

A Formal Theory of Word Order:
A Case Study in West Germanic

Mike Reape

Centre for Cognitive Science
University of Edinburgh



Contents

1 Introduction	1
1-1 Overview	1
1-1.1 Discontinuous constituency and semi-free word order . . .	2
1-1.2 Cross-linguistic variation in word order	10
1-1.3 Formalisation in a homogeneous logical formalism	11
1-1.4 Organisation of the Thesis	12
1-2 Introduction	13
1-3 The Role of Syntax in Linguistic Theory	14
1-3.1 Theoretical nonobservables and modelling of empirical do- mains	15
1-3.2 The Role of Constituent Structure	15
1-3.3 Tree Hacking	17
1-3.4 Tectogrammatic and Phenogrammatic Structure	18
1-3.5 Functor-argument Structure	19
1-3.6 Compositionality	21
1-3.7 Syntactic Categories	21
1-3.8 Unavoidable Syntax	21
1-4 Justification for the Rejection of the Conventional Role of Syntax	22
1-4.1 Empirical motivation and empirical adequacy in syntax . .	22
1-4.2 Syntax as a theory internal construct	28
1-4.3 Empirical justification for our approach	29
1-5 Generalisations and Formalisation	30
1-5.1 Theory-internal vs. Empirical Generalisations	30

1-5.2	Vertical Locality	31
1-5.3	Getting Things in Order	31
1-5.4	Markedness and Canonality	32
1-6	Syntactic Categories in West Germanic	32
1-7	Synchronic and Diachronic Syntax	36
1-8	Scientific Methodology	38
1-8.1	The Isomorphic Formalism Requirement	39
1-8.2	Theory Presentations and Theories	42
1-8.3	The relationship between a formalised theory, the mathematical domain and the empirical domain	43
1-8.4	Prediction	44
1-8.5	Psycholinguistics	45
1-8.6	Computational Linguistics	46
1-8.7	Descriptive Felicity	47
1-9	Cross-linguistic Variation	47
1-9.1	Universal Grammar, Theories and Subtheories	47
1-10	An Example from German	49

2 The Theoretical Framework – from Phrase Structure Grammar to Word Order Domains 53

2-1	Context-Free Phrase Structure Grammar	53
2-2	Categorial Grammar	55
2-3	Generalized Phrase Structure Grammar	56
2-3.1	‘Subject-aux inversion’ without subject, aux or inversion	60
2-4	Liberation Metarules	62
2-5	Uszkoreit (1987)	64
2-5.1	Grammar: Basics and Scope	65
2-5.2	The Modified LP Framework	68
2-5.3	<i>Linear Precedence in Discontinuous Constituents: Complex Fronting in German</i>	71
2-6	Head-Driven Phrase Structure Grammar	76
2-6.1	Inverted Clauses	76

2-6.2	Principles of Constituent Order	77
2-7	Bech (1955)	79
2-7.1	<i>status</i>	79
2-7.2	<i>oriëntation</i>	80
2-7.3	<i>Verbalfeld</i>	80
2-7.4	<i>kohärenzfeld</i>	81
2-8	Evers (1975)	83
2-9	Domain Union	90
2-10	A theory of linear precedence constraints	94
2-10.1	Linear precedence constraints over word order domains . .	94
2-10.2	Partially Ordered LP constraints	94
2-10.3	Partially Ordered Sets of LP Constraints	97
2-10.4	Partially Ordered Sets of Partially Ordered Sets of LP con- straints	100
2-10.5	Markedness	101
2-10.6	Statistics and a Discovery Procedure for Linear Precedence Constraints	102
2-11	Dowty (1991)	105
2-11.1	Introduction	105
2-11.2	The Theoretical Framework	105
2-11.3	An Example from Finnish	107
2-11.4	Examples from German and Dutch	109
2-11.5	Discussion	109

3 An HPSG Formalisation of German Word Order and Syntax **113**

3-1	Introduction	113
3-2	Semantic Interpretation	123
3-3	Semantic Representations	126
3-3.1	The HPSG Treatment of Semantic Interpretation	126
3-3.2	The Syntax of Semantic Representations	128
3-3.3	First Order Translation of Semantic Representations . . .	131

3-3.4	The Syntax of InL	133
3-3.5	The Semantics of InL — InL and DRT	135
3-3.6	InL Translation of Semantic Representations	136
3-4	The Syntax and Semantics of Adjuncts	138
3-4.1	The HPSG Treatment of Adjuncts	138
3-4.2	An Alternative Treatment of Adjunct Syntax and Semantics	141
3-5	The Syntax and Semantics of Specifiers	147
3-5.1	The HPSG Treatment of Specifiers	147
3-5.2	An Alternative Treatment of Specifiers	149
3-6	The HPSG Treatment of Quantifier Scope	150
3-6.1	Pollard and Sag (1987)	150
3-6.2	Pollard and Sag (1992)	152
3-7	The Syntax and Semantics of Quantifiers	154
3-8	Scope of Quantification	157
3-9	Scope of Modification	163
3-9.1	Scope of modification: unchanged by permutation	164
3-9.2	Scope of modification: affected by permutation	165
3-9.3	Scope of modification: ambiguous	166
3-10	Raising	167
4	An Account of German Word Order	170
4-1	Introduction	170
4-2	Netter (1991)	177
4-3	Some assumptions about German verbs	178
4-4	Extraposition	179
4-5	Subjectless Constructions	181
4-6	Pronominalization	183
4-7	Scrambling	185
4-8	Scope of Adjuncts and Negation	194
4-9	Topicalization and Extraposition	208
4-10	Intrapolation of VPs in the Mittelfeld	216

4-10.1	The Empirical Facts	217
4-10.2	A Domain Union Analysis	219
4-10.3	Intraposited VPs and Argument Inheritance	221
4-11	Against Argument Inheritance	223
4-11.1	The Empirical Facts	223
4-12	Raising	227
4-13	Conclusion	230
5	A Comparison of German, Dutch, Zurich German and English	231
5-1	An Analysis of Dutch	231
5-1.1	Standard Dutch	231
5-1.2	The Groningen Dialect	234
5-2	An Analysis of Zurich German	238
5-2.1	Basic Clause Order	238
5-2.2	\approx VP Extraposition	240
5-3	Comparative Clause Structure	246
6	Conclusions	250
6-1	A Selective Summary	250
6-2	Directions for Future Research	251
A	Formalisation of the HPSG Formalism	255
A-1	Introduction	255
A-2	The Language L	256
A-2.1	Syntax	257
A-2.2	Semantics	257
A-2.3	Proof Theory	257
A-3	The Language L^+	258
A-3.1	Motivation	258
A-3.2	Predicate Modal Logic	261

A-3.3	Syntax	262
A-3.4	Semantics	263
A-3.5	Some examples	264
A-3.6	Proof theory	265
A-4	Intensionality and Extensionality	267
A-5	Nonwellfounded Structures	270
A-6	Other Issues	270
A-7	Nonwellfounded Sets	272
A-8	The translation of the HPSG formalism into L^+	274
A-8.1	The translation of the feature structure notation of HPSG into L^+	274
A-8.2	The translation of axioms in HPSG into L^+	276

Acknowledgements

I owe a great many people for their help and support while I worked on this thesis. I'd like to thank the following people: Ewan Klein, my supervisor, for supervisory guidance and providing me with jobs to keep going; Elisabet Engdahl, for support during difficult times; Chris Mellish for support and interest in the formal aspects of my work; Professor John Byrne of Trinity College Dublin for making my life easier in many ways while I finished the thesis; Betty Hughes (University of Edinburgh) and Irene O'Neill (Trinity College Dublin) for helping to make my life easier than it would have been otherwise; Klaus Netter of Universiteit Stuttgart and DFKI, Saarbrücken, Mark Hepple (now of Cambridge University), Professor Gabriel Bès of the Université Clermont-Ferrand, Professor Hans Uszko-reit of DFKI, Saarbrücken, Pete Whitelock (now of Sharp Laboratories Europe) and Max Volino for discussions of the theory; Henk Zeevat (now of Universiteit van Amsterdam) and Jo Calder (now of Simon Fraser University) for general discussion; Michael Newton (aka Kimba) for discussion of abstraction and formalisation; Mark Hepple, Kathrin Cooper and Gosse Bouma of Rijksuniversiteit Groningen for discussions of German and Dutch data; Glyn Morrill for discussion of linguistic methodology; Jerry Seligman (now of Indiana State University) and Patrick Blackburn (now of Rijksuniversiteit Utrecht) for discussion of issues of formalisation and methodology; Peter Blok (Rijksuniversiteit Groningen), Gosse Bouma (Rijksuniversiteit Groningen), Kathrin Cooper, Jack Hoeksema (Rijksuni-versiteit Groningen), Andreas Ludwig, Marc Moens and Klaus Netter for acting as informants and Professor David Dowty, Carl Pollard and Bob Kasper of Ohio State University for their general interest in my work.

I would also like to thank Professor Naomi Sager (New York University) and Lynette Hirshman for introducing me to computational linguistics and teaching me the importance of generalisations. I'd also like to thank Professor Richard Costello of the New York University Linguistics Department for teaching me the wonders of comparative and historical linguistics.

I'd also like to thank Linda Shockey and Henry Thompson (both formerly of The Centre for Speech Technology Research) for providing me with employment and encouragement.

I'd also like to thank the anonymous reviewers of *German Grammar in HPSG* for comments on a version of Ch. 3.

I would especially like to thank Klaus Netter of DFKI (Saarbrücken) for helpful

discussions about German data and syntactic theory over a long period of time. This thesis would not have been possible without his help and interest. I would also like to specially thank Mark Hepple for reading early drafts of Chs. 3-5 and for his interest in my research and the many useful discussions I had with him.

In addition, I would like to thank Professor Frans Zwarts of Rijksuniversiteit Groningen, Professor Hans Uszkoreit of DFKI, Saarbrücken and Professor Remko Scha of Universiteit van Amsterdam for their support and encouragement at various times.

I'd also like to thank the members of all the colloquiums, symposiums and conferences at which I have presented some of this work. They include the Centre for Cognitive Science Parsing Workshop, the Centre for Cognitive Science Parametric Variation Workshop, the Spring, 1989 LAGB meeting in Exeter, the Linguistics Colloquium and the Departmental Colloquium at Vakgroep α -informatica, Universiteit van Amsterdam, the German Centre for Artificial Intelligence (DFKI) in Saarbrücken, the *German in HPSG* workshop at DFKI, the Symposium on Discontinuous Constituency in Tilburg, the 1989 Meeting of the European Association for Computational Linguistics and the Information Sciences Institute, USC.

I'd also like to thank all of my friends and colleagues at the Centre for Cognitive Science, University of Edinburgh and the Department of Computer Science, Trinity College Dublin for providing such pleasant work environments.

I'd also like to thank Ronan Waldron and Paul Harrington of Trinity College Dublin for helping me solve various technical problems when I first arrived at Trinity.

I owe a big debt of gratitude to my office mate of four years, Jo Calder. Jo has been a good friend, a great colleague and a constant source of new ideas. Furthermore, he patiently put up with my idiosyncracies without a single complaint. He also knows every key binding to every Unix application ever written. I miss him dearly.

I would also like to thank my brothers, Dave, Bill and Sean, my brother-in-laws, John and Tim, my father, Bill, my father-in-law, Bud, and my Uncle Pat for working on my house in Webster Groves. Without their help, I would never have been able to afford to pursue postgraduate study.

I'd also like to thank my in-laws, Bud and Rose Yowell, for moral and occasional financial support and I'd like to thank my parents, Bill and Mary, for encouraging my educational goals for as long as I remember.

I also owe an enormous amount to my daughter Jamie and my son Chris. They have had to suffer through years of a father who spent more time with his keyboard than he did with them. I hope I can make it up to them now that this is over.

Finally, I'd like to thank my wife Sue for her constant love and support through everything we've been through together during the pursuit of this thesis. I will

love her always.

Chapter 1

Introduction

“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 obscure 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.”
Noam Chomsky, *Syntactic Structures*, 1957

“... true formalization is rarely a useful device in linguistics.”
Noam Chomsky, *The Kyoto Lectures*, 1987

1-1 Overview

This thesis has three primary goals. The first goal is to provide a treatment of *bounded* discontinuous constituency and word order in general and semi-free word order in particular. Bounded discontinuous constituency is taken to mean the kind of category-bounded discontinuity typically found in the German *Mittelfeld* and in Dutch “cross-serial” dependency constructions. It does not mean unbounded dependencies such as *wh*-movement. The second goal is to provide an alternative account of cross-linguistic variation in word order (particularly in West Germanic) to the *Principles and Parameters* approach of Government and

Binding Theory. The third goal is to formalise the account in a single homogeneous logical formalism which is not based on rewrite rules or other formal language theoretic machinery. These three goals are discussed in more detail below.

1-1.1 Discontinuous constituency and semi-free word order

Virtually every modern theory of syntax assumes that phrases are formed by the concatenation of their constituents. That is, subphrases are adjacent in a phrase. For example, in the sentence *the boy kicked his dog*, the Det *the* and the N *boy* are lexical sister subphrases of the NP *the boy*, the Det *his* and the N *dog* are lexical sister subphrases of the NP *his dog*, the V *kicked* and the NP *his dog* are sister subphrases of the VP *kicked his dog* and the NP *the boy* and the VP *kicked his dog* are sister subphrases of the S *the boy kicked his dog*. This “adjacency principle” follows immediately from the assumption of phrase structure trees as the basis of syntax. The phrase structure tree for *the boy kicked his dog* is Fig. 1-1.

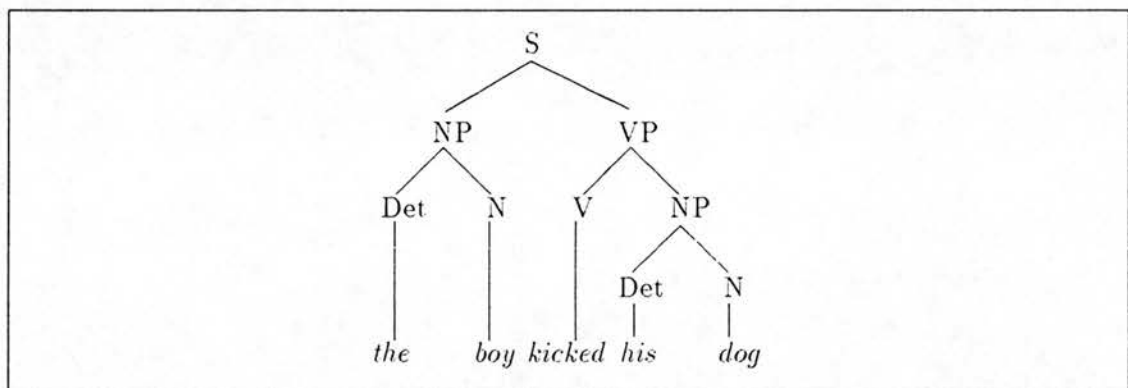


Figure 1-1: Syntax tree for *the boy kicked his dog*

Here we see that *the boy kicked his dog* is formed by reading the leaves of the tree (i.e., the *terminal yield*) from left to right. In this case, the leftmost Det and N are sisters so their terminal symbols *the* and *boy* are adjacent in the tree. The rightmost Det and N are also sisters and so the terminal symbols *his* and *dog* are adjacent in the tree. The V and the rightmost NP are sisters so the terminal yield of the V, *kicked*, and the rightmost NP, *his dog* are adjacent. Finally, the terminal yield of the leftmost NP, *the boy*, and the VP, *kicked his dog*, are adjacent. In each subtree, the terminal yield of all of the daughters are adjacent.

Discontinuous constituency refers to cases where apparent subphrases of a phrase are not all adjacent. For example, consider the subordinate clause *daß es ihm jemand zu lesen versprochen hat* ‘that someone promised him to read it’ contains several “discontinuities”. *es* ‘it’ is dependent on the verb *zu lesen* ‘to read’ but

is separated from it by the sequence of pronouns *ihm jemand*. *ihm* ‘him’ is dependent on *versprochen* ‘promised’ but is separated from it by the sequence *jemand zu lesen*. *jemand* ‘someone’ is dependent on *hat* ‘has’ but is separated from it by the sequence *zu lesen versprochen*. Relying on intuitions from English we would expect *es* and *zu lesen* to be sisters in a VP, *ihm* and *versprochen* to be sisters in a VP and *jemand* and the VP that *hat* heads to be sisters in an S. However, there is no way to arrange a phrase structure tree so that these sisterhood relations are manifest and the terminal yield of the tree is *es ihm jemand zu lesen versprochen hat*.

If we allow tree branches to “cross” then we could achieve the correct sisterhood relations with the right terminal yield as in Fig. 1-2.

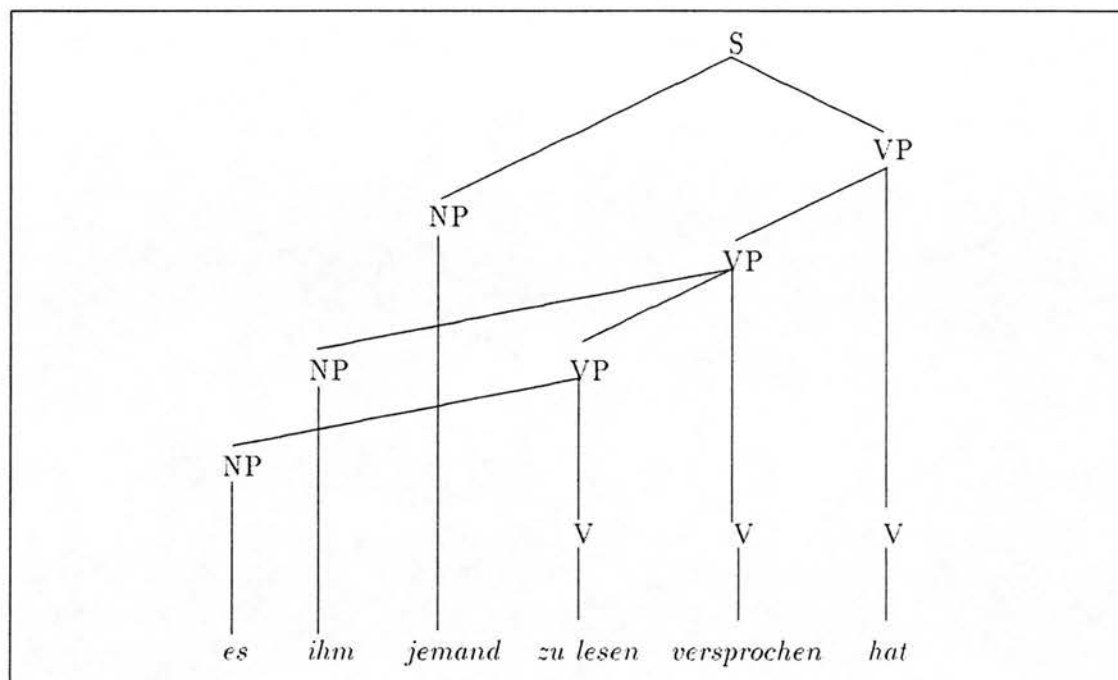


Figure 1-2: Discontinuous tree for *es ihm jemand zu lesen versprochen hat*

Now, all of the sisterhood relations stated above are met and the terminal string is generated correctly (again by reading the leaves of the tree from left to right). Furthermore, the VP *es zu lesen* is a sister to *versprochen* and the VP *es ihm zu lesen versprochen* is a sister to *hat* as we expect from the phrase structure for *someone promised him to read it*.

Such “discontinuous” trees have been proposed for both German and Dutch examples. Fig. 1-3 contains the discontinuous tree for the Dutch subordinate clause *dat Jan Piet Marie die nijlpaarden zag helpen voeren* ‘that Jan saw Pete help Marie feed the hippos’.

We will consider a slightly simpler example however. Fig. 1-4 contains the discontinuous tree for the Dutch subordinate clause *dat Jan Piet Marie zag helpen*

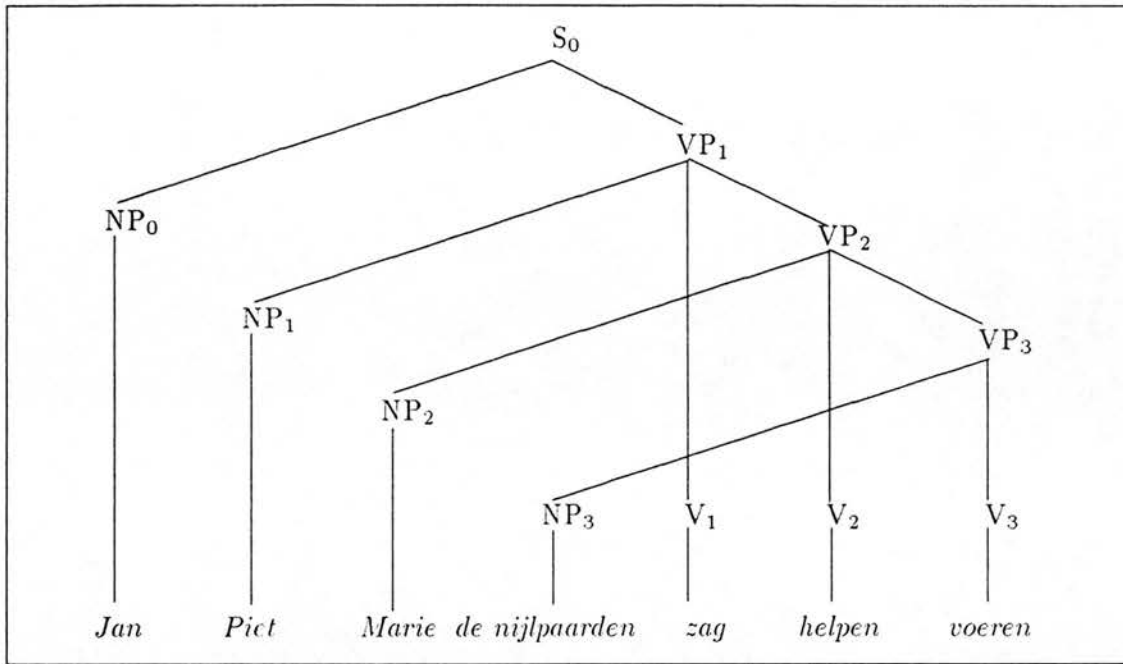


Figure 1-3: Discontinuous tree for *Jan Piet Marie die nijlpaarden zag helpen voeren*

zwemmen ‘that Jan saw Pete help Marie swim’.

