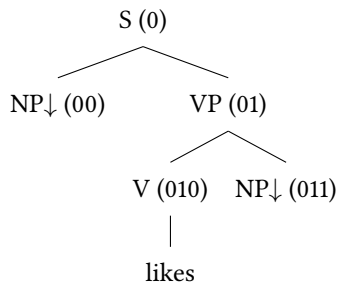


Figure 12.5: Node positions in the elementary tree for *likes*

With these tree positions, the derivation tree for (2) can be represented as in Figure 12.6. The derivation tree expresses the fact that the elementary tree for *likes* was

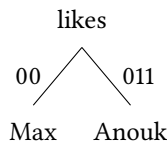


Figure 12.6: Derivation tree for *Max likes Anouk*.

combined with two arguments that were inserted into the substitution positions 00 and 011. The derivation tree also contains information as to what exactly was placed into these nodes.

Kallmeyer & Joshi (2003) use a variant of *Minimal Recursion Semantics* as their semantic representational formalism (Copestake, Flickinger, Pollard & Sag 2005). I will use a considerably simplified representation here, as I did in the HPSG chapter. For the elementary trees *Max*, *likes* and *Anouk*, we can assume the semantic representations in (3)

- (3) Semantic representations for elementary trees:

max(x)	like(x ₁ , x ₂)	anouk(y)
arg: —	arg: ⟨ x ₁ , 00 ⟩, ⟨ x ₂ , 011 ⟩	arg: —

In a substitution operation, a variable is assigned a value. If, for example, the elementary tree for *Max* is inserted into the subject position of the tree for *likes*, then x_1 is identified with x . In the same way, x_2 is identified with y if the tree for *Anouk* is inserted into the object position. The result of these combinations is the representation in (4):

(4) Combination of the meaning of elementary trees:

like(x, y)
max(x)
anouk(y)
arg: —

Kallmeyer & Joshi (2003) show how an extension of TAG, Multi-Component LTAG, can handle quantifier scope and discuss complex cases with embedded verbs. Interested readers are referred to the original article.

12.2 Local reordering

In TAG, there is a family of trees for each word. In order to account for ordering variants, one can assume that there are six trees corresponding to a ditransitive verb and that each of these corresponds to a different ordering of the arguments. Trees are connected to one another via lexical rules. This is parallel to the Categorical Grammar variant developed by Uszkoreit (1986b).

Alternatively, one could assume a format for TAG structures similar to what we referred to as the ID/LP format in the chapter on GPSG. Joshi (1987b) defines an elementary structure as a pair that consists of a dominance structure and linearization constraints. Unlike GPSG, the linearization rules do not hold for all dominance rules but rather for a particular dominance structure. This is parallel to what we saw in Section 10.6.3 on Embodied-CxG. Figure 12.7 shows a dominance tree with numbered nodes. If we com-

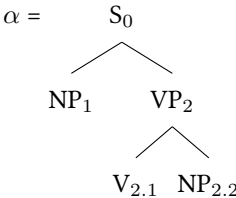


Figure 12.7: Dominance structure with numbered nodes

bine this dominance structure with linearization rules in (5), we arrive at the exact order that we would get with ordinary phrase structure rules, namely $NP_1 \text{ V } NP_2$.

$$(5) \quad LP_1^\alpha = \{ 1 < 2, 2.1 < 2.2 \}$$

If one specifies the linearization restrictions as in (6), all the orders in (7) are admitted, since the empty set means that we do not state any restrictions at all.

$$(6) \quad LP_2^\alpha = \{ \}$$

- (7) a. $NP_1 \text{ V } NP_2$
 b. $NP_2 \text{ V } NP_1$
 c. $NP_1 \text{ } NP_2 \text{ V}$
 d. $NP_2 \text{ } NP_1 \text{ V}$
 e. $\text{V } NP_1 \text{ } NP_2$
 f. $\text{V } NP_2 \text{ } NP_1$

This means that it is possible to derive all orders that were derived in GPSG with flat sentence rules despite the fact that there is a constituent in the tree that consists of NP and VP. Since the dominance rules include a larger locality domain, we do not call such grammars ID/LP grammars (immediate dominance/linear precedence) but rather LD/LP grammars (local dominance/linear precedence) (Joshi et al. 1990).

Simple variants of TAG such as those presented in Section 12.1 cannot deal with re-ordering if the arguments of different verbs alternate in their linear order as in (8).

- (8) weil ihm das Buch jemand zu lesen versprochen hat³
 because him the book somebody to read promised has
 ‘because somebody promised to read him the book.’

In (8), *das Buch* ‘the book’ is the object of *zu lesen* ‘to read’, and *ihm* ‘him’ and *jemand* ‘somebody’ are dependent on the verb. These cases can be analyzed by LD/LP-TAG developed by Joshi (1987b) and FO-TAG (Becker, Joshi & Rambow 1991: 21) since both of these TAG variants allow for crossing edges.

Since certain restrictions cannot be expressed in FO-TAG (Rambow 1994: 48–50), so-called Multi-Component TAG were developed. Joshi, Becker & Rambow (2000) illustrate the problem that simple LTAG grammars have with sentences such as (8) using examples such as (9):⁴

- (9) a. ... daß der Detektiv dem Klienten [den Verdächtigen des Verbrechens zu
 that the detective the client the suspect of the crime to
 überführen] versprach
 convict promised
 ‘that the detective promised the client to convict the suspected perpetrator
 of the crime.’

³ For more on these kind of examples, see Bech (1955).

⁴ The authors use *versprochen hat* ‘has promised’ rather than *versprach* ‘promised’, which sounds better but does not correspond to the trees they use.

- b. ... daß des Verbrechens_k der Detektiv den Verdächtigen_j dem Klienten
 that of.the crime the detective the suspect the client
 [_{-j} _{-k} zu überführen] versprach
 to convict promised

In LTAG, the elementary trees for the relevant verbs look as shown in Figure 12.8. The