Most syntacticians however prefer to handle such cases of “apparent” discontinuity by hypothesising phrase structure trees which have the right terminal yield but in which the tree structure does not correspond to intuitive subcategorisation requirements for VPs. Fig. 1-5 is a tree in the style of Evers ([13]) for *dat Jan Piet Marie zag helpen zwemmen*, Fig. 1-6 is a tree where the VP consists of recursive \bar{V} projections and Fig. 1-7 is a tree in the style of Bresnan, Kaplan, Peters and Zaenen ([5]), henceforth BKPZ.

The trees in both Fig. 1-5 and Fig. 1-6 contain all three NPs as sisters to the VP in contrast to the discontinuous tree in Fig. 1-4 where they are each sister to a different verb projection. The tree in Fig. 1-5 has all the sisters occurring as sisters under the VP while the tree in Fig. 1-6 structures the cluster of verbs into a recursive \bar{V} projection. The tree in Fig. 1-7 is even more radical. The matrix VP is structured into a recursive VP projection while all of the NPs comprise a recursive NP projection.

This thesis rejects both approaches to trying to “save” the phrase structure based account of word order. Instead, we argue that there is a level of syntactic-semantic *functor-argument structure* which is motivated independently of word order. For example, the functor-argument structure for *Jan Piet Marie zag helpen zwemmen* is given in Fig. 1-8.

The first point to note is that functor-argument trees are unordered. Sisters of a subtree may be permuted arbitrarily and all the permutations of any given

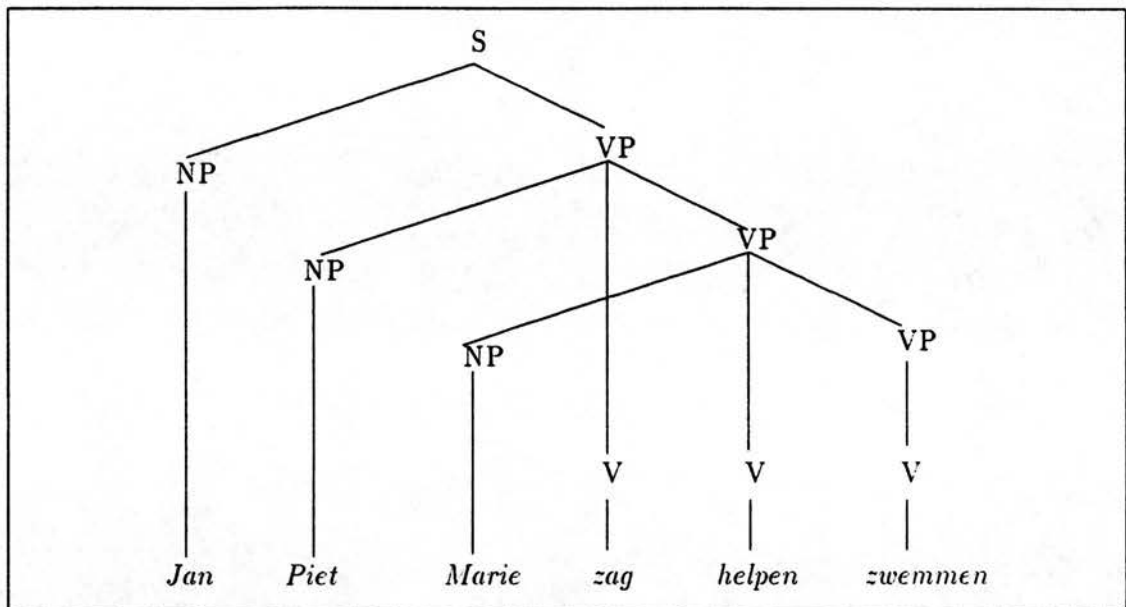


Figure 1-4: Discontinuous tree for *Jan Piet Marie zag helpen zwemmen*

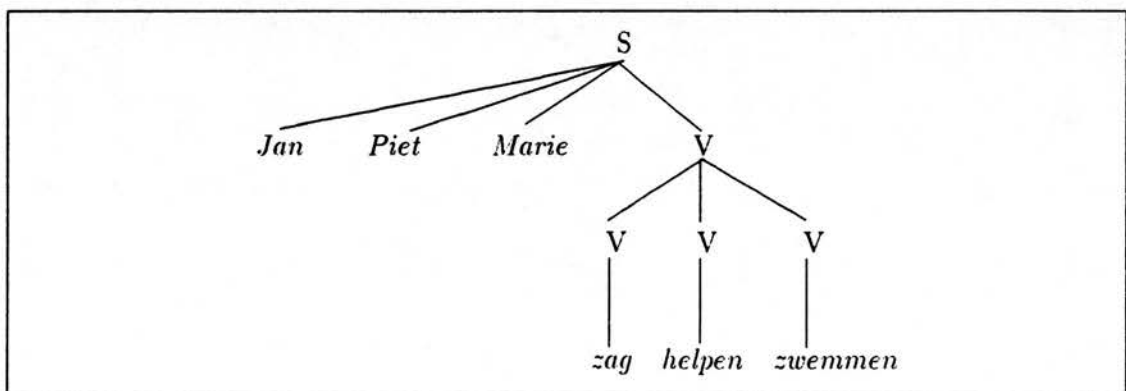


Figure 1-5: Evers style tree for *Jan Piet Marie zag helpen zwemmen*

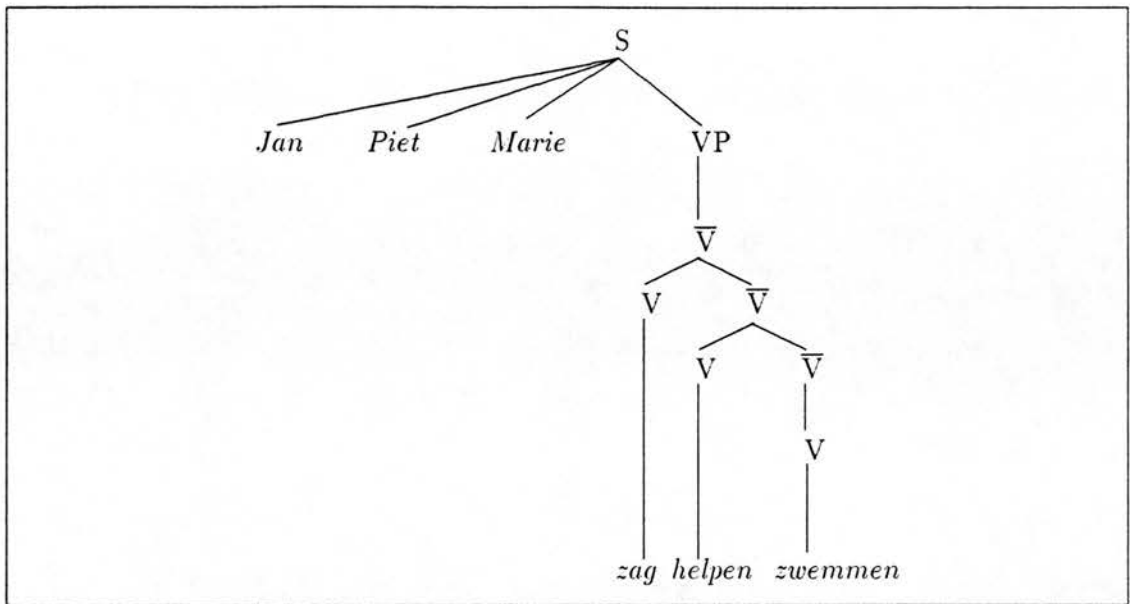


Figure 1-6: Recursive \bar{V} tree for *Jan Piet Marie zag helpen zwemmen*

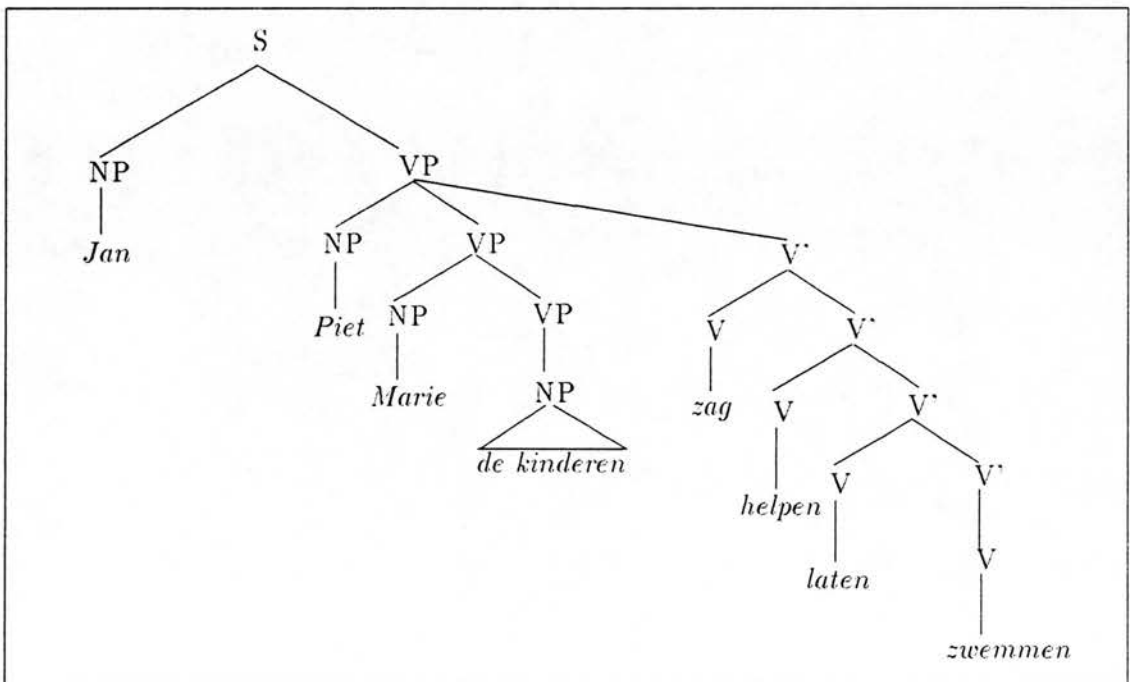


Figure 1-7: BKPZ style tree for *Jan Piet Marie zag helpen laten zwemmen*

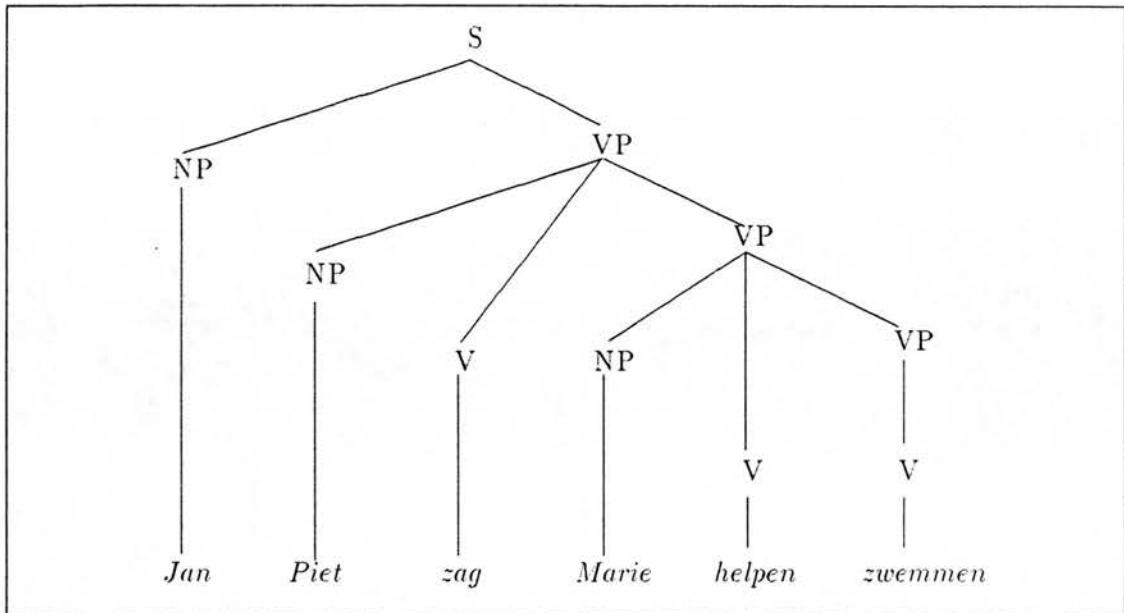


Figure 1-8: Functor-argument tree for *Jan Piet Marie zag helpen zwemmen*

functor-argument structure are equivalent. So for example, the functor-argument tree in Fig. 1-9 is equivalent to the functor-argument tree in Fig. 1-8. (Strictly speaking this means that functor-argument structures are not trees.)

In the tree in Fig. 1-9, the lowest VP dominates a V which dominates *zwemmen*. The second lowest VP dominates the head verb *helpen*. It also dominates the complement NP *Marie* of *helpen* and the complement VP *zwemmen* of *helpen*. The highest VP dominates the head of the VP *zag*. It also dominates the NP complement *Piet* of *zag* and the VP complement *helpen Marie zwemmen* of *zag*. Finally, the highest VP is sister to the NP subject *Jan* and both the highest VP and the NP subject are dominated by S.

Thus, all complements are sisters to their heads and the expected recursive VP structure is evident. This means for example that verbs can assign the morphological form of their VP complements by subcategorisation and simple English-like syntactic rules and that the verbal morphology can be transmitted to the head verbs by something like the Head-Feature Principle of GPSG or the Status Government Principle of Evers ([13]). Notice that functor-argument structures are very much like the *D*-structures of GB except that they are unordered.

Notice that there is a quite direct motivation for functor-argument structures like those in Fig. 1-8 and Fig. 1-9. First, complements are sisters to their heads as required, for example, by the principles of \bar{X} theory. Second, semantic compositionality is straightforward again because complements are sisters to their heads. In the sequel, we shall generalise head-complement structures to functor-argument structures in general, thus the name *functor-argument structure*.

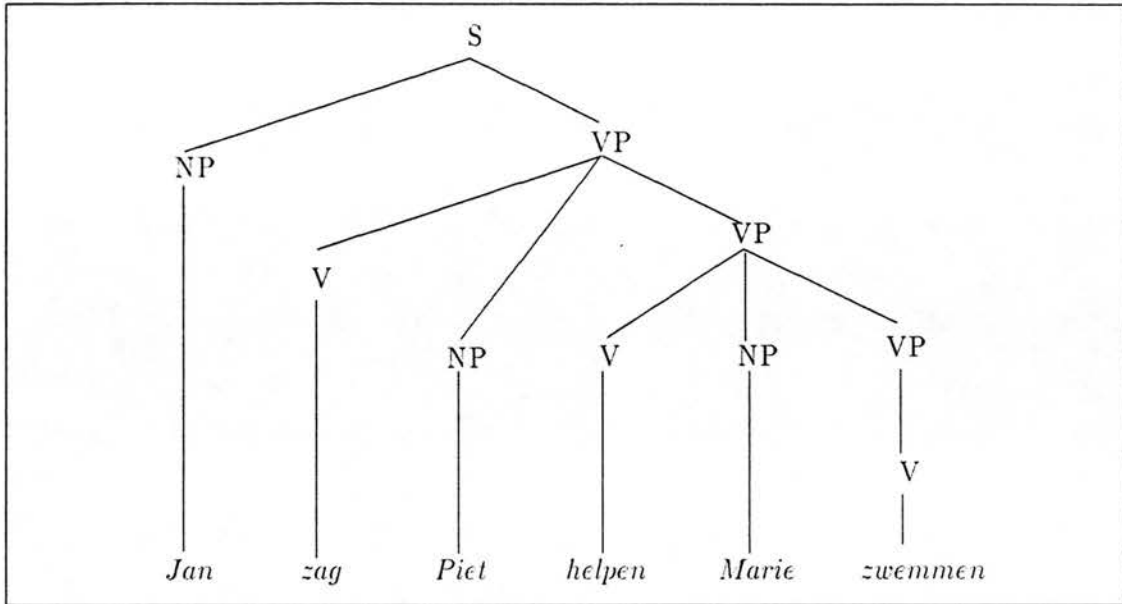


Figure 1-9: A second functor-argument tree for *Jan Piet Marie zag helpen zwemmen*

Given such functor-argument structures, the question which remains is how the correct word order is generated. The answer proposed in this thesis is that word order is determined by *word order domains*. The word order domain of a lexical item is the lexical item itself. The domain of a phrasal constituent is formed from its daughter's domains in one of two ways. Either the domains of a daughter are concatenated in some order which is specified by independent *linear precedence constraints* or *linear precedence statements* (LP statements) or the domains of the daughters are merged or *domain unioned* in such a way that order of the constituent domains is kept in the phrasal domain. Fig. 1-10 presents the *domain tree* for *Jan Piet Marie zag helpen zwemmen*. (Only the phrasal domains are indicated.)

The domain of the second lowest VP is

$$[\text{VP} [\text{NP Marie}] [\text{v helpen}] [\text{v zwemmen}]]$$

It is formed from the concatenation of the three lexical domains $[\text{NP Marie}]$, $[\text{v helpen}]$ and $[\text{v zwemmen}]$. The domain of the highest VP is

$$[\text{VP} [\text{NP Piet}] [\text{NP Marie}] [\text{v zag}] [\text{v helpen}] [\text{VP zwemmen}]]$$

It is formed by merging the lexical domains $[\text{NP Piet}]$ and $[\text{v zag}]$ with the domain of *Marie helpen zwemmen*. Notice that merging allows the constituent domains to appear discontinuously in the phrasal domain. In this example, *Piet* is separated from *zag* to which it is a sister in the functor-argument tree. Furthermore, *Marie*

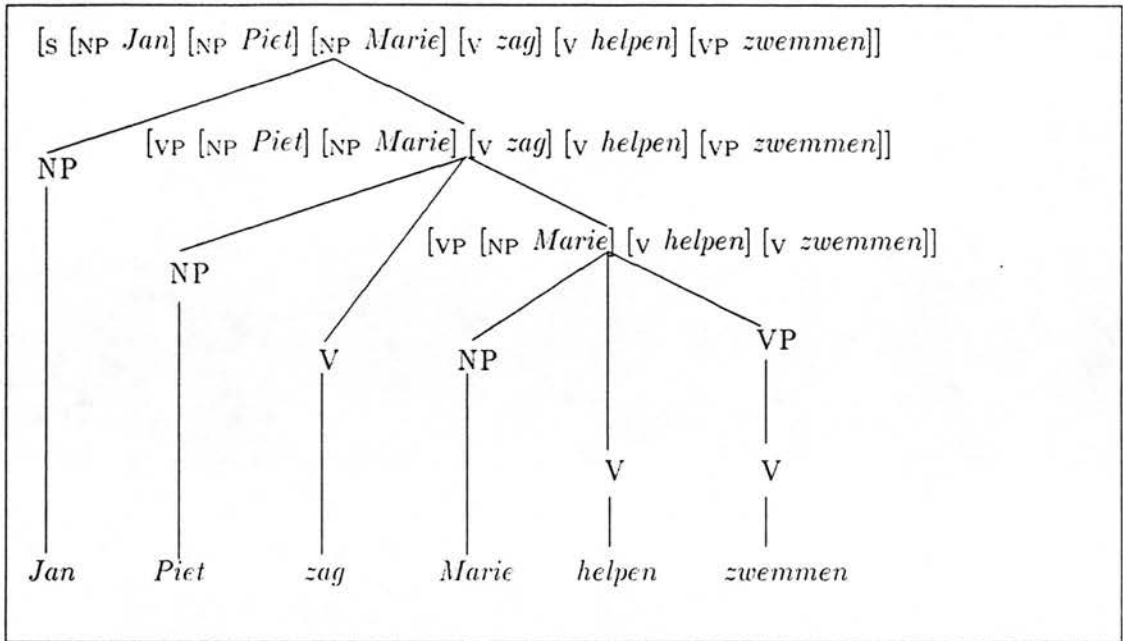


Figure 1-10: Domain tree for *Jan Piet Marie zag helpen zwemmen*

is separated from *helpen* to which it is a sister in the functor-argument tree. Precisely what the constraints are on this merging operation will be discussed in more detail in following sections and chapters. Finally, the domain of the S is

[S [NP *Jan*] [NP *Piet*] [NP *Marie*] [V *zag*] [V *helpen*] [VP *zwemmen*]]

It is formed by merging the lexical domain [NP *Jan*] with the domain of *Piet Marie zag helpen zwemmen*. In this case the domain of *Jan* simply appears in initial position in the domain of the S.

Quite often, languages which exhibit the type of discontinuous constituency discussed in this section also exhibit semi-free word order whose scope corresponds to the merged domains. For example, if we reconsider the German example, *es ihm jemand zu lesen versprochen hat*, we can assign it the schematic pattern

[S NP₃ NP₂ NP₁ V₃ V₂ V₁]

where the subscripts indicate constituent dependency. Now for German we can substitute other NPs such that all of the following orders are obtainable.

[S NP₁ NP₂ NP₃ V₃ V₂ V₁]
 [S NP₁ NP₃ NP₂ V₃ V₂ V₁]
 [S NP₂ NP₁ NP₃ V₃ V₂ V₁]
 [S NP₂ NP₃ NP₁ V₃ V₂ V₁]
 [S NP₃ NP₁ NP₂ V₃ V₂ V₁]
 [S NP₃ NP₂ NP₁ V₃ V₂ V₁]

Of course, all such orders are subject to word ordering preferences. The important thing to note however is that the domain of word order variation in such languages lines up exactly with the domain of apparent bounded discontinuous constituency. It is these issues which we will address in the rest of this thesis.