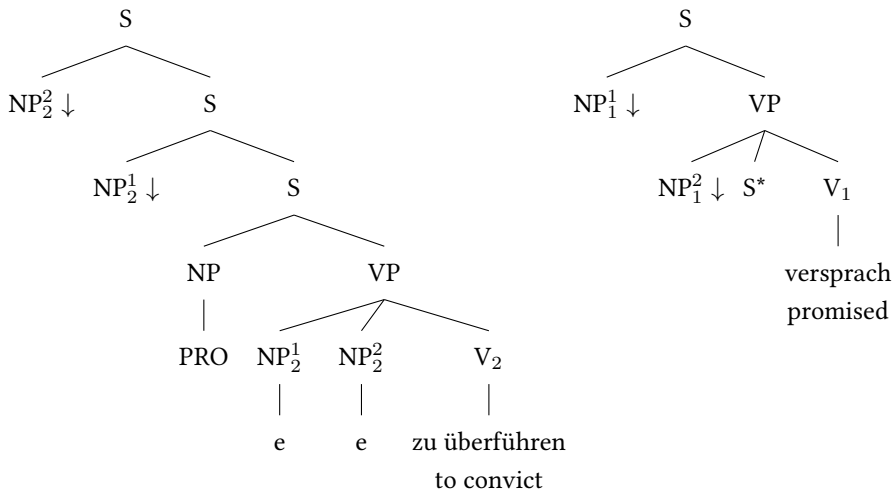


Figure 12.8: Elementary trees of an infinitive and a control verb

verbs are numbered according to their level of embedding. The NP arguments of a verb bear the same index as that verb and each has a superscript number that distinguishes it from the other arguments. The trees are very similar to those in GB. In particular, it is assumed that the subject occurs outside the VP. For non-finite verbs, it is assumed that the subject is realized by PRO. PRO is, like *e*, a phonologically empty pronominal category that also comes from GB. The left tree in Figure 12.8 contains traces in the normal position of the arguments and the relevant NP slots in higher trees positions. An interesting difference to other theories is that these traces only exist in the tree. They are not represented as individual entries in the lexicon as the lexicon only contains words and the corresponding trees.

The tree for *versprach* ‘promised’ can be inserted on any S node in the tree for *zu überführen* ‘to convict’ and results in trees such as those in Figure 12.9. In Figure 12.9, the tree for *versprach* is inserted directly above the PRO NP and in Figure 12.10 above NP₂¹.

It becomes clear that it is not possible this way to derive a tree where an argument of *überführen* ‘to convict’ occurs between the arguments of *versprach* ‘promised’. Joshi, Becker & Rambow (2000) therefore suggest an extension of the LTAG formalism. In MC-TAG, the grammar does not consist of elementary trees but rather finite sets of elementary trees. In every derivational step, a set is selected and the elements of that set are

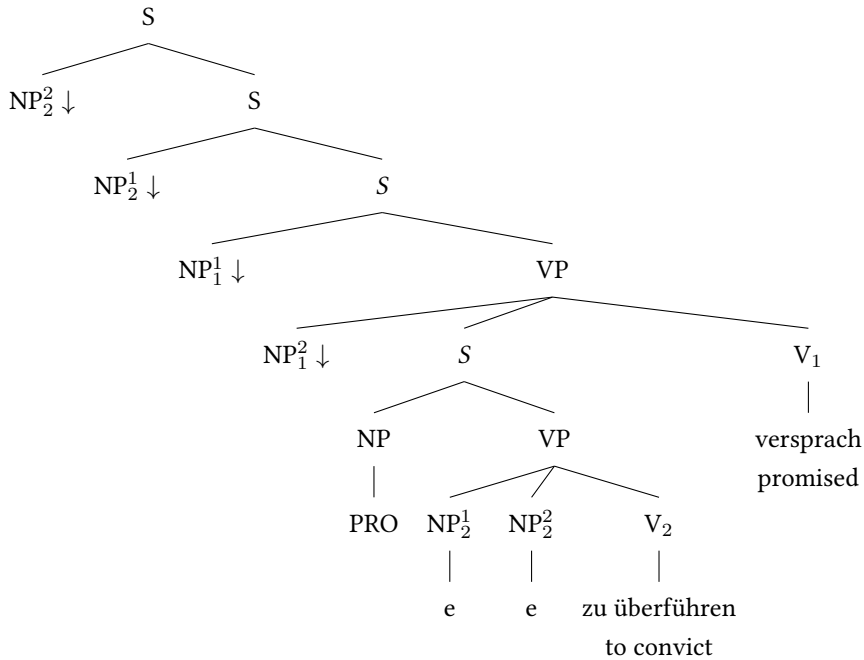


Figure 12.9: Analysis of the order $NP_2^2 NP_2^1 NP_1^1 NP_1^2 V_2 V_1$: adjunction to the lowest S node

simultaneously added to the tree. Figure 12.11 on the next page shows an elementary tree for *versprach* consisting of multiple components. This tree contains a trace of NP_1^1 that was moved to the left. The bottom-left S node and the top-right S node are connected by a dashed line that indicates the dominance relation. However, immediate dominance is not required. Therefore, it is possible to insert the two subtrees in another tree separately from each other and thereby analyze the order in Figure 12.12 on page 405, for example.

Other variants of TAG that allow for other constituent orders are V-TAG (Rambow 1994) and TT-MC-TAG (Lichte 2007).

12.3 Verb position

The verb position can be analyzed in a parallel way to the GPSG analysis with alternative orderings of the verb in a given linearization domain.

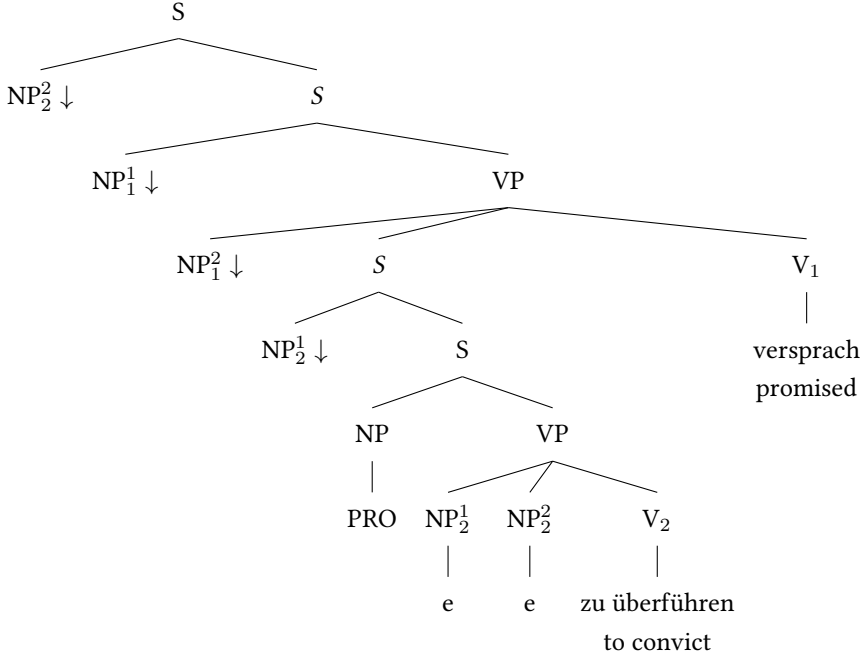


Figure 12.10: Analysis of the order $NP_2^2 NP_1^1 NP_2^2 NP_1^2 V_2 V_1$: adjunction to the S node between NP_2^2 and NP_1^1

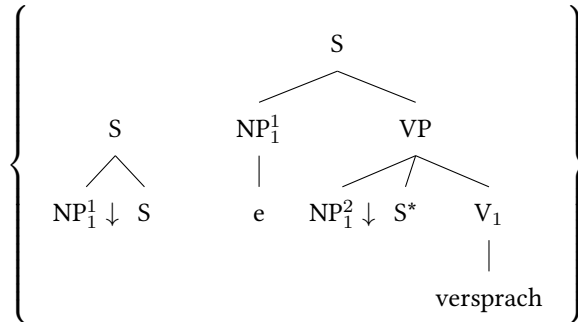


Figure 12.11: Elementary tree for *versprach* consisting of multiple components

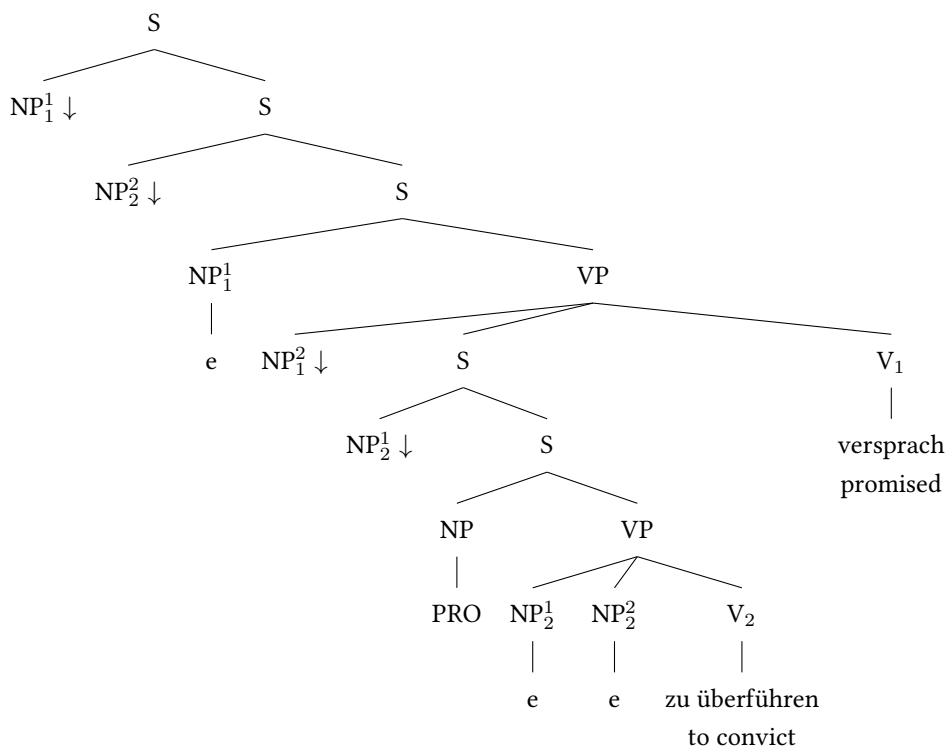


Figure 12.12: Analysis of the order $NP_1^1 NP_2^2 NP_1^1 NP_2^2 V_2 V_1$: adjunction to the S node between NP_2^2 and NP_1^1

12.4 Passive

There is a possible analysis for the passive that is parallel to the transformations in Transformational Grammar: One assumes lexical rules that create a lexical item with a passive tree for every lexical item with an active tree (Kroch & Joshi 1985: 50–51).

Kroch & Joshi (1985: 55) propose an alternative to this transformation-like approach that more adequately handles so-called raising constructions. Their analysis assumes that arguments of verbs are represented in subcategorization lists. Verbs are entered into trees that match their subcategorization list. Kroch and Joshi formulate a lexical rule that corresponds to the HPSG lexical rule that was discussed on page 279, that is, an accusative object is explicitly mentioned in the input of the lexical rule. Kroch and Joshi then suggest a complex analysis of the impersonal passive, which uses a semantic null role for a non-realized object of intransitive verbs (p. 56). This analysis is not necessarily required. It is possible to use the HPSG analysis going back to Haider (1986a) presented in Section 9.2.

There are also proposals in TAG that use inheritance to deal with valency changing processes in general and the passive in particular (Candito (1996) and Kinyon, Rambow, Scheffler, Yoon & Joshi (2006) following Candito). As we saw in Section 10.2 of the Chapter on Construction Grammar, inheritance is not a suitable descriptive tool for valency changing processes. This is because these kinds of processes interact syntactically and semantically in a number of ways and can also be applied iteratively (Müller 2006; 2007c; 2007b: Section 7.5.2; 2013c; 2014a).

12.5 Long-distance dependencies

The analysis of long-distance dependencies in TAG is handled with the standard apparatus: Simple trees are inserted into the middle of other trees. Figure 12.13 on the next page shows an example of the analysis of (10):

- (10) Who_i did John tell Sam that Bill likes _{-i}?

The tree for *WH COMP NP likes* _{-i} belongs to the tree family of *likes* and is therefore present in the lexicon. The tree for *tell* is adjoined to this tree, that is, this tree is inserted in the middle of the tree for *who that Bill likes* _{-i}. This insertion operation can be applied multiple times so that sentences such as (11) where *who* is moved across multiple sentence boundaries can be analyzed:

- (11) Who_i did John tell Sam that Mary said that Bill likes _{-i}?

There is another important detail: Although the tree for (12) has the category S, (12) is not a grammatical sentence of English.

- (12) * who that Bill likes

This has to be captured somehow. In TAG, the marking OA ensures that a tree counts as incomplete. If a tree contains a node with marking OA, then an obligatory adjunction operation must take place at the relevant position.

12.6 New developments and theoretical variants

In Section 12.2, we introduced Multi-Component-TAG. There are a large number of TAG variants with different formal properties. Rambow (1994) gives an overview of the variants existing in 1994. In the following, we will discuss two interesting variants of TAG: Feature Structure-Based TAG = FTAG (Vijay-Shanker & Joshi 1988) and Vector-TAG (V-TAG, Rambow (1994)).