1-1.2 Cross-linguistic variation in word order

Our approach to cross-linguistic variation in word order is based on the assumption of principles of universal grammar (and not parameters as additionally in GB), principles of sets of languages (such as diachronically clearly genetically related languages such as Dutch and German) and principles of particular languages, dialects and idiolects. For example, the principles governing domain construction can be hypothesised to be linguistic universals which define a space of possibilities within which language-specific principles pick out various possibilities.

This abstract statement can be illustrated in terms of our Dutch and German examples. In both examples, all of the NPs in the sentential domains precede all of the Vs in the Mittelfeld. Thus, both Standard German and Standard Dutch require that NPs precede Vs in the Mittelfeld. This leaves open the question of whether verbs can ever precede NPs in the Mittelfeld in West Germanic languages (excluding the verb-initial language English). The answer to this question is yes. As we will see in Ch. 5, Zurich German allows Vs to precede NPs as long as NPs precede the verbs they depend on. We might further ask if there are languages which allow verbs to precede the NPs which depend on them. Again, the answer is yes. Some Zurich German speakers allow the unstressed causative *la* to precede its complement NP. Furthermore, English of course is verb initial and the verb precedes its complements, both NPs and all other complements.

The next question we might ask considers the ordering of the verbs. From our two examples, we see that German verbs canonically “govern” governed verbs to the left and Dutch verbs canonically govern to the right. So, there is a clear choice between left versus right direction of government. Are other possibilities possible? Again, the answer is yes. In the *Ersatzinfinitiv* in German, the government order is 1-3-2. Similarly, in Dutch, the order 2-1 is possible for some speakers when the V_1 is a finite auxiliary and V_2 is a past participle. Both of these possibilities have the property however that each verb governs in one direction every verb which it governs. Further investigation in Ch. 4 will show that this is always the case. Does this property hold universally though? The answer appears to be no. Again, Zurich German allows verbs within a single domain to appear in any order with respect to each other.

Thus, we see that within the space defined by the domain construction principles (to be discussed further in §1.2) there are a number of “decisions” that a language can make which further constrains the word order possibilities. The identification of these decisions is the approach which we will take to cross-linguistic variation in related languages.

1-1.3 Formalisation in a homogeneous logical formalism

There are several reasons for formalising a treatment of bounded discontinuous constituency and semi-free word order in a single homogeneous logical formalism. First, it is desirable to formalise all of linguistic theory within a single formalism. This makes it possible to integrate all the different aspects of grammar. This has the advantage that proving results about the interaction of different components of grammar is greatly simplified. It also means that a unified proof theory can be applied to the entire grammar, rather than applying feature-value logic to complex category symbols, tree domain logic to syntactic descriptions and the proof theory of higher order logic to semantic translations, for example, as is necessary in GPSG.

Second, if the assumption that precise formalisation of linguistic theory is a necessary precondition to progress, then our linguistic descriptions should be written in a logic with a rigorously defined semantics and a sound and, hopefully, complete proof theory. In Appendix A, we present a logic, \mathcal{L}^+ with a rigorously defined semantics and a sound and complete proof theory. The importance of the rigorously defined semantics is to avoid the confusion and misunderstanding which arises from uninterpreted notations. The importance of a sound proof theory is to ensure that inferences which we draw from our theories are in fact semantically true. The importance of a complete proof theory is to be able to detect when our theories (either of universal grammar or of a single language or set of languages) are inconsistent. Without completeness, we cannot guarantee that we can determine if our theories are consistent or inconsistent.

Third, and possibly most importantly for our work here, the availability of a homogeneous logical formalism allows us to escape the “tyranny of trees”. As long as syntactic theory and the theory of word order is based on tree domains, then the ability to state more powerful and more unconventional generalisations about possible word order “domains” remains unavailable. However, if syntactic description is described in the same way as any other aspect of linguistic structure then we can investigate theories of syntax and word order which do not rely on conventional assumptions of adjacency, unique dominance, etc.

A major part of this thesis then has been the formalisation of a homogeneous logic which is expressively powerful enough to allow the description of all levels of linguistic structure including, but not limited to, semantics, syntax and word order. The formalism is very similar to that used in P&S1 ([31]) and P&S2 ([32]). However, whereas functional and relational dependencies remain largely unformalised in those two works, they have been given a rigorous semantics which I believe corresponds to the intuitive use of these two devices and a corresponding logic has been found. The main technical strategy has been to treat feature-value logic (even with function and relation symbols) as hybrid polymodal logics. The results place the formalism of P&S1 firmly within the class of modal logics.

The entirety of this work has been placed in an appendix (App. A) to prevent disruption of the development of the linguistic work which makes up the main

body of the thesis. Readers should refer to the appendix if there are any aspects of the use of the formalism which are not immediately apparent.

1-1.4 Organisation of the Thesis

The thesis contains six chapters and three appendices. The rest of Ch. 1 consists of several parts. §1.2 is an introduction to the basic concept of *word order domain* on which the rest of the thesis rests. §1.3 examines the conventional role of syntax in linguistic theory and its relation to the theory of word order. §1.4 presents arguments for rejecting the conventional role of syntax in linguistic theory. §1.5 discusses issues of generalisation, theoretical prediction and formalisation given our methodological assumptions. §1.6 discusses the common existence of certain syntactic categories in West Germanic. §1.7 discusses the role of diachronic evidence for synchronic syntactic theory. §1.8 discusses the scientific methodology assumed in the thesis. This section is important for placing this work in relation to the rest of modern syntactic theory. §1.9 discusses the approach taken to cross-linguistic variation taken in this thesis in more detail. Finally, §1.10 presents an example derivation of a German subordinate clause which manifests multiple crossing dependencies.

§2-1 – §(2-8) of Ch. 2 examine attempts in the past to deal with discontinuous constituency and (partially) free word order. Problems of inadequacy are discussed for each of the approaches and an attempt is made to show the chronological development of each approach from previous work. The seminal pre-theoretical work of Gunnar Bech is discussed briefly in §2-7 and Evers' work ([13]) which follows Bech's tradition is discussed in §2-8. These two works are especially relevant to the account presented in the thesis. §2-9 formally introduces the central notion of *domain union* on which our account is based. §2-10 then goes on to present a theory of linear precedence constraint preferences. §2-11 discusses the work of David Dowty ([12]) which is similar to the approach presented here.

Ch. 3 presents a formalisation of the account of the German data and also discusses the formalisation of semantic interpretation, semantic representations, the syntax and semantics of adjuncts, the syntax and semantics of specifiers, the syntax and semantics of quantifiers, the scope of quantification, the scope of modification and raising.

Ch. 4 presents an account of basic word order data in German. It also discusses in some detail related topics including extraposition, subjectless constructions, pronominalisation, scrambling, reflexivisation, the scope of adjuncts and negation, topicalisation and extraposition, intraposition of VPs in the Mittelfeld, non-ergative subjects and topicalisation, argument inheritance and raising.

§5.1 first presents a comparative account of basic Dutch word order. It is shown that the grammars of German and Dutch are very similar and that the grammar of Dutch and English deviate only in minor detail. Some data from the Gronin-

gen dialect of Dutch is also considered which again shows that the differences between the grammars of Standard Dutch and the Groningen dialect are very minor although the superficial differences are rather striking. §5.2 first gives a basic account of the Zurich German dialect and again shows that the differences in the grammars of Standard German and Zurich German are very minor indeed although the superficial differences are quite striking. §5.2 then goes on to consider some puzzling data concerning the infinitival particle *z*. Finally, §5.3 discusses the comparative clause structure of German, English and Dutch and shows that the differences are rather minor.

Finally, Ch. 6 discusses various issues which arise during the thesis but which are not discussed extensively in the other chapters.

Appendix A is

1–2 Introduction

Nearly all modern grammatical theories derive word order from the terminal yield of phrase structure trees. This includes theories as disparate as **GB**, **LFG** ([4]), and **GPSG** ([14]). Another way to put this is to say that the word order domain of a constituent is the sequence of the leaves of its constituent structure tree. Therefore, any attempt to explain apparent cases of discontinuous constituency requires that “continuous” surface phrase structure trees be assigned whose terminal yield exhibits the apparent discontinuity. In **GB**, this is done via movement rules which “reorder” *D*-structure trees into *S*-structure trees. In **LFG** and **GPSG** this is done by assigning phrase structure trees to strings which do not necessarily correspond to intuitively-motivated subcategorisation requirements, e.g., Dutch and German control constructions may not include any controlled, infinitival VPs at the level of syntax. This can make the task of giving an interpretive, compositional semantics very difficult. The problem of discontinuous constituency has prompted numerous proposed solutions within and without the **GB** community. In nearly all cases, these proposals either employ operations on trees (e.g., clause union or movement) or a redefinition of the notion of tree (e.g., “tangled” or “discontinuous” trees). That is, the strategy is to somehow produce a tree whose terminal string exhibits the apparent discontinuity. This strategy makes sense as long as one is committed to *surface syntax* as the basis of word order. This has of course been the case in most linguistic theory since the publication of *Syntactic Structures* ([6]).

There is another approach however which is represented to some degree by the dependency grammar tradition and **HPSG** ([31]) and, to a lesser extent, by categorial grammar. This approach denies the existence of, or at least reduces the importance of, an independent level of surface syntactic structure and its role in determining word order. However, most versions of dependency grammar and categorial grammar require a *strict adjacency condition* on strings which is consistent with the phrase structure tradition. That is, the string derived by a phrasal

constituent is still the left-to-right concatenation of the strings of its ordered daughters. To my knowledge, there are only two other approaches which allow word order to be derived from syntactic structure without necessarily requiring the adjacency condition. In theory, this is possible in HPSG since daughters are not ordered and the *Constituent Ordering Principle* allows an arbitrary mapping from syntactic structure to phonological theory in principle. So far, this does not seem to have been taken seriously by the HPSG community yet. The second approach is that taken by David Dowty ([12]). It is similar in many ways to the approach adopted here. I will discuss the relationship between his approach and the one to be advanced here in §2-11 and Ch. 5.¹

I will present an approach which rejects surface syntax and its role in determining word order. In its simplest, most general form, the approach claims that

1. phrasal word order is determined within locally definable *word order domains* which are ordered sequences of constituents,
2. phrasal word order domains are composed compositionally from their daughter word order domains,
3. lexical entries do not have word order domains,
4. the *functor* of a phrasal constituent is an element of its mother's domain and
5. either
 - (a) a nonfunctor daughter is an element of its mother's domain or
 - (b) the elements of a nonfunctor daughter's domain are elements of its mother's domain and furthermore they may appear discontinuously or nonadjacently in the mother's domain provided the relative order of the elements of the daughter domain are preserved in the mother's domain.

When the last option is chosen, we say that the daughter domain has been *domain unioned* into its mother's domain. We can also speak of two or more domains being domain unioned together.

1-3 The Role of Syntax in Linguistic Theory

In this section, I will discuss various aspects of the role of syntax in modern linguistic theory. The position I take is that the theory of syntax and the theory of word order have been incorrectly conflated with damaging consequences for both the theory of syntax and the theory of word order.

¹His research and mine were developed independently without knowledge of the other's work.

1-3.1 Theoretical nonobservables and modelling of empirical domains

One question which any scientific theory must periodically address is whether its stock of nonobservable theoretical constructs is still justifiable. My claim here is that syntax as it is usually formulated is no longer justifiable by the empirical evidence. Pollard and Sag discuss the role as the least justifiable linguistic theoretic construct very elegantly in P&S2. Therefore, I will quote them on this topic ([32, pp7-8]).

“A further methodological principle, shared by the scientific community at large, is that of ontological parsimony: insofar as it is possible without doing violence to the simplicity and elegance of the theory, we do not posit constructs that do not correspond to observables of the empirical domain. Of course, all scientific theories contain such constructs. An obsolete example is the phlogiston that used to form the basis for the theory of combustion; a contemporary one is the quarks that are posited to account for the observed variety of subatomic particles. But the parsimony principle with respect to nonobservable constructs dictates: use only as needed. Perhaps phrase structure itself (variously manifested as, e.g. GB’s S-Structure, LFG’s c-structure, and HPSG’s DAUGHTERS attribute) is the nonobservable linguistic construct that enjoys the widest acceptance in current theoretical work. Surely the evidence for it is far less direct, robust, and compelling than that for phonological structure (e.g. GB’s PF, HPSG’s PHONOLOGY), logical predicate-argument structure (GB’s LF, HPSG’s CONTENT), or underlying grammatical relations (GB’s D-structure, HPSG’s SUBCATEGORISATION attribute, LFG’s f-structure). But for all that a theory that successfully dispensed with a notion of surface constituent structure is to be preferred (other things being equal, of course), the explanatory nature of such a notion is too great for many syntacticians to be willing to relinquish.”

^aNotable exceptions in this respect are Hudson’s ... word grammar, and certain varieties of categorial grammar wherein the particular order in which lexical items are assembled into larger units is viewed as an epiphenomenon lacking in linguistic significance.

1-3.2 The Role of Constituent Structure

Dowty ([12]) presents an appraisal of the role of constituent structure in his paper *Towards a Minimalist Theory of Syntactic Structure* which bears directly on the issues discussed here. I present a quote from it now.

“There are two things that worry me about the situation syntax finds itself in in 1989. Since hierarchical syntactic structure is so often assumed, syntacticians don’t usually ask questions – at least beyond the elementary syntax course – as to what the nature of evidence

for a constituent structure in a particular sentence in a particular language is: we just take whatever structure our favorite syntactic theory would predict as the expected one for the string of words in questions – by current X-bar theory, etc. – unless and until that assumption is contradicted by some particular fact.

My second concern is closely related: I suspect syntacticians today have almost come to think of the “primary empirical data” of syntactic research as phrase structure trees, so firm are our convictions as to what the right S-structure tree for most any given sentence is. But speakers of natural languages do not speak trees, nor do they write trees on paper when they communicate. The primary data of syntax are of course only *strings* of words, and everything else in syntactic description beyond that is part of a theory, invented by a linguist.”

Although Dowty’s appraisal is very clear, I want to emphasise the two points that he makes. First, syntacticians take phrase structure as the basis for determining word order, and, word order as the basis of determining phrase structure, for granted. The phrase structure approach is not God given. Rather it is a theoretical construct which is subject to scrutiny any time there is evidence to suggest that things are amiss. There is plenty of evidence to suggest that phrase structure constituency as the basis of word order and syntax is clearly inadequate. I repeat Dowty’s partial list of examples here.

1. VSO languages
2. Cases for which “Wrapping” operations (Bach 1980, Pollard 1984) have been proposed
3. Free word order and free constituent order languages
4. Various other instances of proposed “flat” structures
5. Clause unions
6. Extrapositions
7. Parenthetical phrases

Second, syntacticians do seem to take phrase structure trees as the primary data. It is not uncommon to hear a syntactician say that a particular tree structure is “unintuitive” or that the approach that we present here or Dowty’s approach is “intuitively unmotivated”. However, trees are theoretical constructs. We do not speak them. We do not hear them. They simply are not in the empirical domain and so are not the type of entity about which we can have intuitions. More generally, we cannot reject proposed theories on the basis of intuitions. We can only reject theories on the basis of intuitions concerning the predictions made by those theories. But even then, such predictions should be checked with the empirical evidence.

1-3.3 Tree Hacking

Much of what passes for syntactic research in modern linguistics is really research on word order. Because standard syntactic theory assumes that word order is determined by phrase structure trees, problems of word order end up becoming problems of syntax. That is, some plausible phrase structure (i.e., syntactic) tree must be found which produces the right word order. I would say that the result is that the theory of syntax has become infested with analyses involving highly unmotivated phrase structures. One could even go so far as to say that syntacticians are “hacking” the trees to get the word order right. However, syntactic structure should be formulated independently of word order and then the relation between the two investigated. As in *D*-structure in GB it is usually fairly clear what the syntactic structure should look like. I claim that everything else is a question of word order.

This requirement to find the right phrase structure tree is often made at the expense of other aspects of grammatical theory. To take a single example, consider once again the tree structure in Fig. 1-11 proposed by Bresnan, Kaplan, Peters and Zaenen for *(dat) Jan Piet Marie de kinderen zag helpen laten zwemmen*.

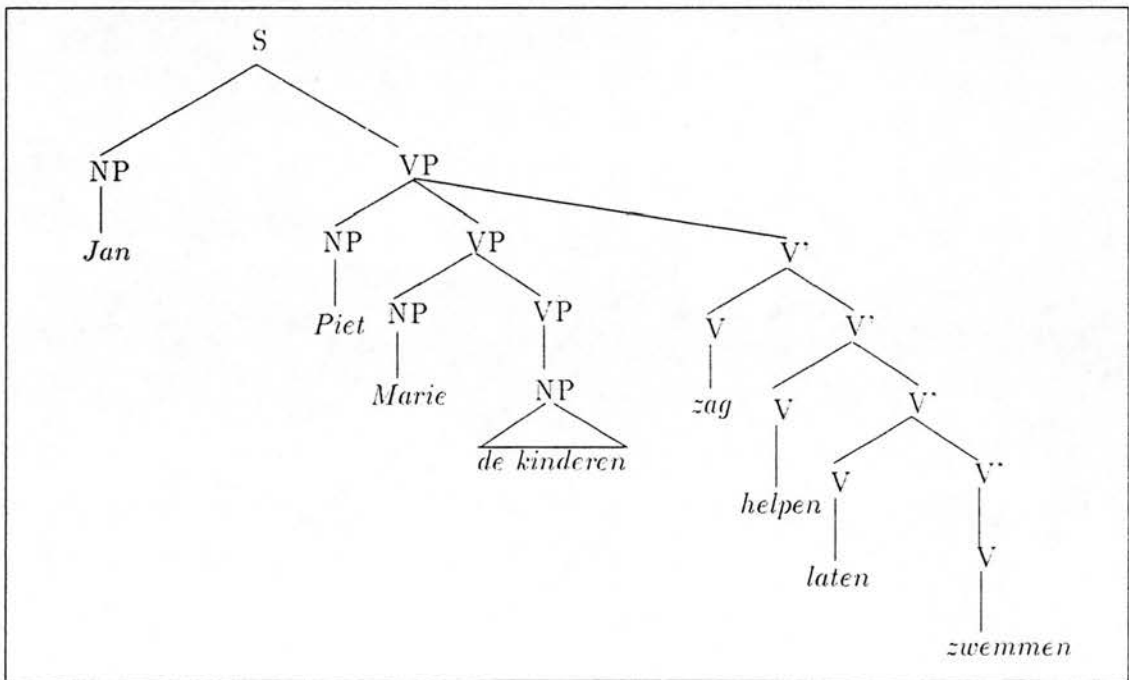


Figure 1-11: BKPZ style tree for *Jan Piet Marie de kinderen zag helpen laten zwemmen*

I would claim that the only motivation for this structure is to get the word order facts right. A full description of the account of [5] is inappropriate here but I will indicate a few suspect aspects of the analysis. First, let's consider the structure of the VPs which comprise a recursive VP structure. Except for the topmost VP, none of the VPs contain their head verb. Furthermore, none of the NP

complements (direct objects), i.e., either *Piet* or *Marie* appear as sisters to the verbs which they depend on. More spectacularly, the NPs are all arranged in a recursive VP structure in which there are no verbs (except for the topmost NP).

This has a number of immediate predictions. Since the level of recursion is theoretically unbounded, we should expect it to be possible to find sequences of the type VP^n for $n \geq 1$. It should then be possible to topicalise such VP structures since VPs are topicalisable in general. That is, we should expect V_2 clauses of the form in (1.1) to be found.

- (1.1) a *Piet Marie zag Jan helpen zwemmen
 b Marie zag Jan Piet helpen zwemmen

(1.1a) should be grammatical since *Piet Marie* is an NP. (1.1b) is grammatical.

Coordination data also indicates that the analysis is wrong. Because of the strict bifurcation between the “VP spine” and the “V’ spine”, there is no possibility that a suffix of the NP sequence and a prefix of the V’ sequence can be conjoined. However, this is in fact possible.

How did this state of affairs arise? In the early days of transformational grammar, transformations applied directly to strings. This was eventually modified so that transformations applied to trees and now we have the situation where a string is derived from *D*-structure by successive applications of move- α and adjunction. Thus, syntacticians were naturally led from treating strings as the primary data to treating trees as the primary data.

1–3.4 Tectogrammatic and Phenogrammatic Structure

Haskell B. Curry ([9]) introduced a distinction between two levels of structure: *tectogrammatical structure* and *phenogrammatical structure*. The distinction between these two levels of structure is precisely the distinction between our functor-argument structure and word order domain structure. Dowty ([12]) introduces this distinction for a very similar organisation of grammar. I quote his definitions here.

“... the sense of syntactic structure we have just been talking about is *tectogrammatical structure*: the steps by which a sentence is built up from its parts, but without regard to the actual form that these combinations of parts takes.”

“... *phenogrammatical structure*: *how* the words and phrases are combined, in what order, whether word order is fixed or free, whether inflectional morphology marks the syntactic organization or not, whether the tectogrammatic groups ... come out continuously or discontinuously in the sentence itself, and so on.”

Dowty then goes on to consider the necessity of this distinction.

“One might ask at this point whether it really matters whether phenogrammatical structure is string-like or not, as long as tectogrammatic constituent structure exists; doesn't the latter do everything we want phrase-markers to do anyway? The answer is, it would not matter at all, if all languages had fixed word order and purely concatenative syntactic formation rules: The phenogrammatic structure of every sentence like (2) would in that case be a straightforward mapping of the tree's leaves into a linear string. The problem, of course, is that languages do not always work this way: they have discontinuous word order, extraposition, and all the other phenomena in (1) that are problematic for the context-free phrase structure model. My claim is therefore that when one gets around to describing such phenomena, one may well find that they are better formulated in terms of syntactic operations applying to strings of words than to phrase markers.”

Of course, we formulate *phenogrammatic* structure in terms of word order domains and not simply strings. This is because strings do not contain enough “structural information” to allow the formulation of the *phenogrammatic* operations. Thus, word order domains are not *phonological* domains but something “inbetween” syntax and phonology. In fact, they are very similar to the *Phonological Form* (PF) of GB.

1-3.5 Functor-argument Structure

The treatment of functor-argument structure is obviously greatly influenced by bidirectional categorial grammar (BiCG). In pure BiCG, there are only two rules of syntax called *forward* and *backwards* application. They are of the following form.

$$\begin{aligned} X &\rightarrow X/Y Y \\ X &\rightarrow Y X \backslash Y \end{aligned}$$

In the first rule X/Y is the *functor* and Y is the *argument*. In the second rule, $X \backslash Y$ is the functor and Y is the argument. Notice that in both cases, X is specified in the functor and is also the mother in the rule. The slash (either forwards or backwards) indicates what the argument of the functor is. Multiple arguments are handled by nested slashes on the functor, e.g., $(S \backslash NP) / NP / NP$ for a ditransitive verb “looking for” two NP object arguments to the right and an NP subject argument to the left. Since we handle word order via word order domains these two rules could be simplified to be unidirectional with the following rule

$$X \rightarrow X|Y, Y$$

where the comma indicates that the rule only specifies dominance information and not precedence information.

Both bidirectional categorial grammar and unidirectional categorial grammar only allow *binary* combination. That is, a functor combines with only one of its arguments at a time. However, for technical convenience, it is necessary for us to allow simultaneous multiple argument combination. Assume that the feature **ARGS** encodes the list of arguments of a functor. Then all of our functor-argument rules are of the schematic form:

$$[\mathbf{ARGS} \langle Y_1, \dots, Y_m \rangle] \rightarrow [\mathbf{ARGS} \langle Y_1, \dots, Y_n \rangle], Y_{m+1}, \dots, Y_n$$

In fact, all our functor-argument rules are of the following schematic form. (In fact, the functor-argument rules are very similar to Rules 1–3 of HPSG.)

$$\begin{aligned} [\mathbf{ARGS} \langle \rangle] &\rightarrow [\mathbf{ARGS} \langle Y \rangle], Y \\ [\mathbf{ARGS} \langle Y_1 \rangle] &\rightarrow [\mathbf{ARGS} \langle Y_1, \dots, Y_n \rangle], Y_2, \dots, Y_n \\ [\mathbf{ARGS} \langle \rangle] &\rightarrow [\mathbf{ARGS} \langle Y_1, \dots, Y_n \rangle], Y_1, \dots, Y_n \end{aligned}$$

The first rule combines a functor with one argument with that argument to result in a functor with an empty argument sequence. The second rule combines a functor with all of its arguments except the first. The third rule combines a functor with all of its arguments.

Functor-argument structure is independently motivated with respect to both syntax and semantics in any framework. With respect to syntax, the functor, be it a determiner, adjunct or head, specifies the arguments that are lexically specified on the functor. This notion generalises subcategorisation, selection of heads by adjuncts and selection of heads by determiners, and specifiers more generally. Furthermore, functor-argument structure is stated in terms of basic or underlying argument requirements. For example, in the Dutch example (*dat*) *Jan Piet Marie zag helpen zwemmen*, functor-argument structure recognises that *helpen* subcategorises for an NP and a VP regardless of how these argument constituents get realised in the surface string.

With regard to semantics, the functor is the semantic functor over its arguments. For example, determiners are semantic functors over the head nouns they select for, adjuncts are semantic functors over the heads that they select for and heads are semantic functors over the complements they subcategorise for. This level of functor-argument *specification* exists independently whether it is taken advantage of in any particular framework or not.

The preceding discussion should sound like an argument for *D*-structure and in fact the parallels are very close. The major difference is that I assume that unbounded dependencies are base generated and that there are no movement rules such as *move- α* . Basically, the approach taken here is similar to a representationalist account of GB with *D*-structure, phonological form (PF) and base generation of unbounded dependencies but no *S*-structure.

In the context of this framework nearly all of syntax falls out directly from the functor-argument structure. Aside from the number of arguments which are saturated in any one rule, the primary role of the three schematic syntactic rules shown above is to simply saturate the argument specifications. Thus, the syntac-

tic component has been reduced to a minimum which is really mostly epiphenomenal. For reasons of correct domain construction, a strictly binary approach or an approach which always combines all arguments at once is not possible without greatly complicating the domain construction rules.

1-3.6 Compositionality

In this approach, semantic functor-argument structure and syntactic functor-argument structure are exactly parallel. X is a syntactic functor with syntactic arguments Y_1, \dots, Y_n if and only if X is a semantic functor with semantic arguments Y_1, \dots, Y_n . Thus there is no need for an interpretive semantics. Semantic compositionality is specified entirely in terms of functor-argument structure. This is strikingly different than those approaches which assign nonstandard phrase structure trees for discontinuous structures and then have to work very hard to assign a correct interpretive semantics.

1-3.7 Syntactic Categories

There are two kinds of syntactic information. The first is the conventional phrase structural notion of syntax where a subtree has syntactic category X . The second is the syntactic categories subcategorised for by a head. HPSG explicitly makes this distinction for example by allowing subcategorisation for elements which subcategorise for some element(s) while forbidding subcategorisation for an element which has a particular phrase structural configuration. The latter distinction is equivalent to treating subcategorisation information as part of the syntactic category information and HPSG again makes this explicit by making the subcategorisation feature `SUBCAT` a syntactic feature `SYN`. It is this level of syntactic information that we are only interested in. So, even if there is discontinuous constituency, we can still talk about syntactic categories as if there were none.

1-3.8 Unavoidable Syntax

Although nearly all syntax falls out of the functor-argument structure and the constraints on how many arguments are saturated at once, there are some elements of syntax of German and Dutch which appear to require the existence of independently stated rules. The primary rule is the rule which establishes the topic position in V2 clauses. One instance of this rule is the filler-gap rule which says that a clause can consist of a topic followed by an inverted clause containing a gap corresponding to the topic. This perhaps might be avoidable through type raising of some kind but this would not account for cases like *Es essen viele Leute Käse* where the topic is the expletive *es* which does not correspond to a gap in the inverted clause. Thus the topic position must arise by rule.

Another case where syntactic rules seem unavoidable is in cases where phrases

of various types can act as adjuncts and it is impossible to specify the adjunct behaviour on the lexical head of the phrasal adjunct. *that*-less relative clauses are of this type.

1-4 Justification for the Rejection of the Conventional Role of Syntax

In this section, I want to argue for the rejection of the conventional role of syntax in modern linguistic theory. The argument consists of three subsections.

First, I try to argue that the conventional role of phrase structure is empirically unmotivated and empirically inadequate. This is an enterprise that is fraught with problems. To do it properly, I would have to argue against every variant of every theoretical framework. Even then, there would always be a theoretician ready to come to the rescue of any given framework. Therefore, I will have to restrict my examinations to two representative examples of the two leading schools of linguistics: **GPSG**, as a leading proponent of monostratal, unification-based approaches to phrase structure grammar, and a conservative version of Government and Binding Theory, i.e., one which assumes *move- α* and adjunction and which rejects clause union, argument inheritance, representational approaches and the like. In the case of **GPSG**, we will have some success in showing empirical inadequacy and greater success in showing theory internal problems. In the case of **GB**, our criticism will be based on lack of empirical motivation and theory internal criticisms.

Second, I will argue that the conventional view of syntax is ripe for rejection since it is a purely theory internal construct and not an empirical one. That is, there is no necessary theoretical basis for the assumption of syntax whatsoever.

Third, I will argue that the functor-argument and word order domain structure approach is both empirically motivated and empirically adequate.

1-4.1 Empirical motivation and empirical adequacy in syntax

We will examine **GPSG** as an example of a monostratal, unification-based theory of grammar. As a monostratal theory based on phrase structure, it systematically fails to deal with domains of discontinuity and partially free word order. First, at a theory internal level it has problems dealing with any language which does not allow the construction of phrase structure trees where the verbs are sisters to the complements that depend on them. In such cases, the basis of the “shake ‘n’ bake” semantics is lost. This might be rescued by readopting the rule-to-rule hypothesis but this runs into problems as well as discussed below. These problems are basically a reflex of the fact that all such monostratal phrase structure approaches are anglocentric. English, of course, allows the assignment of such trees (if we

ignore such phenomena such as quantifier float, relative clause extraposition, relatively free placement of adverbials, etc.). Of course other structures can be assigned. For example, the tree in Fig. 1-12 is empirically adequate as a tree for *Jan Piet Marie zag helpen zwemmen*. (I.e., it derives the right word order.)

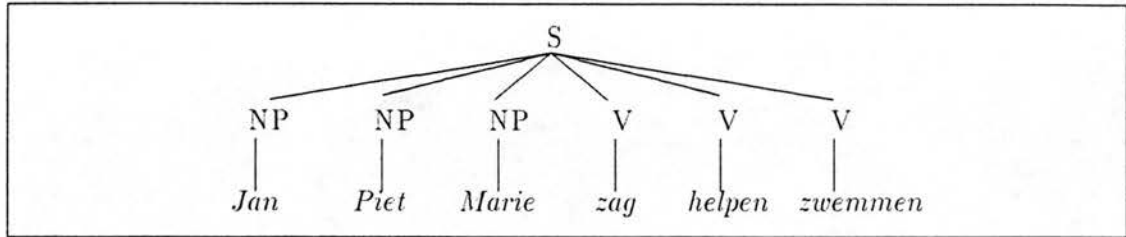


Figure 1-12: GPSG style “flat” tree for *Jan Piet Marie zag helpen zwemmen*

The problem is to explain how it is generated within the theory. The major problem is to coordinate the subcategorisation requirements of the verbs with the number and type of NPs in the clause. Obviously, simple rules of the form $S \rightarrow NP^* V^*$ are inadequate because too many or too few NPs can be generated with respect to the verbs in the verb sequence. Furthermore, there is no finite set of context-free phrase structure rules which can generate the correct structure.² In GPSG at least, there is only one way out. That is to use *binary* metarules which take two phrase structure rules as input and produce one output rule. There are several criticisms of such an approach however. First, in general, there will be multiple NPs and multiple verbs and the shake ‘n’ bake approach to semantics will fail. If the rule-to-rule hypothesis is used instead, then there will be appropriate semantic coindexing however. Secondly, however, the set of rules generated will be infinite thus violating the Finite Closure Property. This is almost sure to make the grammar non-context free, one of the major goals of GPSG.

A problem which is not theory internal but empirical is the use of ordering constraints. No set of GPSG style LP statements will guarantee that the verbs will appear in the (canonical) order that they do. (The problem of NP order is really one of semantic coindexation.) Because of the semantics of LP statements, verbs cannot be ordered with respect to verbs. Furthermore, the theory is too impoverished to provide any information (e.g., concerning verbal government) which might be used to order the verbs. That a finite auxiliary and a single past participle sometimes invert is no help since when there are three verbs they must come in the government order 1-2-3.

An even worse example for GPSG is the Swiss German subordinate clause *er wil*

²This might be made to work in LFG through the interaction of the *coherence* and *completeness* principles and *functional uncertainty* and such approaches have been proposed by Netter, Schuurman and Vilks among others. However, these accounts tend to overgenerate rather badly when a full complement of data is considered.

sini Chind la Mediziin schtudiere.

- (1.2) (das) er wil sini Chind la Mediziin schtudiere
 (that) he wants his child let medicine study
 ‘(that) he wants to let his child study medicine’

There are five additional permutations of this clause which can be characterised as follows: *er* is initial and NP complements precede the verb that subcategorised for them. Thus a flat structure is called for again as in the tree in Fig. 1-13.

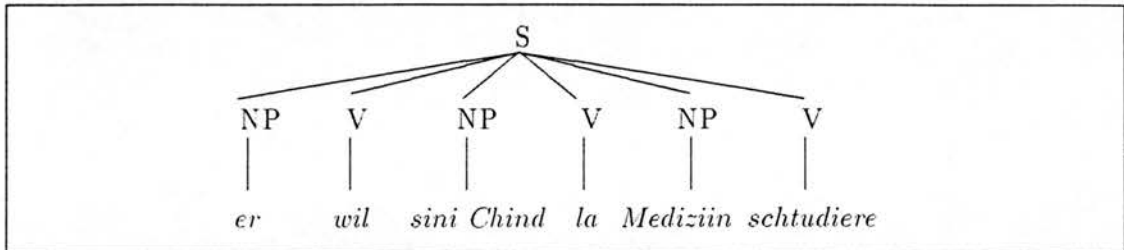


Figure 1-13: GPSG style “flat” tree for *er wil sini Chind la Mediziin schtudiere*

The problem for GPSG once again is to explain how the ordering generalisation concerning verbs and their dependents can be enforced. Again, the framework is too impoverished to allow this type of LP statement.

Any monostratal theory which proposes such flat structures will have similar problems. To summarise, the problems are of two types: first, the generation of the flat structures from a finite set of phrase structure rules, and second, the formulation of appropriate LP statements. If more hierarchical structures are proposed, then an adequate treatment of semantics is difficult, if not impossible to achieve. LFG analyses which make use of functional uncertainty are partly successful in this respect, but they typically overgenerate in systematic ways. It is of course, impossible to prove that the whole class of theories is empirically inadequate or riddled with theory internal problems, but the evidence suggests that all such theories are fundamentally inadequate. At a more basic level, we can make a similar observation. Monostratal phrase structure based accounts must be empirically inadequate because they do not consider domains of discontinuity and partially free word order and because they rely on adjacency of heads and complements (and more generally functors and arguments) to make semantic compositionality feasible.

GB begins with a *D*-structure which is similar to our functor-argument structure so it is empirically adequate at that level. For a sentence like *es ihm jemand zu lesen versprochen hat*, the problem is to explain the positions of *es* and *ihm* since they are not internal to the VPs which they originate in. One possible answer is that there are two NPs Chomsky adjoined to *S* which the NPs move into as in Fig. 1-14.

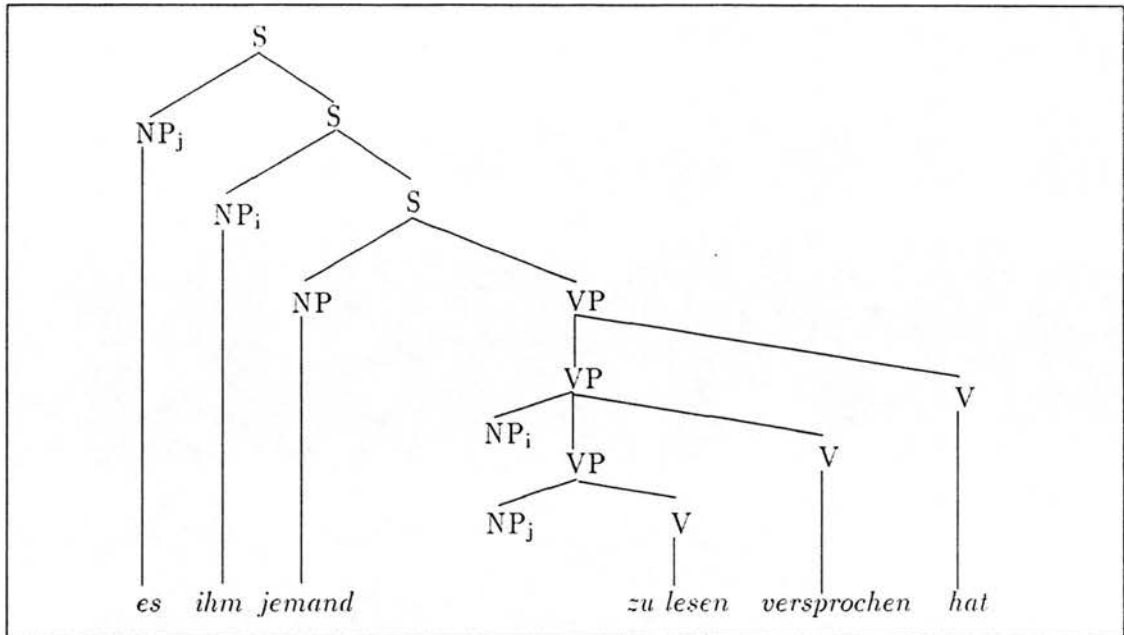


Figure 1-14: GB style scrambling tree for *es ihm jemand zu lesen versprochen hat*

The problem that faces GB on a theory internal level is what licenses the adjunction to S and the movement of the two NPs. The assumption in GB is that nothing moves in *D*-structure unless it has to. Barring an appeal to phonological requirements or some reconstruction of preferential ordering constraints it is not at all clear why the NPs should move *on strictly syntactic grounds* since other clauses with the same *D*-structure do not exhibit any movement at *S*-structure.

A related analysis is given to the clause *Jan Piet Marie zag helpen zwemmen*. In this case, the problem is to explain how the verbs *helpen* and *zwemmen* appear in the positions they do and not in the VP internal positions that they originate in in *D*-structure. The answer is that first there is a Chomsky adjoined \bar{V} position to the \bar{V} dominating *helpen* which *zwemmen* moves to. Then there is a Chomsky adjoined \bar{V} position to the \bar{V} dominating *zag* which the \bar{V} *helpen zwemmen* moves to as in Fig. 1-15.

Again, the problem that faces GB is what licenses the adjunctions to \bar{V} and the movement of the verbs to the two adjoined positions. This example is a bit more problematic than the last because presumably it would be possible to adjoin to the \bar{V} dominating *zag* first and then to move *zwemmen* to it followed by another level of adjunction \bar{V} to the \bar{V} dominating *zag zwemmen* and then moving *helpen* to it thus ultimately deriving the string *zag zwemmen helpen* which is ungrammatical.

To bring out the issues a bit more clearly consider the tree in Fig. 1-16.

In this case, there is no movement whatsoever and the *D*-structure is the same as the *S*-structure. The question this time is why do *ihm* and *das Buch* not move.

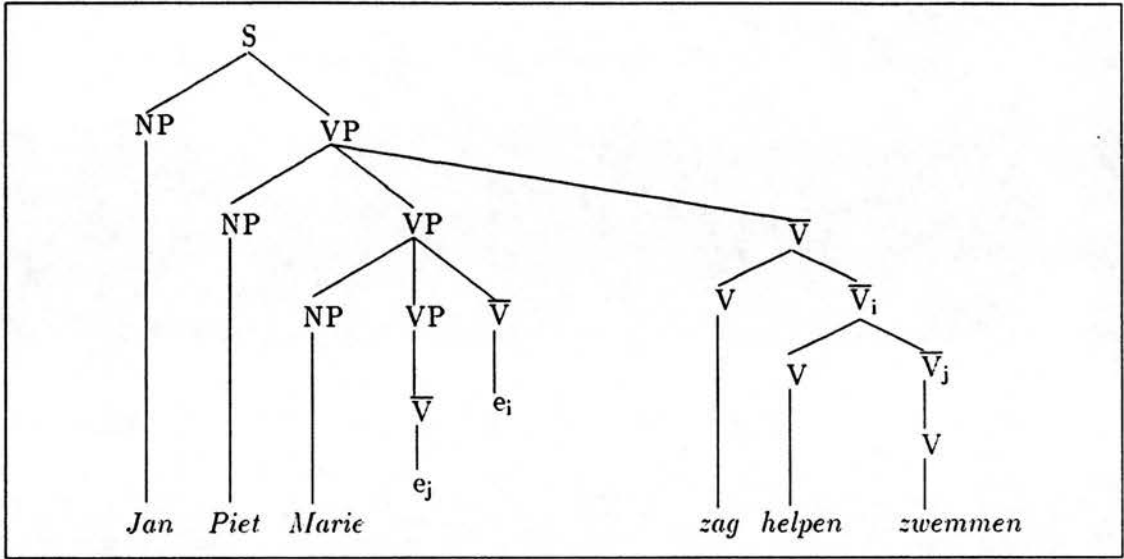


Figure 1-15: GB style verb raising tree for *Jan Piet Marie zag helpen zwemmen*

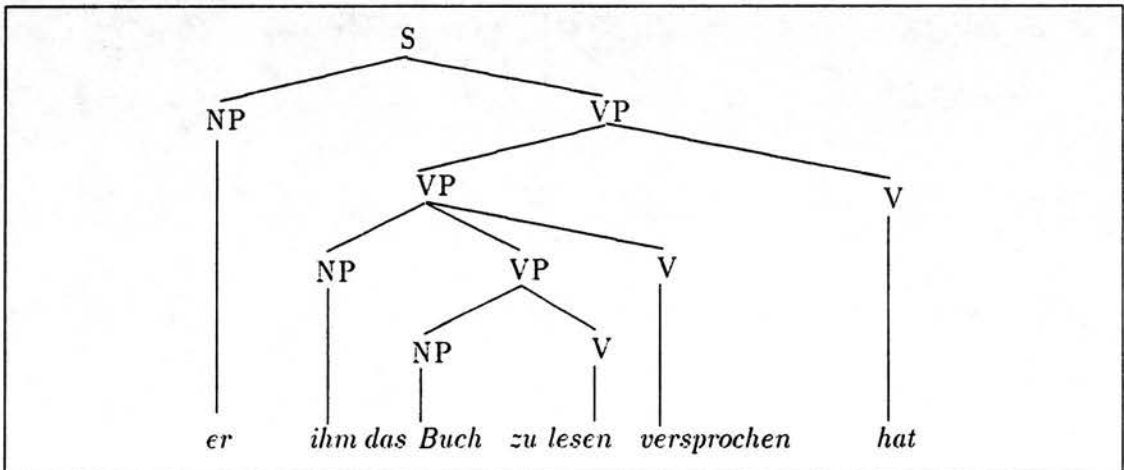


Figure 1-16: GB style tree for *er ihm das Buch zu lesen versprochen hat*

Finally, consider the tree in Fig. 1-17 for *es ihm jemand zu lesen versprochen hat*.