12.6.1 FTAG

In FTAG, nodes are non-atomic (N, NP, VP or S) and instead consist of feature structures. With the exception of substitution nodes, each node has a top structure and a bottom

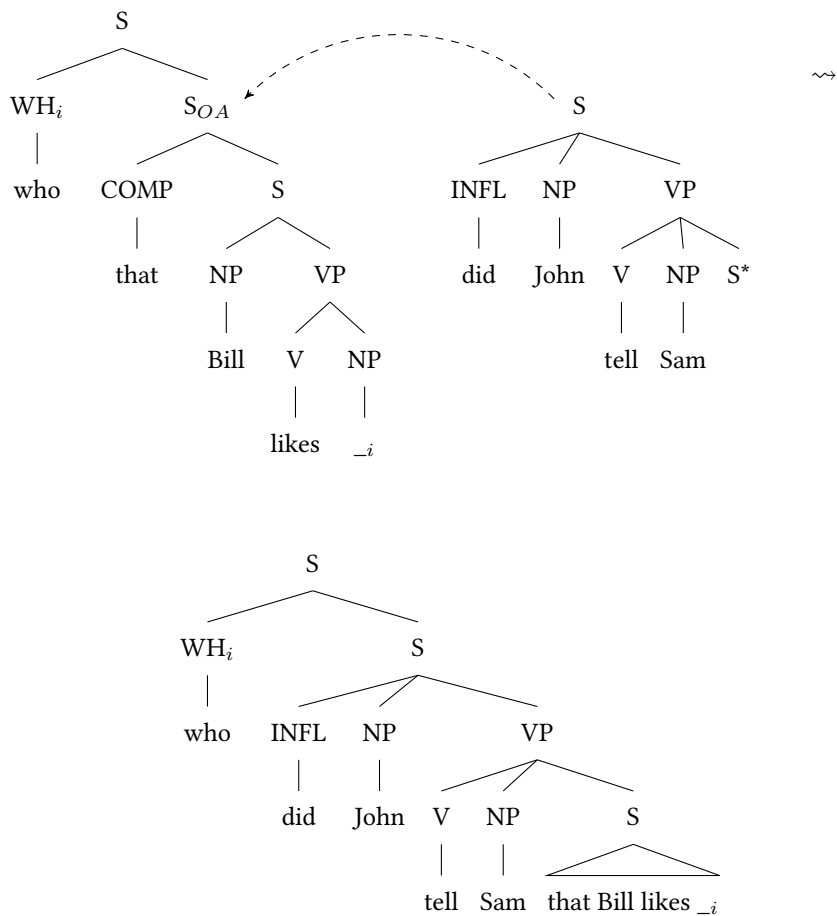


Figure 12.13: Analysis of long-distance dependencies in TAG

structure. The top structure says something about what kind of properties a given tree has inside a larger structure, and the bottom structure says something about the properties of the structure below the node. Substitution nodes only have a top structure. Figure 12.14 shows an example tree for *laughs*. A noun phrase can be combined with the

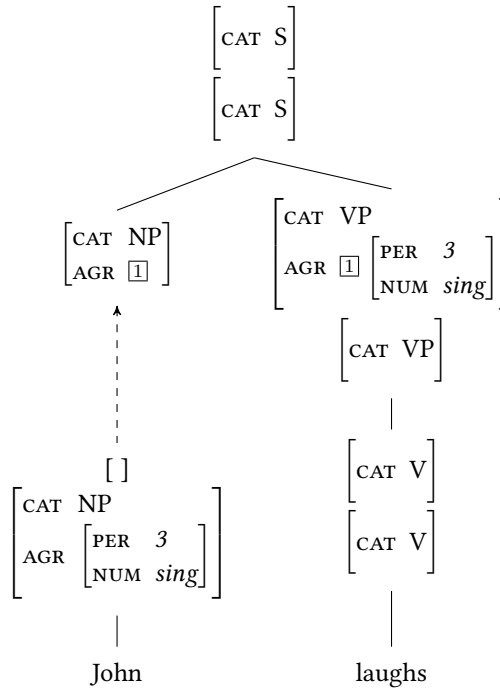


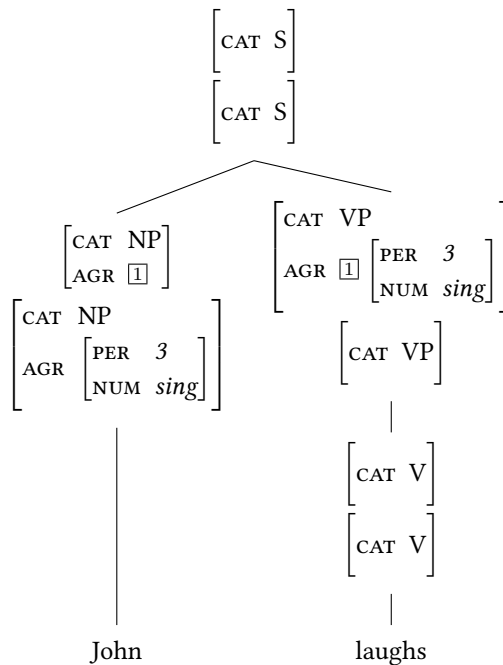
Figure 12.14: Elementary trees for *John* and *laughs* in FTAG

tree for *laughs* in Figure 12.14. Its top structure is identified with the NP node in the tree for *laughs*. The result of this combination is shown in Figure 12.15 on the next page.

In a complete tree, all top structures are identified with the corresponding bottom structures. This way, only sentences where the subject is in third person singular can be analyzed with the given tree for *laughs*, that is, those in which the verb matches the subject in its agreement features.

For adjunction, the top structure of the tree that is being inserted must be unifiable with the top structure of the adjunction site, and the bottom structure of the node marked “*” in the inserted tree (the so-called foot node) must be unifiable with the adjunction site.

The elementary trees discussed so far only consisted of nodes where the top part matched the bottom part. FTAG allows for an interesting variant of the specification of nodes that makes adjunction obligatory in order for the entire derivation to be well-formed. Figure 12.16 on page 410 shows a tree for *laughing* that contains two VP nodes

Figure 12.15: Combination of the trees for *John* and *laughs* in FTAG

with incompatible *MODE* values. In order for this subtree to be used in a complete structure, another tree has to be added so that both VP nodes are separated. This happens by means of an auxiliary tree as shown in Figure 12.16. The upper VP node of the auxiliary tree is unified with the highest VP node of *laughing*. The node of the auxiliary tree marked with “*” is unified with the lower VP node of *laughing*. The result of this is given in Figure 12.17 on page 411.

If a tree is used as a final derivation, the top structures are identified with the bottom structures. Thus, the *AGR* value of the highest VP node is identified with that of the lower one in the tree in Figure 12.17. As such, only NPs that have the same *AGR* value as the auxiliary can be inserted into the NP slot.