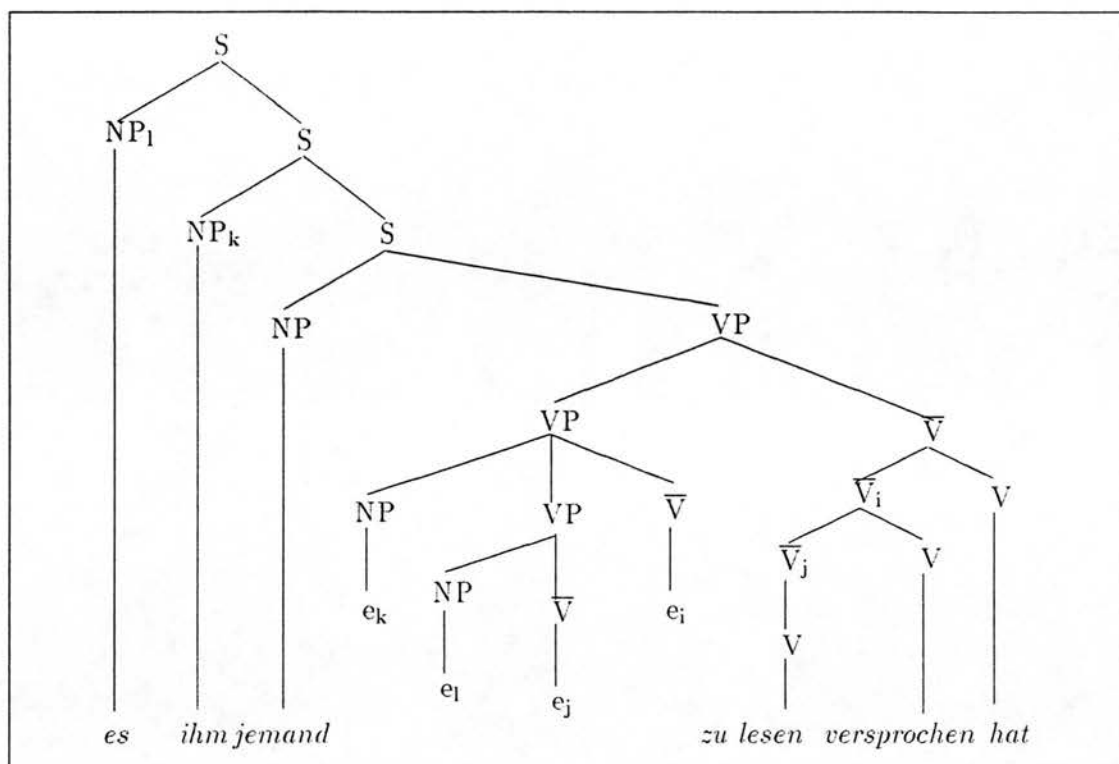


Figure 1-17: GB raising style tree for *es ihm jemand zu lesen versprochen hat*

Here we have movement of the NPs *es* and *ihm* and of the verbs *zu lesen* and *versprochen*. Perhaps this is the correct structure if we want to generalise across the syntax of German and Dutch for in this case we “raise” the verbs in the German example as well as in the Dutch example. Presumably Chomsky’s prohibition of string vacuous movement comes into play here but if it does then there is no explanation of why movement is necessary in the Dutch case (except to get the word order right). Presumably Dutch and German have the same structures. They just differ in the direction of adjunction of \bar{V} s.

In the end, GB generates phrase structures which have no empirical basis. The traces are purely theoretically motivated and have no motivation in the empirical data. There has been some work on the empirical “existence” of traces in English, notably on “wanna” contraction. The data is dubious however as many speakers of English accept sentences like “Who do you wanna win?” as easily as they accept “Where do you wanna go?”. However, to my knowledge there has been no comparable work on establishing the empirical evidence for traces in German and Dutch clauses exhibiting discontinuities and the lack of it in German clauses with a properly nested structure at both *D*-structure and *S*-structure.

Furthermore, there is no explanation of why there is adjunction and movement in German but not in Dutch and “verb raising” in Dutch but not in German.

At the theory internal level, the grammars of Dutch and German look quite different, something which is rather fatal for a theory which endeavours to explain differences between natural languages by universal grammar. I think it is not too controversial to say that universal grammar does not really come into the key aspects of the analyses of these types of constructions in German and Dutch.

It is difficult to argue against GB on grounds of empirical inadequacy since there are so many practitioners with so many different assumptions and the framework is so powerful that for any given problem someone will almost always come up with an analysis. But this strikes at the heart of whether GB is an empirically falsifiable theory. Many would say that it is not. Its very lack of formalisation makes the entire issue very treacherous. Instead we must argue on the basis of the empirical consequences of any given analysis and show that they are false or that they fail to satisfy other desirable consequences of a theory of the empirical data. We have tried to do this above.

1-4.2 Syntax as a theory internal construct

First, as Pollard and Sag point out, syntax is the least justifiable part of linguistic theory and it would be better if we could do without it. Our proposals are a big step in that direction by introducing the tectogrammatic-phenogrammatic distinction which allows us to massively simplify the tectogrammatic (functor-argument structure) component. (The tectogrammatic component is the component which most closely resembles conventional phrase structure. Actually, it is more similar to the *D*-structure of GB except it is unordered.) The phenogrammatic component (word order domain structure) is argued below to be empirically justifiable. Since phrase structure is a theory internal construct and since we are proposing a new theory it is unnecessary to take phrase structure on board. Phrase structure is only justifiable in theory internal terms, not empirical ones. Therefore, it is a candidate for replacement.

I would argue that phrase structure is largely a historical artefact of the development of modern linguistic theory. It has its origins in the early days of the formalisation of linguistic structure (in the work of Zellig Harris and Noam Chomsky's *Syntactic Structure*) as the primary way of characterising the set of strings in a language. As far as I know, there has been no serious attempt within modern linguistics to eliminate phrase structure to date. There have been many proposals (which we will discuss in Ch. 2) to modify the treatment of phrase structure to deal with discontinuous constituency and free word order but no proposals to abandon it completely. Even Hudson's Word Grammar, while purporting to eschew phrase structure, requires a strict adjacency condition on constituents which allows it to be reformulated as conventional phrase structure (i.e., there are no overlapping dependency arcs). The reasons for phrase structure's robust position within the theory of syntax are a matter for speculation, but it is very clear that the essential properties of phrase structure have changed very little since *Syntactic Structures*.

One problem with the theory of phrase structure is that it is clearly anglocentric. Most early work in generative grammar was done on English. By the time other languages came to be considered in detail, phrase structure was firmly entrenched within the theory. This anglocentric basis is something which ought to be reconsidered in the light of languages which systematically exhibit discontinuous constituency and (partial) free word order but this has not been done. Thus, phrase structure as it is conventionally assumed is an anglocentric artefact of modern linguistic theory.

1-4.3 Empirical justification for our approach

Our approach to syntax consists of two parts: syntactic functor-argument structure and word order domain structure. Syntactic functor-argument structure is clearly empirically motivated. For example, consider yet again *es ihm zu lesen versprochen hat*. *zu lesen* subcategorises for an accusative object (*es*). *versprochen* subcategorises for a dative object (*ihm*) and a **zu**-VP complement which can be seen by considering the extraposed version *ihm jemand versprochen hat, es zu lesen*. Finally *hat* subcategorises for a nominative subject (*jemand*) and a past participle VP *es ihm zu lesen versprochen* which can be seen by considering the V2 clause *jemand hat es ihm zu lesen versprochen*.

Word order domains can be shown to have an empirical basis in a way that phrase structure cannot. Consider again the German example (*daß*) *es ihm jemand zu lesen versprochen hat*. We know that the three NPs can be permuted under substitution. Therefore they must be in the same domain. Furthermore, there is some order freedom among the three verbs. Therefore they must be in the same domain. The question is whether the NP sequence and the verb sequence are part of the same domain. Although there is no scrambling of NPs and verbs, there is evidence that they are part of the same domain. If we consider the domain *es zu lesen* (which might occur in extraposed or topicalised position for example) and the domain *es ihm zu lesen versprochen* (which might appear clause finally in a V2 clause) we see that the domain *es zu lesen* appears discontinuously in the domain *es ihm zu lesen versprochen* and that *ihm versprochen* (which might appear clause pre-finally with its VP argument extraposed) also appears discontinuously in it. Thus, if we make the minimal assumption that domains are combined recursively then we can see that the domain of *es zu lesen* and the domain of *ihm versprochen* are subdomains of the domain of *es ihm zu lesen versprochen*. Furthermore, we can consider one more cycle of domain construction and see that *es ihm zu lesen versprochen* appears in the domain *es ihm jemand zu lesen versprochen hat*. Thus, both the NP sequence and the verb sequence must appear in the same domain. Of course, the domain construction rules are theory internal empirical nonobservables however consideration of constituent domains and “compound” domains indicates quite clearly what relation holds between them.

By similar arguments we can show that substrings must be in different domains. For example, in Dutch, the subject never permutes with objects (except in the

case of passives and unaccusative verbs) and objects sometimes permute. Thus, objects must be in the same domain while the subject is in a different domain (in unergative constructions).

To put it another way, we know that *es ihm jemand zu lesen versprochen hat* is in a single domain precisely because there are overlapping discontinuities which span the whole string.

Similarly, we can argue that linear precedence statements are hardly more than direct empirical generalisations stated in terms of these empirically motivated domains.

1-5 Generalisations and Formalisation

1-5.1 Theory-internal vs. Empirical Generalisations

Another motivation for the approach presented here is to make generalisations about universal grammar, families of languages and individual languages which are more empirically motivated and less theory internally motivated. This is not to say that such generalisations are “superficial” either in the sense that they are not profound generalisations or in the sense that they refer to “surface” features of the empirical data. Neither does it mean that this work is an exercise in structural linguistics or language typology of the type familiar from Greenberg and others. Rather, the goal is to make generalisations which are primarily motivated by the empirical data and not primarily by theory internal considerations. The generalisations that we make are based on an abstract theory in the same way that GB is for example but are far less prone to arbitrary theory redefinition to accomodate some difficult piece of data. Were some particularly difficult class of examples for our theory to be found, the entire empirical basis of the theory would have to be reexamined to determine precisely which empirical generalisations we were making turned out to be inaccurate.

This goal of more empirically-based generalisations is not only a motivation for our approach but a justification for it. The theory we present is rather simple and based heavily on generalisations which have their foundation in the empirical data. Other more theory internal accounts of the same range of data can be very complex and based heavily on theory internal argumentation, argumentation which is difficult if not impossible to justify empirically.

One must ask the question of such theories with such elaborate theoretical constructs whether the elaborateness of the theory is justified by the empirical evidence, and indeed, by the range of empirical data covered. This question is valid on a language by language basis. That is, it may well be that such a theory purports to explain a vast array of data from a truly wide range of languages. but if the theory purports to be a theory of universal grammar and we find that “explanations” for some language or some set of related languages which are newly being considered involves much theory internal argumentation, then we have to

wonder whether or not there is too much theoretical “machinery” for such a small empirical payoff and whether or not some simpler set of more empirically-based assumptions is not called for. This is a regular process in science and is not a suggestion which is to be dismissed out of hand. Like computer programs, scientific theories begin life pristine with a clear structure which then becomes more and more obscure as changes are made to them until there comes a point where the entire theory must be reformulated from scratch. Generative grammar since the publication of *Syntactic Structures* has undergone several such “overhauls”, the most sweeping probably being the creation of Government and Binding Theory. Therefore, this process is as native to linguistic theorising as it is to the physical sciences.

1-5.2 Vertical Locality

Domain construction only takes into account one level of domain structure. For example, if we have a daughter domain $[x_1 [x_2 Y_1 Y_2]]$, then either the entire domain can occur as an element of the mother’s domain or $[x_2 Y_1 Y_2]$ can occur as an element of the mother’s domain through domain union. However, we cannot “look through” the second level of bracketing (corresponding to X_2) and make Y_1 and Y_2 elements of the mother’s domain. Arbitrary discontinuity arises from recursive domain construction. Therefore, domain construction embodies a kind of *vertical locality* principle since only the internal structure of constituent domains is available to the construction of a mother domain. That is, arbitrary selection of nested domains is not allowed. This is similar to the principle in **HPSG** that subcategorisation for the **DTRS** feature is not allowed but subcategorisation for an element which still contains subcategorised elements is allowed.

1-5.3 Getting Things in Order

Most syntactic approaches to discontinuity start from some notion of *canonical* order or canonical structure and then try to derive noncanonical structures which exhibit the continuity. In **GB** this is done by the use of *move- α* to “rearrange” *D*-structure trees into *S*-structure trees. In **GPSG**, this is done by using metarules to derive additional phrase structure rules from those stated in the ID/LP format. In **LFG**, this is done by allowing very schematic phrase structure rules whose *f*-structures are related by the use of functional uncertainty and the coherence and completeness conditions. In categorial grammar, this is accomplished by the use of type-changing rules which reorder the lexically specified order of arguments on functors. I could go onto to consider other theories here but it suffices to say that all such approaches start with a notion of canonical order and canonical structures and then try to “get things out of order”.

Our approach to order on the other hand is the very opposite. The lowest level constituent domains are ordered only by **LP** statements. When they are combined to form larger domains, other **LP** constraints apply and recursively upwards.

Thus, we see that order is *monotonically increasing* as we move up a domain tree. Once a domain is ordered, its elements maintain that order in any other domain that they occur in. Thus, it can be said that our approach starts with no notion of canonical order or canonical structure and recursively “gets things in order”.

1-5.4 Markedness and Canonality

One characteristic of our approach is that there is no notion of *markedness* and no notion of *canonicity*. Functor-argument structure is unordered. Domain elements are ordered by LP statements which may allow more than one possible order. (This is highly characteristic of German.) Whereas other approaches would hypothesise some initial structure with fixed order and then derive alternative orders from it, there is no sense in which our approach can be said to pick out some particular order as “initial”, or in other words, canonical. Thus “canonicity” and “noncanonicity” are two terms which clearly play no part in our formulation. Markedness similarly lacks a function in our account. It is a concept which is frequently appealed to but almost never defined. In our context, it would have to mean that certain noncanonical structures were assigned as being more “marked”. However, since we have no notion of canonicity to refer to, neither can we have a notion of markedness to refer to.

1-6 Syntactic Categories in West Germanic

If our hypothesis is correct that the grammars of genetically related languages are very similar, then we should expect the same syntactic categories to appear in similar positions for translation equivalents. For example, the syntactic tree for *Jan Piet Marie zag helpen zwemmen* and *Jan saw Pete help Marie swim* should contain the same syntactic categories with roughly the same subcategorisation requirements. In the case of the functor-argument approach this is true. The two functor-argument structures in Fig. 1-18 and Fig. 1-19 (presented as trees) are identical except for the lexical entries. (Of course the order of the constituents within a functor-argument “phrase” is irrelevant.)

Monostratal theories fail across the board on this account. Reconsider the tree in Fig. 1-20. There are no VPs whatsoever.

However, we should expect that where there are VPs in English, there are VPs in both Dutch and German. Whatever analysis a monostratal phrase structure based approach takes, there will be no VPs in Dutch and German when there are verb clusters. This leads us to another point. If we consider extraposition variants of German examples, phrase structure approaches would have to posit two lexical entries for each verb which we consider to take VP complements, one that takes a VP when the VP complement is extraposed and one that subcategorises for a nonmaximal category when there is a verb cluster. For example, consider the

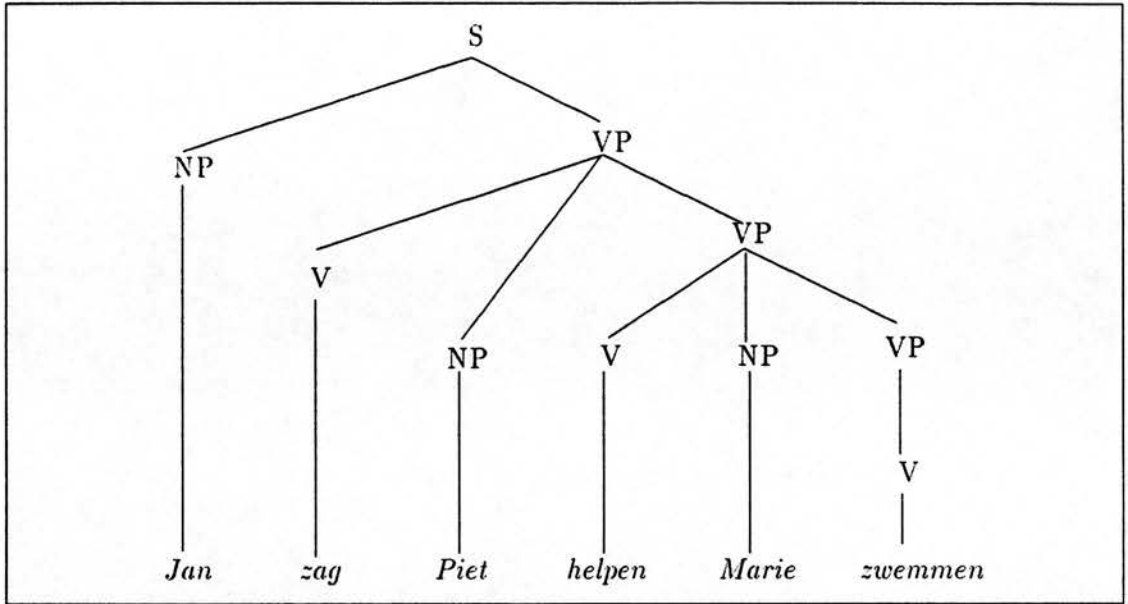


Figure 1-18: Functor-argument tree for *Jan Piet Marie zag helpen zwemmen*

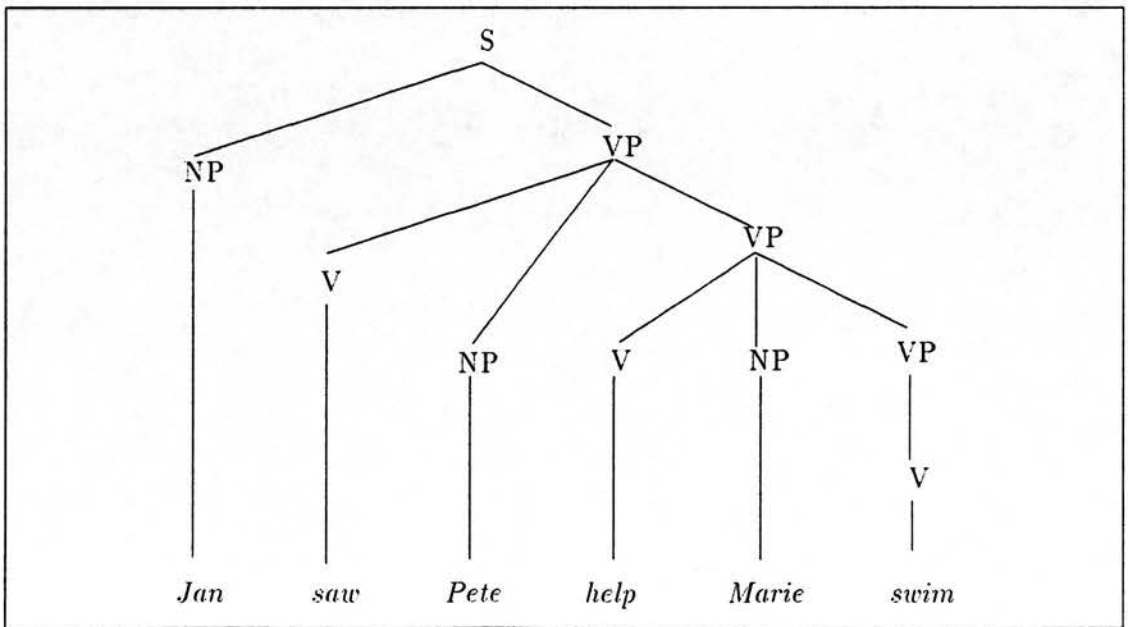


Figure 1-19: Functor-argument tree for *John saw Pete help Marie swim*

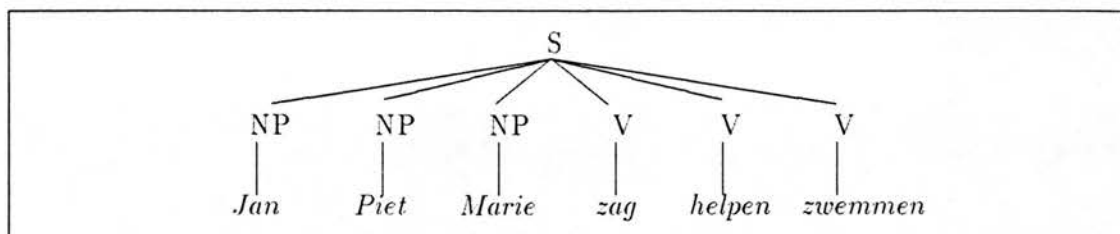


Figure 1-20: GPSG style “flat” tree for *Jan Piet Marie zag helpen zwemmen*

following examples.

- (1.3) a es ihm jemand zu lesen versprochen hat
 it him someone to read promised has
 ‘someone has promised him to read the book’
- b ihm jemand versprochen hat, es zu lesen

In the (a) example, a phrase structure approach must hypothesise a flat structure and thus *versprochen* will subcategorise for a sub-VP verb projection whereas in the VP extraposition variant (b), it must subcategorise for the extraposed VP *es zu lesen*.

The situation gets much worse however if we consider certain categorial analyses. In these analyses, verbs are type raised over other verbs to create complex verbs which have all of the NP subcategorisations required. As an example, I will review the UCG account presented in [47, §5.4]. The following presentation is based very closely on that account.

Zeevat proposes to type raise the complements of infinitives over the infinitives which subcategorise for them. Ignoring issues of order and semantics, the sign for *laat* ‘let’ in subordinate clauses is the following.

- (1.4) *laat*
 $\text{sent}[\text{fin}]/\text{np}[\text{nom}]/\text{np}[\text{obj}]/(\text{sent}[\text{bse}]/\text{np})$

The sign for *zwemmen* ‘swim’ is the following.

- (1.5) *zwemmen*
 $X/(X/(\text{sent}[\text{bse}]/\text{np}))$

Thus the complement *zwemmen* of *laat* is type-raised over the verb that subcategorises for it, *laat*. Therefore, by functional application, we get the following sign.

- (1.6) *laat zwemmen*
 $\text{sent}[\text{fin}]/\text{np}[\text{nom}]/\text{np}[\text{obj}]$

In the case of transitive and ditransitive verbs, things are slightly more complicated. Consider the signs for *drinken* 'drink' and *geven* 'give'.

(1.7) *drinken*
 $X/np[obj]/(X/(sent[bse]/np))$

(1.8) *geven*
 $X/np[obj]/np[obj]/(X/(sent[bse]/np))$

In these cases, the NP subcategorisations are satisfied after the verbs have combined with the bare infinitival VP (i.e., $sent[bse]/np$).

The infinitival version of *laten* is as follows.

(1.9) *laten*
 $X/np[obj]/(sent[bse]/np)/(X/(sent[bse]/np))$

The tree for *(dat) Jan Marie de kinderen bier zag laten drinken* is given in Fig. 1-21.

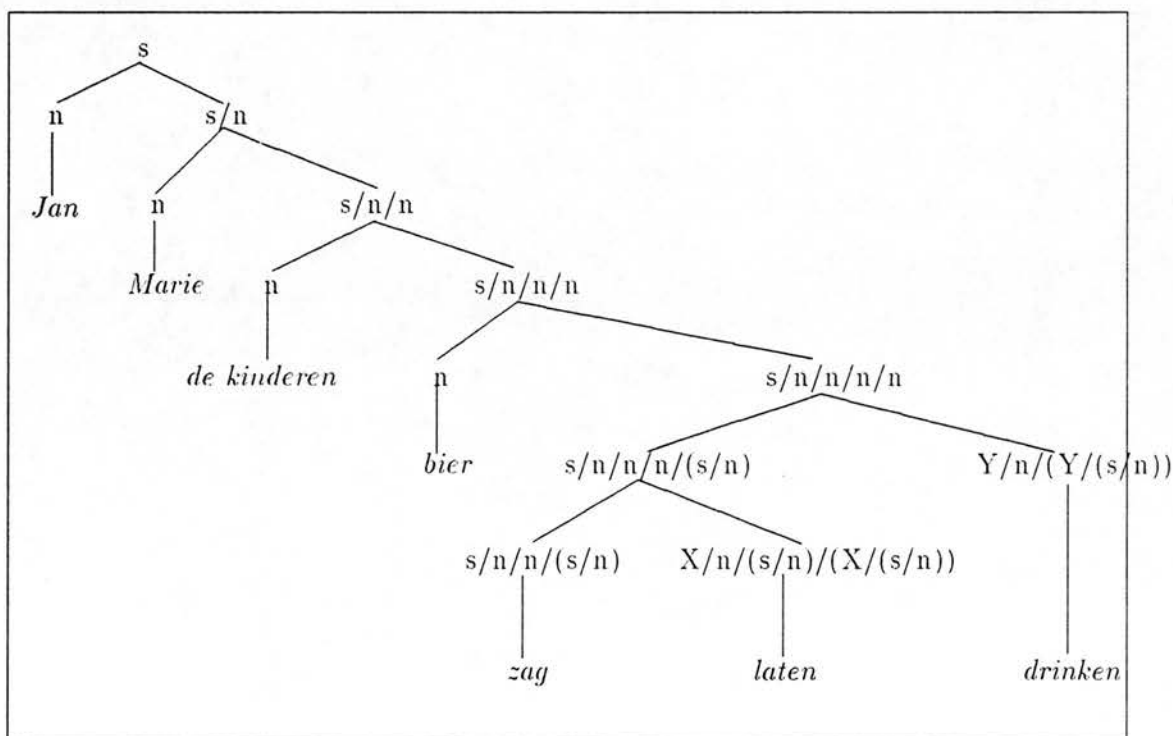


Figure 1-21: UCG tree for *(dat) Jan Marie de kinderen bier zag laten drinken*

My criticism of this approach is that dependent verbs do not subcategorise for their heads nor do verbs subcategorise for their dependent verbs. Rather verbs

which take VP complements uniformly subcategorise for VPs. The problem with phrase structure based approaches is that this route is not available to them when dealing with languages with nonconfigurational surface structure. What is clearly needed is a level of representation at which subcategorisation requirements are expressed uniformly and another level of representation at which surface word order is dealt with. In this respect GB has the matter right by allowing both *D*-structure and *S*-structure but as I have argued elsewhere, GB is empirically inadequate and unmotivated at the level of *S*-structure and adopts an excess of empirically unmotivated theory internal machinery. The functor-argument/domain structure account on the other hand deals with this data in a very simple empirically motivated fashion.

1-7 Synchronic and Diachronic Syntax

Jack Hoeksema in two papers ([18] and [19]) presents evidence of the suppression of a word order pattern in German and Dutch. The verb cluster in German and Dutch is taken to be uninterruptable by nonverbal material. For example, consider the following data.

- (1.10)
- a Zij lachte omdat ik naar adem moest happen
she laughed because I to breath must gasp
'She laughed because I had to gasp for breath'
 - b Zij lachte omdat ik naar adem happen moest
 - c Zij lachte omdat ik moest happen naar adem
 - d Zij lachte omdat ik happen moest naar adem
 - e *Zij lachte omdat ik moest naar adem happen
 - f *Zij lachte omdat ik happen naar adem moest

In (a-d), the two verb clusters *moest happen* and *happen moest* are uninterrupted by nonverbal material. However, in the ungrammatical examples (e-f), the verb clusters contain the PP *naar adem*.

However, dialects of Flemish and Swiss German allow this type of pattern.

- (1.11) a Hij had zich *kunnen* in weelde *ontwikkelen* en een
he had himself can in wealth develop and a
beroemde naam maken zekerlijk
famous name make certainly
'He would have been able to develop in wealth and to make a
name for himself'
- b dass er *het wöuwe chönne* mit dem Velo uf ds Jungfrauoch
that he has want can with the bike to the Jungfrauoch
faare
go
'that he wanted to be able to go by bike to the Jungfrauoch'

The precise analysis of these constructions is not important here.³ What is important is that Hoeksema establishes on diachronic evidence that the word order pattern exhibited in these examples was well established in Middle Dutch dialects dating back to the thirteenth century. Hoeksema also goes on to show that the pattern is not uncommon in the spoken language of native speakers of Dutch and is also common in second language speakers of Dutch. He also conjectures that children overgenerate in this respect and he has some anecdotal evidence to support this.

Hoeksema concludes that this word order pattern used to be a fully productive part of Middle Dutch grammar and that furthermore, it is a fully productive part of modern Dutch grammar which is stylistically marked. He attributes this markedness to the production component of linguistic performance.

While accepting Hoeksema's evidence, I would argue that this word order pattern is no longer a fully productive part of Dutch grammar. Rather it is a semiproductive pattern which is dying out. If we compare diachronic morphological evidence, it is very common to find what can only be classified as derivational morphology which is limited in its application although it will have been fully productive in the past. I would say that the same is true of the word order pattern mentioned above with regard to Standard Dutch and Standard German while still being largely retained in varieties of Flemish and Swiss German dialects.

This topic is relevant to the syntactic process that I will call *raising* after Uszkoreit's term *focus raising*. Raising occurs when one of the dependents of a verb in an extraposed VP appears in the Mittelfeld of the finite clause. Uszkoreit ([45])

³For the record, however, I would analyze them as examples of bare infinitival VP extraposition. This is not a subject I will address in this thesis however.

presents several examples of raising. I will repeat his (365) here.

- (1.12) Letztes Jahr hatte Peter [das große Haus]_i der Stadt
 last year had Peter the big house the city
 versprochen e_i zu reparieren
 promised to repair
 ‘Last year Peter had promised the city to repair the big house’

However, raising is a strictly restricted process. First of all, it only occurs in peripheral position, i.e., in extraposed and topicalised VPs. Second, the distribution of raising from extraposed and topicalised VPs differ. Third, raising is less acceptable if the case of the raised element(s) matches the case of some element of the Mittelfeld. Uszkoreit presents two examples ((374) and (375)) which show this.

- (1.13) ?Ich hatte ihm darum diesen Kindern versprochen zu helfen
 I had him therefore these children promised to help
 ‘I had therefore promised him to help these children’

- (1.14) ?Ich hatte sie deshalb dieses Buch gebeten bis moregen
 I had her therefore this book asked by tomorrow
 zurückbringen
 to return
 ‘I had therefore asked her/them to return this book by
 tomorrow’

Multiple raising is possible but what is not possible is the raising of several complements which share case with the elements of the Mittelfeld. Thus, raising is either a restricted stylistic but productive part of modern German grammar or it is a semi-productive process which has its roots in Old or Middle German. Either way, raising is a restricted, semi-productive process which is not on a par with, say, VP extraposition and what we will descriptively call here verb raising.

1-8 Scientific Methodology

It is this sort of result that is an important goal of the GPSG approach to linguistics: the construction of theories of the structure of sentences under which significant properties of grammars and languages fall out as theorems as opposed to being stipulated as axioms.

Gazdar, Klein Pullum and Sag, 1985

Before we consider an example from German, we need to consider some methodological issues which arise from the formal approach taken in this thesis. In

much linguistic literature, there is an implicit assumption that the program of searching for the structure of universal grammar should take the form of finding a formalism or metalanguage which allows the grammatical description of all and only the human natural languages. Thus, the theory of universal grammar is identified with the expressive power of the formalism or metalanguage. I call this the *isomorphic formalism requirement* or *ISO* for short. I reject this approach to linguistic methodology as both unnecessary and fundamentally flawed. Rather the approach we take might be called *axiomatic*. This is not to be confused with the school of linguistic theory called *Axiomatic Linguistics*. By *axiomatic*, I mean that individual grammars and indeed the theory of universal grammar is formalised as a set of axioms in some formal, interpreted language.

1–8.1 The Isomorphic Formalism Requirement

The isomorphic formalism requirement is usually taken to be both a necessary and sufficient condition for characterising universal grammar. I agree that *ISO* is a sufficient condition for characterising universal grammar but not a necessary one. Before I argue for the elimination of *ISO* I first would like to present the following lengthy quote from GKPS ([14]). Although the quote does not convincingly justify the assumption of *ISO*, it is one of the clearest and most explicit statements of it I have seen. I will discuss the quote below.

“The most interesting contribution that generative grammar can make to the search for universals of language is to specify formal systems that have putative universals as consequences, as opposed to merely providing a technical vocabulary in terms of which autonomously stipulated universals can be expressed. . . .

. . . , that universals are most interesting when embedded as integral parts of a formal system has some nontrivial structure, involves just as clear a break with the approaches adopted in much current work. It goes without saying that the process of searching for grammatical universals initially involves attempting to discover facts about language (as opposed to facts about some particular language or set of languages). But there is a sense in which even a precise formulation of a successful discovery of this sort will not constitute truly interesting results in theoretical linguistics. If the fact needs a special statement, as opposed to following from the very form in which the theoretical reconstruction of the notion ‘natural language’ has been cast, the job is not done.

Thus, for example, one might propose that natural language grammars never exhibit direct grammatical dependencies between elements separated by more than two phrasal categories of a certain sort, or that they never permit a full category in a certain position in the clause, or whatever. But these proposed universals are not accounted for by the mere fact of their having been written down in some uninterpreted

algebraic formalism. The explanatory task has not even begun when a constraint or generalization is merely stated. Only when it can be shown to be a nontrivial consequence of the definition of the notion ‘possible grammar’ can it be regarded as explained, because while it resides in the form of an autonomous statement it can be modified, enhanced, weakened, or even discarded with no consequences for the rest of the theory (cf. Dowty 1982b, pp. 107-8, on this important point). The penalty for failure of such a universal is effectively zero; a new universal saying something carefully hedged to avoid the last known counterexample can be constructed in a moment. Ironically, in view of the fact that such universals are often presented with a considerable fanfare of rhetoric about explanation, they have much the same status as the descriptive universals we find in the typological work that takes its lead from Greenberg (1963) – only these claims, being better researched, generally have a much better half-life.

Our goal in the work that has led to **GPSG** has been to arrive at a constrained metalanguage capable of defining the grammars of natural languages, but not the grammar of say, the set of prime numbers. . . .

$$(1) \quad [\text{VFORM FIN}] \subset [-\text{N}, +\text{V}]$$

. . . If (1) is a universal, then it should not need saying. It ought to be a consequence of the grammatical metalanguage itself – for example, by virtue of a theory of features which (unlike ours) ties tense securely to the semantic notion it expresses and simultaneously restricts its syntactic realization to verbal categories in the theory of grammar. If this were done effectively, the discovery of a language with tensed adjectives would severely compromise the theory of features as a whole and force revisions that would alter the consequences of the theory in other domains. If we simply rest content with the universal stipulation ‘ $[\text{VFORM FIN}] \subset [-\text{N}, +\text{V}]$ ’, we can drop it, modify it to say ‘ $[\text{VFORM FIN}] \subset [+V]$ ’ at no real cost. . . .

We therefore regard universals stated within the metalanguage as inherently less interesting than those which are built into it. We exhibit in this book some claims, for example the Exhaustive Constant Partial Ordering claim about linear precedence in grammars. . . . which follow as consequences of our overall formal system. It is this sort of result that is an important goal of the **GPSG** approach to linguistics: the construction of theories of the structure of sentences under which significant properties of grammars and languages fall out as theorems as opposed to being stipulated as axioms.”

The claim that **ISO** is more interesting than simply expressing generalisations in the metaformalism is a completely subjective claim. What is interesting and what is not is a matter of personal taste. Being interesting is not an important metric in science. Before worrying about achieving **ISO**, we should first try to correctly

model the empirical domain first. The search for a formalism which is equivalent to universal grammar can come after we have found a correct characterisation of universal grammar in the first place. Then we can try to find an isomorphic formalism later if we want. Needless to say, we are a long way from finding a correct characterisation of universal grammar. There is another reason to doubt the program of ISO however. That is, we have no reason to believe that the program is achievable at all. To take a similar example, the theory of Boolean algebras is expressed in a standard first order logic with equality but to my knowledge there is no language which captures precisely the class of Boolean algebras and allows the formalisation of all and only the Boolean algebras. The claim that this sort of thing is possible holds a large promissory note with no evidence that it is in fact achievable.

GKPS make the following point in the quote above.

“Only when it can be shown to be a nontrivial consequence of the definition of the notion ‘possible grammar’ can it be regarded as explained, because while it resides in the form of an autonomous statement it can be modified, enhanced, weakened, or even discarded with no consequences for the rest of the theory (cf. Dowty 1982b, pp. 107-8, on this important point). The penalty for failure of such a universal is effectively zero; a new universal saying something carefully hedged to avoid the last known counterexample can be constructed in a moment.”

It is true that a universal in the form of an “autonomous statement” can be adjusted at any time. Linguistic universals as expressed by linguists are not etched in titanium. However, the same can be said of the ISO approach. Putative universals are changed on a daily basis by practicing linguists and GPSG is as guilty of this criticism as any other linguistic theory. This is just the nature of science. It is not true though that the cost of changing a universal is effectively zero. If the universal has any substance then it will interact with other principles of universal grammar and so small changes to any one principle will have large consequences for the theory as a whole. But this is true of both the ISO approach and the approach which rejects it. Finally, creating a new universal which carefully avoids the “last known counterexample” is just bad science but again is as problematic with ISO as with the axiomatic approach.

Linguistics seems to be the only science requiring anything remotely resembling ISO. In all the physical sciences, scientists avail themselves of whatever mathematical machinery they deem necessary. For example, physicists typically use as much of the differential and integral calculus as they feel they need and if there is not a suitable notation they invent a new one (e.g., tensors). The closest analogy to linguistics seems to be computer science. In both cases, the scientists invent new languages to allow the encoding of various types of knowledge. In the case of computer science, the argument usually takes the form “X is the best programming language”. Of course the response to this type of assertion is

that there is no “best” programming language but that the choice of language depends very much on the task at hand. In linguistics, there is only one task, the description of natural languages. Thus the basis for argumentation over the correct formalism can only be based on descriptive felicity (see below). In other words, the question of the “correct” formalism is one of descriptive ease rather than empirical superiority.

The non-ISO approach has often been criticised on the grounds that the formalism that the axioms are expressed in is in fact the theory of universal grammar, and since the formalism is so powerful, it says nothing substantive about universal grammar at all. If we take another example from physics, we can see the fallacy of that argument. The theory of heat flow is expressed in the language of partial differential equations but that does not mean that there is no theory of heat flow or that the theory of heat flow is just the theory of partial differential equations. In precisely the same way, it is unnecessary to identify the metaformalism with the theory of universal grammar.

If one accepts that ISO is unnecessary, then one can immediately see that ISO is a masochistic activity. It makes the task of formalising universal grammar massively more difficult than it has to be. Besides which, we will *never* know what universal grammar is owing to the ubiquity of language and the limited resources that linguistic theory as a sociological enterprise has to bring to bear on the topic. There will always be forthcoming data which punctures our theories. Surely every practicing linguist is familiar with this problem. Since this is the case, there is good reason to postponing the activity of finding a metaformalism equivalent to UG since it is doubtful that we will ever have a definitive theory (short of experimentation on children).

Finally, ISO is hypocritical. The metaformalism of every theory of natural language ever proposed is clearly inadequate as a theory of universal grammar. We are kidding ourselves if we say that we should encode UG as a metaformalism and then propose nothing but metaformalisms which are clearly inadequate (e.g., have bad computational complexity or are capable of generating unnatural languages). We know at the outset that what we are proposing as a theory of UG is hopelessly inadequate. If we really believe in ISO, we should never present it anyway.

1-8.2 Theory Presentations and Theories

A *theory presentation* is a pair $\langle \Sigma, L \rangle$ where L is some formal, interpreted language with a *consequence relation* \vdash and Σ is a set of *sentences* or *formulas* of L . (I am intentionally informal about what constitutes a formal, interpreted language because I do not want to unnecessarily restrict the class of languages under consideration. Basically, I have in mind formal logics, programming languages with a rigorous semantics and the like.) Theory presentations are often called *theories* by logicians and nonlogicians alike. However, a *theory* of L is a set of sentences of L which is closed under logical consequence. When we use theory

in the strict sense, it is to this definition that we refer. If a theory contains the formula \perp then it contains every formula of L . In this case, we say that the theory is *inconsistent*. If a theory does not contain \perp it is *consistent*. The elements of a theory presentation are usually called *axioms*.

The *theory closure* ($\vdash(P)$) is the set $\{\phi \mid \Sigma \vdash \phi\}$. Thus, the theory closure $\vdash(P)$ of a theory presentation \mathbf{P} contains \mathbf{P} as a subset and also anything which is a logical consequence of \mathbf{P} . $\vdash(P)$ will usually be much bigger than \mathbf{P} . For example, if it is classical, it will be infinite in size even if \mathbf{P} is finite.

An important point to be made which will figure in our discussion of methodology below is that two theory presentations can be *logically equivalent* in the sense that the closure of both presentations is the same set. To take a trivial example consider the two theory presentations, $\mathbf{P}_1 = \{P, Q, R\}$ and $\mathbf{P}_2 = \{P \wedge Q \wedge R\}$ (where P , Q and R are propositional variables). Then $\vdash(\mathbf{P}_1) = \vdash(\mathbf{P}_2)$.

1-8.3 The relationship between a formalised theory, the mathematical domain and the empirical domain

Pollard and Sag (P&S2) present a discussion of the relationship between a formalised theory, the mathematical domain the theory formalises and the empirical domain that is modelled by the mathematical domain. Since they state the situation very elegantly and since I am in complete agreement with them, I will present the following rather lengthy quote on this topic ([32, pp4-5]).

“Let us begin by making explicit some methodological assumptions. In any mathematical theory about an empirical domain, the phenomena of interest are *modelled* by mathematical structures, certain aspects of which are conventionally understood as corresponding to observables of the domain. The theory itself does not talk directly about the empirical phenomena; instead it talks about, or is *interpreted by*, the modelling structures. Thus the predictive power of the theory arises from the conventional correspondence between the model and the empirical domain.

Now an informal theory is one that talks about the model in natural language, say a technical dialect of English, German or Japanese. But as theories become more complicated and their empirical consequences less straightforwardly apparent, the need for formalization arises. In cases of extreme formalization, of course, the empirical hypotheses are cast as a set of axioms in a logical language, where the modelling structures serve as the intended interpretations of the expressions in the logic.

For example, in one kind of standard model of celestial mechanics, the positions and velocities of bodies subject to mutual gravitation are represented by vectors in a higher-dimensional Euclidean space

(“phase space”), the masses of the bodies by positive real numbers, and their motions by paths along certain smooth vectorfields (“flows”) on the space. Of course such a model is not the same thing as what it models (e.g. the solar system), but certain formal properties of such a model may represent aspects of the solar system of interest to a physicist. Now in a formal theory based on such a model, the underlying logic is just a standard first-order language (e.g. the language of Zermelo-Fraenkel set theory) and the axioms are certain systems of differential equations (e.g. Hamiltonian systems) that the flows are required to satisfy. An observed motion of the solar system is then predicted by the theory insofar as it agrees – under the conventional correspondence – with an admissible flow (i.e. one that satisfies the equations). ...