This example shows that, instead of the marking for obligatory adjunction that we saw in the section on long-distance dependencies, the same effect can be achieved by using incompatible feature specifications on the top and bottom structures. If there are incompatible top and bottom structures in a tree, then it cannot be a final derivation tree and therefore this means that at least one adjunction operation must still take place in order to yield a well-formed tree.

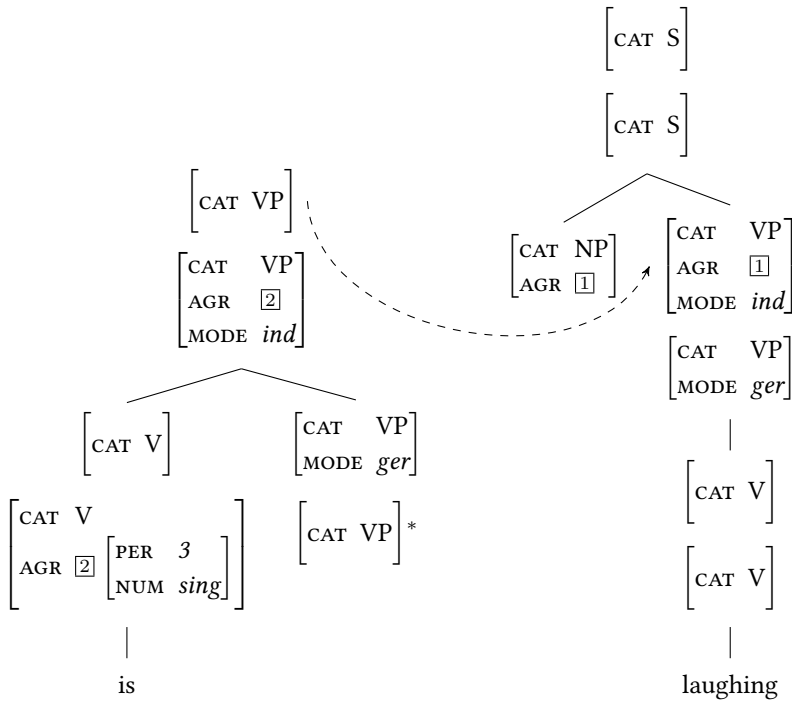


Figure 12.16: Obligatory adjunction in FTAG

12.6.2 V-TAG

V-TAG is a variant of TAG proposed by Owen Rambow (1994) that also assumes feature structures on nodes. In addition, like MC-TAG, it assumes that elementary trees consist of multiple components. Figure 12.18 on the facing page shows the elementary lexical set for the ditransitive verb *geben* ‘give’. The lexicon set consists of a tree for the verb, an empty element of the category VP and three trees where a VP has been combined with an NP. As in MC-TAG, dominance relations are also indicated. The dominance constraints in Figure 12.18 ensure that all lower VP nodes dominate the highest VP node of the tree further to the right. The order of the arguments of the verb as well as the position of the verb is not given. The only fact required is that the noun phrases and the verb dominate the empty verb node. With this lexicon set, it is possible to derive all permutations of the arguments. Rambow also shows how such lexical entries can analyze sentences with verbal complexes. Figure 12.19 on page 412 shows a verbal complex formed from *zu reparieren* ‘to repair’ and *versprochen* ‘promised’ and the relevant dominance constraints. Both of the first NPs have to dominate *versprochen* and the third and fourth NP have to dominate *zu reparieren*. The order of the NPs among each other is not restricted and thus all permutations can be derived.

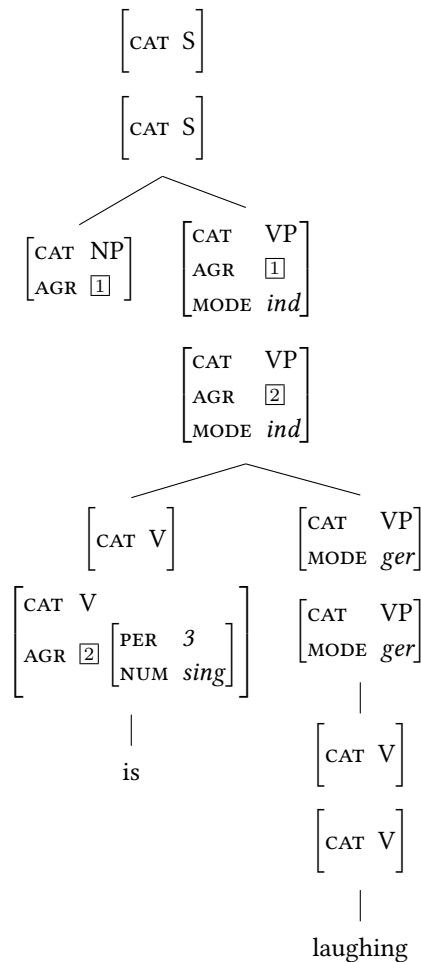


Figure 12.17: Result of obligatory adjunction in FTAG

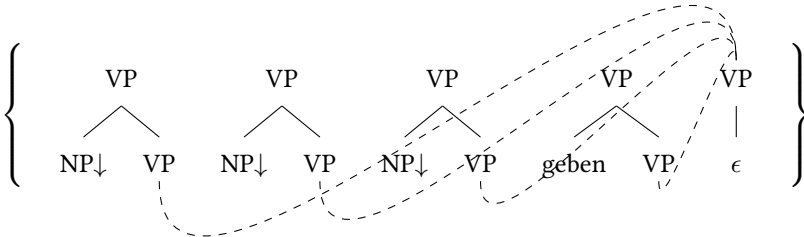
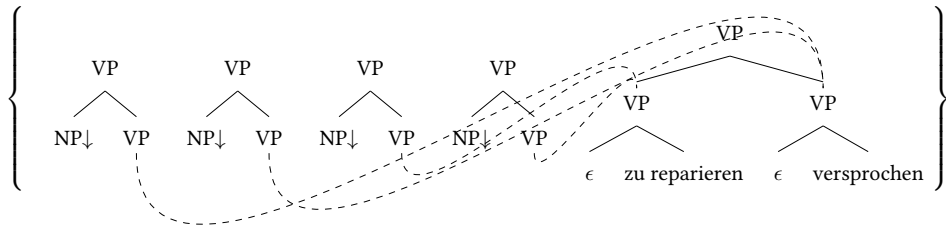


Figure 12.18: Lexicon set for *geben* ‘to give’ in V-TAG according to Rambow (1994: 6)

Figure 12.19: Analysis of the verbal complex *zu reparieren versprochen* in V-TAG

The interesting thing here is that this approach is similar to the one proposed by Berman (1996: Section 2.1.3) in LFG (see Section 7.4): In Berman's analysis, the verb projects directly to form a VP and the arguments are then adjoined.