In our view, a linguistic theory should bear exactly the same relation to the empirical domain of natural language, viz. the universe of possible linguistic objects, as a mathematical theory of celestial mechanics should bear to the possible motions of n-body systems. Thus we insist on being explicit as to what sorts of constructs are assumed (i.e. what ontological categories of linguistic objects we suppose to populate the empirical domain), and on being mathematically rigorous as to what structures are used to model them. Moreover, we require that the theory itself actually count as a theory in the technical sense of precisely characterizing those modelling structures which are regarded as admissible or well-formed (i.e. corresponding to those imaginable linguistic objects which are actually predicted to be possible ones). This does not mean that the empirical hypotheses must be rendered in a formal logic as long as their content can be made clear and unambiguous in natural language (the same holds true in mathematical physics), but in principle they must be capable of being so rendered. Unless these criteria are satisfied, an enterprise purporting to be a theory can not have any determinate empirical consequences. Thus we emphatically reject the current widespread view which holds that linguistic theory need not be formalized. Rather, our position is the same as the one advocated by Chomsky (1957:5) and subsequently abandoned by him ...”

1-8.4 Prediction

The methodological approach that we adopt here has been criticised as having no predictive capacity on the basis that the metaformalism makes no predictions about universal grammar. This criticism follows immediately from ISO since the metaformalism is necessarily a theory of universal grammar under ISO. However, as we have shown ISO is unnecessary. The criticism that our approach has no predictive capacity could hardly be more false. First, every time we make a statement in the formalism, there is a whole range of theoretical consequences

which are predicted. For example, if we state that NPs precede Vs in word order domains, then this means that all NPs must precede all verbs in every possible domain structure. Typically, it will not be immediately apparent whether this is true or not. For example, does it apply to topicalised VPs, extraposed VPs, intraposed VPs, small clauses? Conversely, such a statement limits the class of putative domain structures to those which satisfy the ordering constraint. But domain structure indirectly reflects syntactic structure and so on. The ramifications can be far reaching.

Second, because our theories are formalised in a logic (formal language) with a consequence relation means that adding an axiom ϕ to a theory \mathbf{T} means that there will immediately be other sentences of the logic that can be derived. Formally speaking, $\vdash (T) \subseteq \vdash (T \cup \{\phi\})$. But this is simply the Galois connection. As you add more axioms to a theory, the more inferences one can draw.

From a theoretical perspective, this second type of predictive adequacy is extremely important for the following reason. Once we formalise a theory of \mathbf{UG} in some theory presentation, every theoretical consequence and empirical prediction follows automatically from considering the theory closure of the theory presentation of \mathbf{UG} . Thus, the common concept of prediction disappears, or at least becomes trivialised, since the theory closure contains everything there is to know.

1-8.5 Psycholinguistics

ISO holds that universal grammar is identified with the metalanguage or formalism in which possible human languages are expressed. The reasoning runs that if a linguistic formalism has intractable computational properties then it is an untenable theory of universal grammar since humans process natural languages very quickly, the theory of universal grammar or the equivalent metalanguage must be computationally tractable. This has been taken as evidence by Berwick, Barton and Ristad among others that most, if not all, modern grammatical theories are inadequate as theories of universal grammar because of the computational properties of their associated formalisms. (**GB** is conspicuously absent from such criticism since it is not sufficiently formalised to test its computational properties.)

However, if **ISO** is dropped then this metric for the evaluation of linguistic theories is unavailable. Rather, it is the theory of universal grammar (in the technical sense) whose computational properties should be investigated. The properties of the formalism or formal language in which universal grammar is expressed are irrelevant. Just as the quicksort algorithm has a computational complexity of $\mathcal{O}(n \log n)$ regardless of the programming language or specification language that it is encoded in, so universal grammar has an inherent computational complexity which is independent of the language it is expressed in. The theory of universal grammar can be written in any formal language which is expressive enough. The fact that we may choose to write it in a formal language whose satisfiability prob-

lem is harder than that for the theory of universal grammar is of no consequence for psycholinguistics.

What are the consequences for psycholinguistics of the axiomatic approach then? The answer is none. The axiomatic approach is agnostic with respect to psycholinguistic issues. Universal grammar, if it is ever to be discovered, will be formalised as a theory presentation. That is, for some language L and some set Σ of sentences of L , universal grammar will be the theory presentation $\langle \Sigma, L \rangle$. However, there will be many equivalent theory presentations to $\langle \Sigma, L \rangle$, many of them expressed with respect to formal languages other than L . The theory closures of all such theory presentations will be equivalent under appropriate morphisms between languages. Thus, given a theory presentation of universal grammar, it is not even the theory presentation whose properties should be investigated but the equivalent theory presentation which has the best computational properties. This is more likely to be a job for a mathematical logician or a theoretical computer scientist than a psycholinguist.

What then can psycholinguistics offer to this approach to universal grammar? Again, the answer is nothing. Given independent existence of the set of equivalent theory presentations of universal grammar, it is impossible to infer anything about this set from psycholinguistic evidence. The theory of universal grammar may be implemented in any number of ways, none of which are externally determinable. The task of inferring the structure of universal grammar or the grammar of any particular human language from external behaviour is like trying to infer the organisation of a Lisp program by watching its input-output behaviour while the program runs on a Lisp interpreter with an overactive garbage collector. We *will* be fooled by the garbage collector and if the garbage collector has any regular behaviour we will devise a theory about such behaviour which has a very tenuous connection, if it has any connection at all, with the organisation of the program which is the real object of study.

1-8.6 Computational Linguistics

Just as the complexity of the metaformalism is irrelevant to psycholinguistics, it is irrelevant to computational linguistics as well. A theory of universal grammar or of a particular language should not be criticised because it is formalised in a formalism with intractable computational properties. It is the task of the computational linguist to discover other equivalent formalisations or subsets of the formalism which are expressively adequate which have tractable properties. Work of this type was carried out on GPSG where several communities of researchers devised revisions of the GPSG formalism which were computationally tractable but nevertheless allowed a formulation of grammars which was still true to the fragment in GKPS.

To take another example, the satisfiability problem for nonempty theories of the language \mathcal{L}^+ defined in §A is undecidable and intractable in certain respects. However, there is a subset of the relation symbols, namely those which are *deter-*

ministic, and a subset of the axioms, namely those which do not contain negated variables in the context of a universal relation, which defines a subset of the language and a subset of the theories such that satisfiability is tractable (although still undecidable).

1-8.7 Descriptive Felicity

Although the metaformalism used to formalise universal grammar or the grammar of a particular human language is irrelevant as long as it is expressive enough to allow such formalisation, we should nevertheless formalise our theories in formalisms which have *descriptive felicity*. That is we should choose a formalism which is appropriate to the task. Just as we might choose Lisp for list processing applications, Prolog for deductive reasoning and Fortran for intensive numeric calculations, we should also choose a metaformalism for linguistic formalisation which allows linguistic generalisations to be stated in as simple a manner as possible. We might call this requirement *descriptive transparency*. Within the computational linguistics community and within certain branches of theoretical convergence there has been a convergence towards feature-value logics augmented with sorts and function and relation symbols. (The language \mathcal{L}^+ defined in §A is such a language.) Many linguists have found such languages convenient for writing grammars of natural language. This trend should be continued.

1-9 Cross-linguistic Variation

In this section we will discuss the formal notions of *theory* and *subtheory* and discuss their application to formalising a principle-based approach to universal grammar.

1-9.1 Universal Grammar, Theories and Subtheories

An algebra $\langle L, \sqcup, \sqcap \rangle$ s.t.

$$(a \sqcup b) \sqcup c = a \sqcup (b \sqcup c)$$

$$(a \sqcap b) \sqcap c = a \sqcap (b \sqcap c)$$

$$a \sqcup b = b \sqcup a$$

$$a \sqcap b = b \sqcap a$$

$$a \sqcup a = a$$

$$a \sqcap a = a$$

$$a \sqcup (a \sqcap b) = a$$

$$a \sqcap (a \sqcup b) = a$$

for all $a, b, c \in L$. is a *lattice*.

A lattice $\langle L, \sqcup, \sqcap \rangle$ s.t.

$$a \sqcap (b \sqcup c) = (a \sqcap b) \sqcup (a \sqcap c)$$

$$a \sqcup (b \sqcap c) = (a \sqcup b) \sqcap (a \sqcup c)$$

for all $a, b, c \in L$, is a *distributive lattice*.

An algebra $\langle L, \sqcup, \sqcap, -, 0, 1 \rangle$ s.t. $\langle L, \sqcup, \sqcap \rangle$ is a distributive lattice and

$$a \sqcup 0 = a$$

$$a \sqcap 1 = a$$

$$a \sqcup -a = 1$$

$$a \sqcap -a = 0$$

for all $a \in L$, is a *Boolean algebra*.

All three algebras are formalised in terms of implicitly quantified universally quantified first order formulas with the function symbols \sqcup , \sqcap and $-$ and the equality relation. Thus they are equivalent to the following three sets of first order sentences.

$$\forall a, b, c. ((a \sqcup b) \sqcup c = a \sqcup (b \sqcup c))$$

$$\forall a, b, c. ((a \sqcap b) \sqcap c = a \sqcap (b \sqcap c))$$

$$\forall a, b, c. (a \sqcup b = b \sqcup a)$$

$$\forall a, b, c. (a \sqcap b = b \sqcap a)$$

$$\forall a, b, c. (a \sqcup a = a)$$

$$\forall a, b, c. (a \sqcap a = a)$$

$$\forall a, b, c. (a \sqcup (a \sqcap b) = a)$$

$$\forall a, b, c. (a \sqcap (a \sqcup b) = a)$$

$$\forall a, b, c. (a \sqcap (b \sqcup c) = (a \sqcap b) \sqcup (a \sqcap c))$$

$$\forall a, b, c. (a \sqcup (b \sqcap c) = (a \sqcup b) \sqcap (a \sqcup c))$$

$$\forall a. (a \sqcup 0 = a)$$

$$\forall a. (a \sqcap 1 = a)$$

$$\forall a. (a \sqcup -a = 1)$$

$$\forall a. (a \sqcap -a = 0)$$

Call the sets of sentences \mathbf{T}_1 , \mathbf{T}_2 and \mathbf{T}_3 respectively. Then let $T_{lattice} = T_1$, $T_{distributive} = T_1 \cup T_2$ and $T_{Boolean} = T_1 \cup T_2 \cup T_3$. Then $\mathbf{T}_{lattice}$ is the *theory* of

lattices, $\mathbf{T}_{distributive}$ is the theory of distributive lattices and $\mathbf{T}_{Boolean}$ is the theory of Boolean algebras.⁴ Then obviously, $T_{lattice} \subset T_{distributive} \subset T_{Boolean}$.⁵

The theory of universal grammar and the grammar of a particular language stands in precisely the same type of subset relation as the theory of lattices does to the distributive lattices or Boolean algebras. Assume that universal grammar is formalised as a theory presentation $\langle \Sigma, L \rangle$ whose theory closure is \mathbf{T}_{UG} . Then necessarily the theory of universal grammar, \mathbf{T}_{UG} is a subset of \mathbf{T}_{Eng} , i.e., $T_{UG} \subset T_{Eng}$. That is, if UG exists in our heads as a theory (presentation) then the grammar of any particular language must be a superset of UG.

This means that cross-linguistic variation exists as a continuum. Just as we can add arbitrary universally quantified sentences to the theory of lattices to define new theories of varieties (in the nontechnical sense) of lattices, we can also add arbitrary axioms to the theory of universal grammar to define the grammar of yet other possible natural languages. That is, *any* extension of universal grammar defines a possible human language. This may seem counterintuitive to some but it can be no other way given the axiomatic approach to universal grammar adopted here and, implicitly, in the work of Pollard and Sag on HPSG.

Furthermore, it is also true under widespread assumptions concerning the form of universal grammar as a formalised theory. Many linguists would hold that universal grammar is by its nature equivalent to some metaformalism. That is, the substantive hypotheses about natural language are “built into” the metaformalism itself. Let us call the metaformalism that UG is equivalent to \mathbf{UG} . Then the theory presentation for \mathbf{UG} is just $\langle \emptyset, UG \rangle$. That is, it is the empty theory whose consequences follow only from the formalism (i.e., \mathbf{UG}) itself. Then by conventional assumptions, any theory expressed in the metaformalism of \mathbf{UG} is the grammar of a possible natural language. But any such theory presentation will be $\langle \Delta, UG \rangle$ for some set $\Delta \subset UG$. Therefore $\emptyset \subseteq \Delta$ and so the conventional assumptions lead to the same conclusion. Therefore cross-linguistic variation is a continuum whatever set of assumptions are made, so long as language-specific *principles* of grammar are allowed in addition to parameters.

1-10 An Example from German

Before proceeding with a detailed analysis of German verb projections, I will present a simple example of a German “cross-serial” subordinate clause. The analysis assumes the framework described in §1-2. In (1.15), the subscripts indicate the head-complement dependencies. Each NP is separated from its head by other constituents. *es* ‘it’ is the direct object of the verb *zu lesen* ‘to read’, *ihm* ‘him’ is the dative object of the past participle *versprochen* ‘promised’ and *jemand* ‘someone’ is the subject of the finite auxiliary *hat*. *zu lesen* subcategorises

⁴These “theories” are properly *theory presentations* in the language of first order logic with equality and not *theories* but this is a common abuse of terminology.

⁵The theory presentations of the theories stand in the same subset relations.

for a nominative subject and an accusative object, *versprochen* subcategorises for a nominative subject, a dative object and a *zu*-infinitival VP and *hat* subcategorises for a nominative subject and a past participle VP.

- (1.15) daß es₃ ihm₂ jemand₁ zu lesen₃
 that it (ACC) him (DAT) someone (NOM) to read
 versprochen₂ hat₁
 promised has
 ‘that someone promised him to read it’

To analyze this clause, we only need to make four assumptions beyond those made in §1–2. First, the configuration [_S NP[NOM] VP] is ungrammatical in German. Instead, verbs take all their complements as sisters in clauses. This does not mean that there are no VPs, just that they never form a clause with a subject. Second, we assume the LP constraint NP < v. Third, a verb follows any verb that it governs.⁶ A verb V_i governs a verb V_j iff V_j is the head verb of a VP complement of V_i or V_i governs V_k and V_k governs V_j. (This is expressed using \preceq .) Fourth, all of the verbs in (1.15) domain union their VP complements. (This is artificially trivial but is all we need to analyze this example.)

The verb *zu lesen* ‘to read’ subcategorises for a direct object so we can form the VP *es zu lesen* ‘to read it’ with domain (1.16).

- (1.16) [VP [NP es] [v zu lesen]]

The NP *es* ‘it’ precedes the verb as required by the LP statement NP < v. The past participle *versprochen* ‘promised’ subcategorises for an indirect object and a *zu*-infinitival VP. *es zu lesen* is such a VP and *ihm* ‘him’ is the masculine, third person dative pronoun. If we union the domain of *es zu lesen* into the VP domain of which *versprochen* is the head we can form the VP *es ihm zu lesen versprochen* ‘promised him to read it’ with domain (1.17).

- (1.17) [VP [NP es] [NP ihm] [v zu lesen] [v versprochen]]

Since the order of the two NPs with respect to each other is unconstrained they may appear in this order.⁷ Both NPs precede both verbs as required and the governed verb *zu lesen* precedes its governing verb *versprochen* as required. *hat* ‘has’ subcategorises for a past participial VP. *es ihm zu lesen versprochen* is such a VP and *jemand* ‘someone’ is a nominative pronoun. If we union the domain of *es ihm zu lesen versprochen* into the finite clause domain we can form the clause

⁶This is basically the concept that von Stechow ([46]) calls “status government”.

⁷Actually one would want to develop a theory of ‘weak’ ordering constraints on NPs, as discussed by [45]. A theory of such constraints which builds on Uszkoreit’s work is presented in Ch. 2.

es ihm jemand zu lesen versprochen hat 'someone has promised him to read it' with domain (1.18).

(1.18) [VP [NP *es*] [NP *ihm*] [NP *jemand*] [v *zu lesen*] [v *versprochen*] [v *hat*]]

Again, all NPs precede all verbs and *zu lesen* and *versprochen* precede the governing auxiliary *hat* so the domain is wellformed. The assumptions made above account for the possible permutations of NPs in the Mittelfeld and the canonical order of verbs in the verb sequence. Figure 1-22 is the syntax tree of (1.15) and Figure 1-23 is the domain tree of (1.15). Domain trees have domains as the nodes of trees instead of categories. In a local domain tree, the mother domain node is constructed from the daughter domain nodes according to the domain construction rules introduced in §1-2.

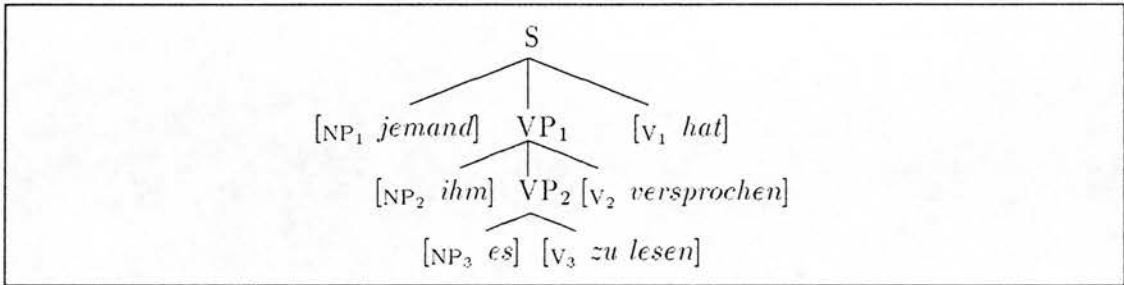


Figure 1-22: Syntax tree for (1.15)

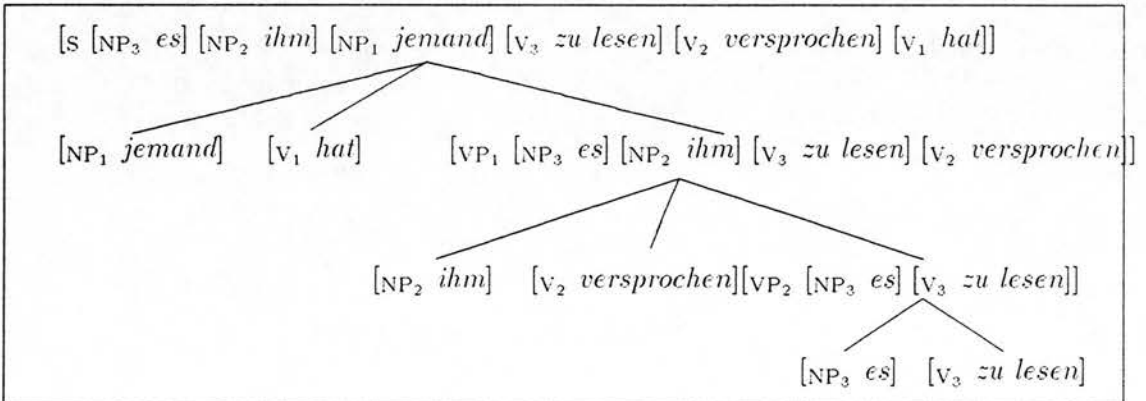


Figure 1-23: Domain tree for (1.15)

Both syntax trees and domain trees are unordered. Furthermore, the structure of the trees in Figure 1-22 and Figure 1-23 is the same (modulo the order of the daughters which is irrelevant). This isomorphism between syntax trees and domain trees always holds given the rules for domain construction in §1-2. This



means that domain construction is strictly compositional in the Montagovian sense. However, unlike composition of meaning translation in Montague semantics where the composition rules are all functional, domain construction is relational (because of the relational character of domain union and the nondeterminism in applying domain union or not). So, for (1.15) we could derive a total of six domain trees corresponding to the clauses in (1.19) since the order of NPs is unconstrained.

- (1.19) es ihm jemand zu lesen versprochen hat
es jemand ihm zu lesen versprochen hat
ihm es jemand zu lesen versprochen hat
ihm jemand es zu lesen versprochen hat
jemand es ihm zu lesen versprochen hat
jemand ihm es zu lesen versprochen hat

Chapter 2

The Theoretical Framework – from Phrase Structure Grammar to Word Order Domains

“Let them eat cake.”

Marie Antoinette

“Let them have trees.”

Ron Kaplan

In this chapter, I will review previous proposals to account for phrase structure in general and languages with discontinuous constituency and partially free word order in particular. I will try to show the chronological development of the various proposals and how they have contributed to and differ from the account presented in this thesis. There are two basic strands of development. First, phrase structure grammar developed into categorial grammar and **GPSG**. From **GPSG**, Hans Uszkoreit ([45] and [44]) proposed various modifications to accommodate a fragment of German grammar. In many ways, the work in this thesis carries on Uszkoreit's initial proposals. A further development is **HPSG** from which this work borrows a lot of the organising principles of grammar and its formalism. The second strand begins with the classic work of Gunnar Bech on the German infinitive, *Studien über das Deutsche Verbum Infinitum* ([3]). This work is followed up by Evers' classic PhD thesis, *The Transformational Cycle in Dutch and German* ([13]). Many of the ideas underlying both Bech's and Evers' analyses are shared by the account we present here.

2–1 Context-Free Phrase Structure Grammar

Phrase structure grammars take a very simple form. Given a set of nonterminal symbols, N , and a set of terminal symbols, T (which we will assume without loss

of generality are disjoint), phrase structure grammars consist of sets of *rewrite rules* of the form

$$(2.1) \alpha \rightarrow \beta_1, \dots, \beta_n$$

or simply

$$(2.2) \alpha \rightarrow \epsilon$$

where $\alpha \in N$, $\beta_1, \dots, \beta_n \in N \cup T$ and $\epsilon \in T$ is the *empty string*.

A *derivation* consists of replacing a left hand symbol by its right hand side and then recursively substituting the right hand sides of the rules whose left hand side appear in the string. For example, if we have the following set of rules

$$\begin{array}{l} S \rightarrow NP VP \\ NP \rightarrow Det N \\ Det \rightarrow the \\ N \rightarrow boy \\ N \rightarrow dog \\ VP \rightarrow V NP \\ V \rightarrow kicked \end{array}$$

then we can produce the following derivation.