A difference to other analyses discussed in this book is that there is always an empty element in the derived trees regardless of verb position.

12.6.3 The competence–performance distinction and the generative capacity of tree-local MC-LTAG

In many of the theories discussed in this book, a distinction is made between competence and performance (Chomsky 1965: Section I.1). Competence theories are supposed to describe linguistic knowledge, whereas a performance theory should explain how linguistic knowledge is used and why we make mistakes during speech production and comprehension, etc.

Joshi, Becker & Rambow (2000) discuss examples of center self embedding of relative clauses as those in (13b), and follow Chomsky & Miller (1963: 286) in the assumption that the fact that this kind of embedding is only possible up to three levels should not be described by grammar, but is rather due to processing problems with the hearer independent of their principle abilities with regard to grammar.

- (13) a. dass der Hund bellt, der die Katze jagt, die die Maus gefangen hat
 that the dog barks that the cat chases, that the mouse caught has
 b. Der Hund, [₁ der die Katze, [₂ die die Maus gefangen hat, [₂] jagt [₁]
 the dog that the cat that the mouse caught has caught
 bellt.
 barks
 ‘that the dog that chases the cat that caught the mouse barks.’

What is interesting in this context is that it is possible to construct examples of center embedding so that they are easier to process for the hearer. In this way, it is possible to increase the number of center embeddings possible to process by one and therefore show that all grammars that formulate a restriction to two embedded relative clauses are incorrect. The following example from Hans Uszkoreit is easier to process since all

embedded relative clauses are isolated and the verbs are separated by material from the higher clause.

- (14) Die Bänke, [₁ auf denen damals die Alten des Dorfes, [₂ die allen the benches on which back.then the old.people of.the village that all Kindern, [₃ die vorbeikamen ₃], freundliche Blicke zuwarfen ₂], lange children that came.by friendly glances gave long Stunden schweigend nebeneinander saßen ₁], mussten im letzten Jahr hours silent next.to.each.other sat must in last year einem Parkplatz weichen. a car.park give.way.to
 ‘The benches on which the older residents of the village, who used to give friendly glances to all the children who came by, used to sit silently next to one another had to give way to a car park.’

For other factors that play a role in processing, see Gibson (1998).

Joshi et al. (2000) discuss verbal complexes with moved arguments. The general pattern that they observe has the form shown in (15):

- (15) $\sigma(\text{NP}_1 \text{NP}_2 \dots \text{NP}_n) \text{V}_n \text{V}_{n-1} \dots \text{V}_1$

Here, σ stands for any permutation of noun phrases and V_1 is the finite verb. The authors investigate the properties of Lexicalized Tree Adjoining Grammar (LTAG) with regard to this pattern and notice that LTAG cannot analyze the order in (16) if we want to take the semantics seriously.

- (16) $\text{NP}_2 \text{NP}_3 \text{NP}_1 \text{V}_3 \text{V}_2 \text{V}_1$

Instead, they propose the extension of TAG discussed in Section 12.2; so-called *tree-local multi-component LTAG* (Tree-local MC-LTAG). They show that tree-local MC-TAG can analyze (16) but not (17) with the correct semantics. They claim that these orders are not possible in German and argue that in this case, unlike the relative clause examples, one has both options, that is, the unavailability of such patterns can be explained as a performance phenomenon or as a competence phenomenon.

- (17) $\text{NP}_2 \text{NP}_4 \text{NP}_3 \text{NP}_1 \text{V}_4 \text{V}_3 \text{V}_2 \text{V}_1$

If we treat this as a performance phenomenon, then we are making reference to the complexity of the construction and the resulting processing problems for the hearer. The fact that these orders do not occur in corpora can be explained with reference to the principle of cooperativeness. Speakers normally want to be understood and therefore formulate their sentences in such a way that the hearer can understand them. Verbal complexes in German with more than four verbs are hardly ever found since it is possible to simplify very complex sentences with multiple verbs in the right sentence bracket by extraposing material and therefore avoiding ambiguity see Netter (1991: 5) and Müller (2007b: 262))

The alternative to a performance explanation would involve using a grammatical formalism which is just powerful enough to allow embedding of two verbs with reordering, but rules out embedding of three verbs with the corresponding reordering. Joshi et al. (2000) opt for this solution and therefore attribute the impossibility of the order of arguments in (17) to competence.

In HPSG (and also in Categorical Grammar and in some GB analyses), verbal complexes are analyzed by means of argument composition (Hinrichs & Nakazawa 1989a; 1994). Under this approach, a verbal complex behaves exactly like a simplex verb and the arguments of the verbs involved can be placed in any order. The grammar does not contain any restriction on the number of verbs that can be combined, nor any constraints that ban embedding below a certain level. In the following, I will show that multiple movements are ruled out by communication rules that are applied even with cases of two simple transitive verbs. The impossibility of embedding four or more verbs in fact should be explained as a performance issue.

Before I present arguments against a competence-based approach to (17), I will make a more general comment: Corpora cannot help us here since one does not find any instances of verbs with four or more embeddings. Bech (1955) provides an extensive collection of material but had to construct the examples with four embedded verbs. Meurers (1999b: 94–95) gives constructed examples with five verbs that contain multiple auxiliaries or modal verbs. These examples are hardly processable and are not relevant for the discussion here since the verbs in (17) have to select their own arguments. There are therefore not that many verbs left when constructing examples. It is possible to only use subject control verbs with an additional object (e.g. *versprechen* ‘to promise’), object control verbs (e.g. *zwingen* ‘to force’) or AcI verbs (e.g. *sehen* ‘to see’ or *lassen* ‘to let’) to construct examples. When constructing examples, it is important to pay attention to the fact that all the nouns involved differ as much as possible with regard to their case and their selectional restrictions (e.g. animate/inanimate) since these are features that a hearer/reader could use to possibly assign moved arguments to their heads. If we want to have patterns such as (17) with four NPS each with a different case, then we have to choose a verb that governs the genitive. There are only a very small number of such verbs in German. Although the example constructed by Joshi et al. (2000) in (9b) fulfills these requirements, it is still very marked. It therefore becomes clear that the possibility of finding a corresponding example in a newspaper article is extremely small. This is due to the fact that there are very few situations in which such an utterance would be imaginable. Additionally, all control verbs (with the exception of *helfen* ‘to help’) require an infinitive with *zu* ‘to’ and can also be realized incoherently, that is, with an extraposed infinitive complement without verbal complex formation. As mentioned above, a cooperative speaker/author would use a less complex construction and this reduces the possibility that these kinds of sentences arise even further.

Notice that tree-local MC-LTAG does not constrain the number of verbs in a sentence. The formalism allows for an arbitrary number of verbs. It is therefore necessary to assume, as in other grammatical theories, that performance constraints are responsible for the fact that we never find examples of verbal complexes with five or more verbs. Tree-

local MC-LAG makes predictions about the possibility of arguments to be reordered. I consider it wrong to make constraints regarding mobility of arguments dependent on the power of the grammatical formalism since the restrictions that one finds are independent of verbal complexes and can be found with verbs taking just two arguments. The problem with reordering is that it still has to be possible to assign noun phrases to their heads. If this assignment leads to ambiguity that cannot be resolved by case, selectional restrictions, contextual knowledge or intonation, then the unmarked constituent order is chosen. Hoberg (1981: 68) shows this very nicely with examples similar to the following:⁵

- (18) a. Hanna hat immer schon gewußt, daß das Kind sie verlassen will.
 Hanna has always PRT known that the child she leave wants
 ‘Hanna has always known that the child wants to leave her.’
 b. # Hanna hat immer schon gewußt, daß sie das Kind verlassen will.
 Hanna has always PRT known that she the child leave wants
 Preferred reading: ‘Hanna has always known that she wants to leave the child.’
 c. Hanna hat immer schon gewußt, daß sie der Mann verlassen will.
 Hanna hat immer schon gewußt, daß sie the.NOM man verlassen will.
 ‘Hans always known that the man wants to leave her.’

It is not possible to reorder (18a) to (18b) without creating a strong preference for another reading. This is due to the fact that neither *sie* ‘she’ nor *das Kind* ‘the child’ are unambiguously marked as nominative or accusative. (18b) therefore has to be interpreted as Hanna being the one that wants something, namely to leave the child. This reordering is possible, however, if at least one of the arguments is unambiguously marked for case as in (18c).

For noun phrases with feminine count nouns, the forms for nominative and accusative as well as genitive and dative are the same. For mass nouns it is even worse. If they are used without an article, all cases are the same for feminine nouns (e. g. *Milch* ‘milk’) and also for masculines and neuters with exception of the genitive. The following examples from Wegener (1985: 45) show that it is not immediately possible to switch the dative and accusative objects, whereas this is possible if the nouns are used with articles:

- (19) a. Sie mischt Wein Wasser bei.
 she mixes wine water PRT
 ‘She mixes water into the wine.’
 b. Sie mischt Wasser Wein bei.
 she mixes water wine PRT
 ‘She mixes wine into the water’

⁵ Instead of *das* ‘the’, Hoberg uses the possessive pronoun *ihr* ‘her’. This makes the sentences more semantically plausible but one then gets interference from the linearization requirements for bound pronouns. I have therefore replaced the pronouns with the definite article.

- c. Sie mischt dem Wein das Wasser bei.
 she mixes the.DAT wine the.ACC water PRT
 ‘She mixes water into the wine.’
- d. Sie mischt das Wasser dem Wein bei.
 she mixes the.ACC water the.DAT wine PRT
 ‘She mixes the water into the wine.’

The two nouns can only be switched if it is clear from the context what it means (e.g. through explicit negation of the opposite) and if the sentence carries a certain intonation.

The problem with verbal complexes is now that with four noun phrases, two almost always have the same case if one does not wish to resort to the few verbs governing the genitive. A not particularly nice-sounding example of morphologically unambiguously marked case is (20):

- (20) weil er den Mann dem Jungen des Freundes gedenken
 because he.NOM the.ACC man the.DAT boy of.the.GEN friend remember
 helfen lassen will
 help let wants
 ‘because he wants to let the man help the boy remember his friend.’

Another strategy is to choose verbs that select animate and inanimate objects so that animacy of the arguments can aid interpretation. I have constructed such an example where the most deeply embedded predicate is not a verb but rather an adjective. The predicate *leer fischen* ‘to fish empty’ is a resultative construction that should be analyzed parallel to verbal complexes (Müller 2002a: Chapter 5).

- (21) weil niemand₁ [den Mann]₂ [der Frau]₃ [diesen Teich]₄
 because nobody.NOM the.ACC man the.DAT woman this.ACC pond
 leer₄ fischen₃ helfen₂ sah₁
 empty fish help saw
 ‘because nobody helped the woman to see the man catch fish in the pond until it was empty.’

If one reads the sentences with the relevant pauses, it is comprehensible. Case is unambiguously marked on the animate noun phrases and our word knowledge helps us to interpret *diesen Teich* ‘this pond’ as the argument of *leer* ‘empty’.

The sentence in (21) would correctly be analyzed by an appropriately written tree-local MC-LTAG and also by argument composition analyses for verbal complexes and resultative constructions. The sentence in (22) is a variant of (21) that corresponds exactly to the pattern of (17):

- (22) weil [der Frau]₂ [diesen Teich]₄ [den Mann]₃ niemand₁
 because the.DAT woman this.ACC pond the.ACC man nobody.NOM
 leer₄ fischen₃ helfen₂ sah₁
 empty fish help saw
 ‘because nobody saw the man help this woman fish the pond until it was empty.’

(22) is more marked than (21), but this is always the case with local reordering (Gisbert Fanselow, p. M. 2006). This sentence should not be ruled out by the grammar. Its markedness is more due to the same factors that were responsible for the markedness of reordering of arguments of basic verbs. Tree-local MC-LTAG can not correctly analyze sentences such as (22), which shows that this TAG variant is not sufficient for analyzing natural language.

There are varying opinions among TAG researchers as to what should be counted as competence and what should be counted as performance. In this way, Rambow (1994: 15) argues that one should not exclude reordering that cannot be processed by means of competence grammar or the grammatical formalism. In Chapter 6, he presents a theory of performance that can explain why the reordering of arguments of various verbs in the middle field is harder to process. One should therefore opt for TAG variants such as V-TAG or TT-MC-TAG (Lichte 2007) that are powerful enough to analyze the diverse reorderings and then also use a performance model that makes it possible to explain the graded differences in acceptability.

An alternative to looking for a grammatical formalism with minimal expressive power is to not restrict the grammatical formalism at all with regard to its expressive power and instead develop as restrictive linguistic theories as possible. For further discussion of this point, see Chapter 18.

12.7 Summary and classification

In sum, we have seen the following: LTAG is lexicalized, that is, there is at least one lexical element in every tree. There are not any trees that correspond to the rule $S \rightarrow NP, VP$ since no words are mentioned in this rule. Instead, there are always complex trees that contain the subject NP and the VP. Inside the VP, there can be as much structure as is necessary to ensure that the verb is contained in the tree. As well as the head, elementary trees in LTAG always contain the arguments of the head. For transitive verbs, this means that both the subject and the object have to be components of the elementary tree. This is also true of the trees used to analyze long-distance dependencies. As shown in Figure 12.13, the object must be part of the tree. The fact that the object can be separated from the verb by multiple sentence boundaries is not represented in the elementary tree, that is, recursive parts of grammar are not contained in elementary trees. The relevant effects are achieved by adjunction, that is, by insertion of material into elementary trees. The elementary tree for extraction in Figure 12.13 differs from the elementary tree for *likes* given in Figure 12.4 for the use in normal SVO clauses. Every minimal construction, in which *likes* can occur (subject extraction, topicalization, subject relative clauses, object relative clauses, passive, ...) needs its own elementary tree (Kallmeyer & Joshi 2003: 10). The different elementary trees can be connected using lexical rules. These lexical rules map a particular tree treated as underlying to other trees. In this way, it is possible to derive a passive tree from an active tree. These lexical rules are parallel to transformations in Transformational Grammar, however, one should bear in mind that there is always a lexical element in the tree, which makes the entire

grammar more restricted than grammars with free transformations.

An interesting difference to GB and variants of LFG, CG, and HPSG that assume empty elements is that the variants of TAG presented here⁶ do not contain empty elements in the lexicon. They can be used in trees but trees are listed as a whole in the lexicon.

Elementary trees can be of any size, which makes the analysis of idioms interesting (see Chapter 19). Since recursion is separate, trees can contain elements that appear very far away from each other in the derived tree (extended domains of locality).

Kasper, Kiefer, Netter & Shanker (1995) show that it is possible to transfer HPSG grammars that fulfill certain requirements into TAG grammars. This is interesting as in this way one arrives at a grammar whose complexity behavior is known. Whereas HPSG grammars are generally in the type-0 area, TAG grammars can, depending on the variant, fall into realm of type-2 languages (context-free) or even in the more expansive area of the mildly context-sensitive grammars. Yoshinaga, Miyao, Torisawa & Tsujii (2001) have developed a procedure for translating FB-LTAG grammars into HPSG grammars.

Comprehension questions

1. How are long-distance dependencies analyzed in TAG? Does one need empty elements for this?
2. Is it possible to analyze the reordering of arguments of multiple verbs using standard TAG processes?

Exercises

1. Analyze the following string in LTAG:

(23) der dem König treue Diener
 the the.DAT king loyal servant
 ‘the servant loyal to the king’

Further reading

Some important articles are Joshi, Levy & Takahashi (1975), Joshi (1987a), and Joshi & Schabes (1997). Many works discuss formal properties of TAG and therefore are not particularly accessible for linguistically interested readers. Kroch & Joshi (1985) give a good overview of linguistic analyses. An overview of linguistic and computational linguistic works in TAG can be found in the volume edited by Abeillé and Rambow from 2000. Rambow (1994) compares his TAG variant (V-TAG) with Karttunen’s *Radical*

⁶ See Rambow (1994) and Kallmeyer (2005: 194), however, for TAG analyses with an empty element in the lexicon.

Lexicalism approach, Uszkoreit's GPSG, Combinatorial Categorical Grammar, HPSG and Dependency Grammar.

Shieber & Johnson (1993) discuss psycholinguistically plausible processing models and show that it is possible to do incremental parsing with TAG. They also present a further variant of TAG: synchronic TAG. In this TAG variant, there is a syntactic tree and a semantic tree connected to it. When building syntactic structure, the semantic structure is always built in parallel. This parallelly built structure corresponds to the level of Logical Form derived from S-Structure using transformations in GB.

Rambow (1994: Chapter 6) presents an automaton-based performance theory. He applies it to German and shows that the processing difficulties that arise when reordering arguments of multiple verbs can be explained.

Kallmeyer & Romero (2008) show how it is possible to derive MRS representations directly via a derivation tree using FTAG. In each top node, there is a reference to the semantic content of the entire structure and each bottom node makes reference to the semantic content below the node. In this way, it becomes possible to insert an adjective (e.g. *mutmaßlichen* 'suspected') into an NP tree *alle Mörder* 'all murderers' so that the adjective has scope over the nominal part of the NP (*Mörder* 'murderers'): For adjunction of the adjective to the N node, the adjective can access the semantic content of the noun. The top node of *mutmaßlichen* is then the top node of the combination *mutmaßlichen Mörder* 'suspected murderers' and this ensures that the meaning of *mutmaßlichen Mörder* is correctly embedded under the universal quantifier.

Grammatical theory

This book introduces formal grammar theories that play a role in current linguistics or contributed tools that are relevant for current linguistic theorizing (Phrase Structure Grammar, Transformational Grammar/Government & Binding, Generalized Phrase Structure Grammar, Lexical Functional Grammar, Categorical Grammar, Head-Driven Phrase Structure Grammar, Construction Grammar, Tree Adjoining Grammar). The key assumptions are explained and it is shown how the respective theory treats arguments and adjuncts, the active/passive alternation, local reorderings, verb placement, and fronting of constituents over long distances. The analyses are explained with German as the object language.

In a final chapter the approaches are compared with respect to their predictions regarding language acquisition and psycholinguistic plausibility. The Nativism hypothesis that assumes that humans possess genetically determined innate language-specific knowledge is examined critically and alternative models of language acquisition are discussed. In addition this chapter addresses issues that are discussed controversially in current theory building as for instance the question whether flat or binary branching structures are more appropriate, the question whether constructions should be treated on the phrasal or the lexical level, and the question whether abstract, non-visible entities should play a role in syntactic analyses. It is shown that the analyses that are suggested in the respective frameworks are often translatable into each other. The book closes with a section that shows how properties that are common to all languages or to certain language classes can be captured.

“With this critical yet fair reflection on various grammatical theories, Müller fills what was a major gap in the literature.” Karen Lehmann, *Zeitschrift für Rezensionen zur germanistischen Sprachwissenschaft*, 2012

“Stefan Müller’s recent introductory textbook, ‘Grammatiktheorie’, is an astonishingly comprehensive and insightful survey for beginning students of the present state of syntactic theory.” Wolfgang Sternefeld und Frank Richter, *Zeitschrift für Sprachwissenschaft*, 2012

“This is the kind of work that has been sought after for a while. [...] The impartial and objective discussion offered by the author is particularly refreshing.” Werner Abraham, *Germanistik*, 2012

ISBN 978-3-944675-21-3



9 783944 675213