S
 NP VP
 Det N VP
 the N VP
 the boy VP
 the boy V NP
 the boy kicked NP
 the boy kicked Det N
 the boy kicked the N
 the boy kicked the dog

Corresponding to a derivation is its *derivation tree*. The derivation tree for 'the boy kicked the dog' is given in Fig. 2-1. Derivation trees are also called *phrase structure trees*.

It is easy to see that the order of right hand sides of rules is preserved in the phrase structure tree because of the close connection between derivations and their trees. For example, S rewrites as the sequence NP followed by VP and in the phrase structure tree for 'the man kicked the dog' the NP subject node occurs to the left of the VP node. Therefore, word order is determined entirely by the ordering

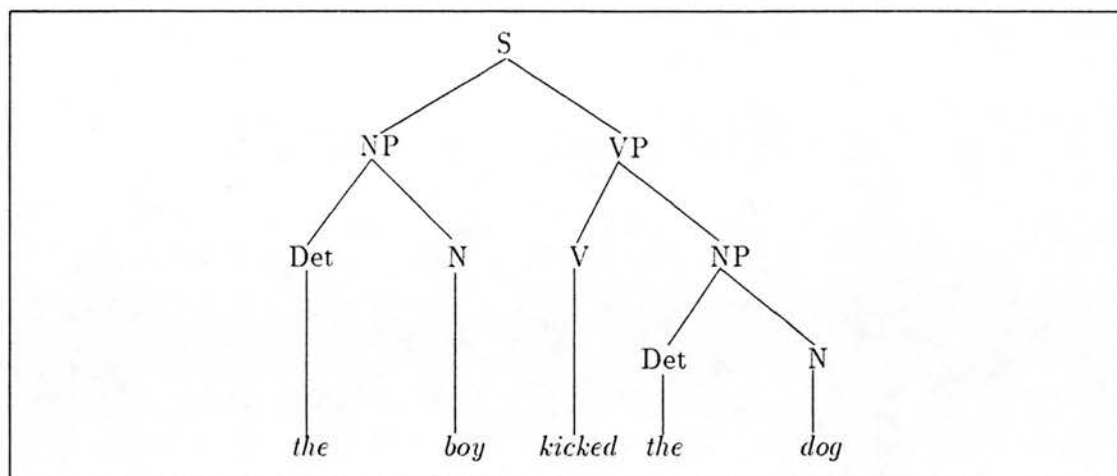


Figure 2-1: Phrase structure tree for *the boy kicked the dog*

of right hand side elements. This is sufficient as long as there is no variation in order. In this case, alternate rules have to be written which allow the other orders. Furthermore, if there are generalisations about order which hold of many or all rules of a grammar, there is no way to express them. A phrase structure grammar of English will contain many rules of the form $VP \rightarrow NP \dots VP$, i.e., we can make the generalisation that NPs precede VPs. However, there is no way to express this generalisation in phrase structure grammar.

Phrase structure grammar is also wholly inadequate with respect to discontinuous constituency of all types. There is no way available to “link” the displaced elements with the constituent from which they were displaced.

2-2 Categorical Grammar

In this section, I will discuss simple bidirectional categorial grammar. There have been many extensions to categorial grammar proposed to handle various phenomena but I will not detail them here.

A bidirectional categorial grammar is defined in the following way. Let *Atomic* be a set of atomic categories. Then $Atomic \subset Cat$ and $\alpha/\beta \in Cat$ and $\alpha \setminus \beta \in Cat$ for $\alpha, \beta \in Cat$. *Cat* is the set of categories.

In pure bidirectional categorial grammar there are only two rule schemata of the following form.

$$\begin{aligned} X &\rightarrow X/Y Y \\ X &\rightarrow Y X \setminus Y \end{aligned}$$

Thus two instances of the rule schemata are the following.

$$\begin{aligned} s \backslash np &\rightarrow (s \backslash np) / np \quad np \\ s &\rightarrow np \quad s \backslash np \end{aligned}$$

Elements of *Cat* are assigned to lexical entries. Assume the following category assignments: 'Jamie' np, 'Chris' np, 'a' np/n, 'sweetie' n and 'gave' (s \ np) / np. Then the tree in Fig. 2-2 is the derivation tree for 'Jamie gave Chris a sweetie'.

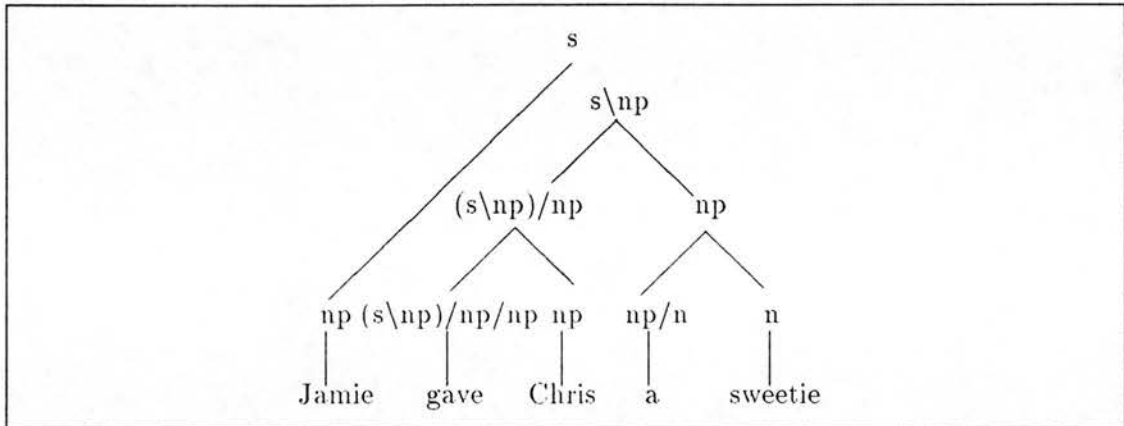


Figure 2-2: Derivation tree for *Jamie gave Chris a sweetie*

In pure categorial grammar, this is all there is. Categorial grammar is partially successful in handling word order problems since functor categories (those with a slash) determine whether their arguments (the slashed elements) appear to the left or to the right and in what order. However, for languages with verb clusters and other types of discontinuous constituency, the formalism must be extended. There have been many proposals of various types but all of them seem to overgenerate with respect to the data.

2-3 Generalized Phrase Structure Grammar

Gazdar, Pullum, Klein and Sag develop a unification-based approach to context free phrase structure grammars in their book *Generalized Phrase Structure Grammar* ([14], henceforth referred to as **GKPS**). The theoretical framework is rather elaborate. Here we are interested in only two aspects of the theory: the so-called *immediate dominance/linear precedence* format, or ID/LP for short, and the use of a metarule to handle subject-auxiliary inversion.

The following exposition is based directly on **GKPS**.

ID/LP format factors out immediate dominance and linear precedence information from *local* trees. (A local tree is a tree of depth one.) A normal phrase structure rule such as $A \rightarrow B C D$ states that a tree with mother A is licensed which has daughters B , C and D appearing in that order. To state the same

information in ID/LP format we would need an *immediate dominance* rule of the form $A \rightarrow B, C, D$ and two *linear precedence* statements of the form $A \prec B$ and $B \prec C$. The reason for switching to ID/LP format is to capture generalisations about word order which cannot be captured in simple phrase structure grammars.

Consider what local trees the ID rule (4) admits

$$(4) \quad A \rightarrow B, C, D$$

(4) licenses the set of local trees in Fig. 2-3. The rule licenses local trees with mother A and daughters B, C and D occurring in any order. It is easy to verify that all six permutations are licensed.

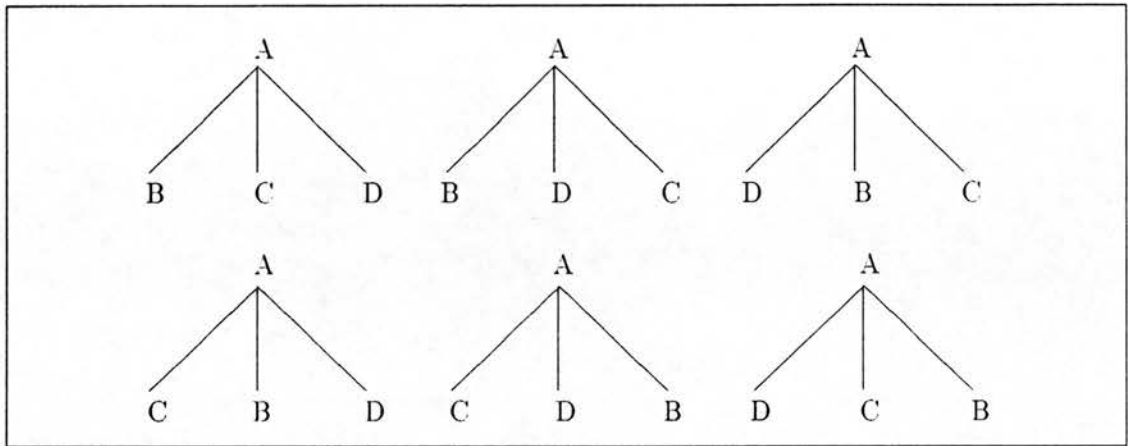


Figure 2-3: GKPS Fig. 5

If we now add the single LP statement $B \prec C$, then the number of possibilities is halved and only the trees in Fig. 2-4 are licensed. In each case, B precedes C.

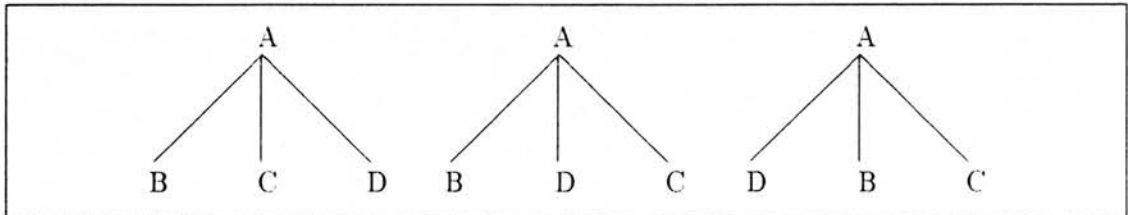


Figure 2-4: GKPS Fig. 6

For another example, consider the grammar (7).

- (7) i A → B, C, D
 B → A, C, D
 C → A, B, D
 D → A, B, C
- ii A < B < C < D

Taken together, (7i) and (7ii) are equivalent to the grammar (4) except that they capture the generalisation that the daughters appear in alphabetical order whereas (4) does not.

- (1) A → B C D
 B → A C D
 C → A B D
 D → A B C

Consider the toy grammar in (8). It fails to express at least two generalisations: AUX and V are always constituent initial and NPs precede VPs.

- (8) S → NP VP
 S → AUX NP VP
 VP → AUX VP
 VP → V VP
 VP → V NP
 VP → V NP VP

An equivalent grammar in ID/LP format is (9).

- (9) i S → NP, VP
 S → AUX, NP, VP
 VP → AUX, VP
 VP → V, VP
 VP → V, NP
 VP → V, NP, VP
- ii AUX < NP
 V < NP
 NP < VP

This grammar may seem to be more verbose than (8) but the economy in expression can be seen from considering a more realistic grammar fragment. (10)

presents almost all the rules necessary to characterise Makua VPs.

- (10) i
- VP → V
 - VP → V S
 - VP → V NP
 - VP → NP V
 - VP → NP V NP
 - VP → NP NP V
 - VP → V NP NP
 - VP → V NP S
 - VP → V S NP
 - VP → NP V S
 - VP → NP V PP
 - VP → PP V NP
 - VP → PP NP V
 - VP → NP PP V
 - VP → V PP NP
 - VP → V NP PP

(10) is equivalent to the much more compact ID/LP grammar (11).

- (11) i
- VP → V
 - VP → V, NP
 - VP → V, S
 - VP → V, NP, NP
 - VP → V, NP, PP
 - VP → V, NP, S
- ii V < S

Not all context free grammars can be rewritten in ID/LP format however. (12) and (13) are exactly as complex as (8) and (10) respectively yet neither of them can be expressed in ID/LP format.

- (12)
- S → NP VP
 - S → AUX NP VP
 - VP → AUX VP
 - VP → V VP
 - VP → V NP
 - VP → V VP NP

- (13)
- VP → V
 - VP → V S
 - VP → V NP
 - VP → NP V
 - VP → NP V NP
 - VP → NP NP V
 - VP → V NP NP
 - VP → V NP S
 - VP → S V VP
 - VP → NP V S
 - VP → NP V PP
 - VP → PP V NP
 - VP → PP NP V
 - VP → NP PP V
 - VP → V PP NP
 - VP → V NP PP

The property of grammars which can be expressed in ID/LP format is characterised by the *Exhaustive Constant Partial Order* property (ECPO). Informally, ECPO means that the set of expansions of any one category observes a partial ordering that is also observed by the expansions of every other category.

One such LP statement for the grammar of English is (14). It states that elements which are defined for the SUBCAT value precede elements which are not defined for SUBCAT. Among other things, the effect of this LP statement is to force lexical heads to precede their complements.

- (14) SUBCAT < ~[SUBCAT]

Another (complex) LP constraint is given below. It states that nominal and adjectival projections precede prepositional phrases which in turn precede verb phrases.

- (2.3) +N < P² < V²

These are the two major LP statements used for head-complement and head-adjunct syntax in GPSG.

2-3.1 'Subject-aux inversion' without subject, aux or inversion

One of the devices that GPSG uses to capture generalisations is the *metarule*. A metarule is of the form

$$A \rightarrow B, \dots \Rightarrow C \rightarrow D, \dots$$

The idea is that the input rule to the left of the \Rightarrow is used to produce an output rule of the form specified to the right of the arrow. Metarules can contain variables which range over sets of categories. A metarule called the 'subj-aux metarule' is used to generate subject-auxiliary constructions. The basic insight into such constructions is illustrated by (9).

- (9)
- a will do it
 - b has gone away
 - c should leave
 - d doesn't care
 - e are swimming
-
- a' Will you do it?
 - b' Has she gone away?
 - c' Should we leave?
 - d' Doesn't he care?
 - e' Are they swimming?

We can see in (9) that in subject-auxiliary constructions, the constituent which follows the subject NP is whatever would have followed the verb in an assertion clause. Therefore, a metarule something like (10) is required.

- (10) $VP \rightarrow W \Rightarrow S[+INV] \rightarrow W, NP$

(10) states that if there is a VP rule expanding to a set of categories W then there is also an inverted clause rule expanding to the set W plus a NP.

Another device that GKPS use to capture generalisations is the *Feature Cooccurrence Restriction* (FCR). One class of FCRs are implications of the form $\alpha \supset \beta$ meaning that if a category satisfies α then it must satisfy β .

FCR 10 (11) prevents VPs from ever being [+SUBJ]. (VPs are [-SUBJ] by definition. This FCR "defines" inverted clauses as being [+SUBJ].)

- (11) FCR 10: [+INV, BAR 2] \supset [+SUBJ]

Conversely, FCR 11 (12) requires that any category that is [+SUBJ] be a phrasal verb. (Both clauses and VPs are considered to be phrasal verbs. The SUBJ feature distinguishes them.)

- (12) FCR 11: [+SUBJ] \supset [+V, -N, BAR 2]

Finally we are ready for the 'Subject-Aux Inversion (SAI) Metarule' (13).

- (13) 'Subject-Aux Inversion (SAI) Metarule
 $V^2[-SUBJ] \rightarrow W \Rightarrow V^2[+INV, +SUBJ] \rightarrow W, NP$

(13) states that if there is a VP rule which rewrites to a set of categories W then there is an inverted clause rule which rewrites to W plus an NP. The effect is that the head verb, the subject NP and any complement of the verb will all be sisters in the inverted clause rule. Thus ordering constraints will order the head before the subject and the subject before the complement of the verb.

Consider the VP rule (15). (13) can apply to it to yield (16).

- (15) a VP[+AUX] \rightarrow H[46], VP[-AUX,BSE]
 b do
 c did see Laurie

- (16) $V^2[+AUX,+INV,+SUBJ] \rightarrow H^0[46], V^2[-AUX,-SUBJ,BSE], N^2$

Given the LP statements discussed above, this rule will admit the local tree in Fig. 2-5.

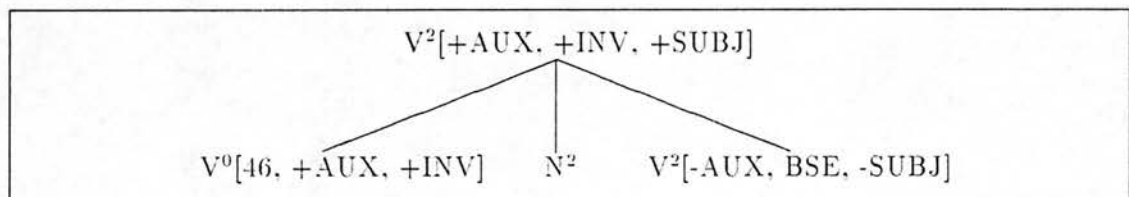


Figure 2-5: GKPS Fig. 17

Our purpose in this section has been to introduce just enough of the GPSG framework to understand various developments of it in following sections. Therefore, we satisfy ourselves with this brief introduction and turn to the next section.

2-4 Liberation Metarules

Geoffrey Pullum ([33]) and Arnold Zwicky¹ introduced a use of metarules that has come to be known as "liberation metarules" in an attempt to describe languages with free word order (as opposed to free constituent order). The idea is best explained by an example. (15) is an example of a liberation metarule. It states that if there is a rule that expands a VP into an NP and some other constituents X then there is a rule that expands VP into Det, N and X . Thus, the Det and the N are "liberated" in the VP.

- (15) $VP \rightarrow NP, X \Rightarrow VP \rightarrow Det, N, X$

¹Arnold M. Zwicky. 1985. *Free Word Order in GPSG*, Paper presented at Linguistic Institute summer meetings, Georgetown University. To appear in OSU WPL.

This allows Dets and Ns to appear discontinuously within VPs. There is one initial point to observe and that is that liberation is specified for the “containing” category, in this case, VP. It is not specified for a whole class of categories.

(16b) is an example of metarule (15) applied to the VP rule (16a). (16a) expands a VP to two NPs so (15) allows the first of these to be rewritten as a DET[ACC] and an N[ACC]. (15) could also apply to (16b) to produce (16c).

- (16) a VP \rightarrow H, NP[ACC], NP[DAT]
 b VP \rightarrow H, Det[ACC], N[ACC], NP[DAT]
- (16) c VP \rightarrow H, Det[ACC], N[ACC], Det[DAT], N[DAT]

One criticism of such an approach is that the liberation metarules are parochial in the sense that there is no expression of the generalisation that the right hand side of one rule is effectively being substituted for one of the right hand side symbols of another rule. This is the only option available within the GPSG approach to metarules. Recognising this problem, Zwicky ([49]) proposes a “Universal Liberation Metarule” metaprinciple which parameterises liberation metarule application to a list of pairs (A, B) such that B is substituted into A.

“My 1985 approach posits instead a Universal Liberation Metarule liberating the daughters of B into A, as in (4), with parochial stipulation localized in the list P of pairs (A, B) subject to this metaprinciple (so that free constituent order is described by stipulating that the pair (S, VP) is in P).”

- (4) Given a pair-set P for a language, then for any pair (A, B) in P.
- a. A and B are phrasal categories
- b. B \neq S, and
- c. If A \rightarrow B, X and B \rightarrow Y then A \rightarrow X, Y,
 where X and Y are any sets of constituents.

As we can see from (4a), Zwicky disallows liberation from an S node. This encodes the widely held assumption that elements cannot be scrambled out of the clause they originate in. Pullum writes:

“...Yet Ross very perceptively included in the specification of his scrambling rule a very plausible condition: constituents are allowed to scramble within, but not out of their clauses. This suggests an important distinction.

Chomsky (1965, 126) claims that “there is no known language” in which “each permutation of the words of each sentence give(s) a grammatical ... paraphrase of the original.” I think Chomsky is correct in this claim. ...”

I agree with Pullum and Chomsky that it is likely that scrambling out of clauses is probably universally impossible. Pullum however, goes onto to make a much stronger claim.

“There are interesting limits to what liberation metarules can do, however. The most significant of these is that a finite system of metarules yielding a finite output sets of rules cannot liberate a constituent of a recursive category into a higher recursive category. ... More generally, I claim that no constituent of a recursive category (one that can immediately dominate itself) can scramble out of that category. ...

I claim that this restriction obtains not because of some functional pressure to avoid ambiguity, but because the metagrammar is a set of statements specifying a *finite* set of CF-PSG rules. If a constituent of a recursive category could scramble up into a containing category, there would be no limit to the number of daughters a particular recursive category might have, and hence no longest constituent type; but the requirement that grammars be finite sets of rules guarantees that there is a finite bound on rule length.”

Pullum observes correctly that a finite set of liberation metarules which do not apply recursively are capable of generating only a finite number of phrase structure rules. He then goes on to state that “no constituent category ... can scramble out of that category” and takes the reason to be the fact that the metagrammar generates only a finite number of phrase structure rules.

It is here that I part company with Pullum’s observations. As I have claimed earlier, the forms of discontinuity in German and Dutch arise because of the recursive application of domain union to VP domains. I claim that this process is potentially unbounded. Indeed, Bech ([3]) cites examples where there are up to five levels of VP combination. While it is true that there is a tendency to extrapose VPs when the number of them becomes high, potentially we have to conclude that the process is unbounded. Were we to try to encode our insights about the “surface structure” (read domain structure) of German using liberation metarules we would have to allow recursive rule application. (Zwicky’s Universal Liberation Metarule appears to allow recursive application by the way.)

2-5 Uszkoreit (1987)

In [45], Hans Uszkoreit provides a fragment grammar of German in GPSG. His primary concerns are the structure of finite clauses, the position of the finite verb and the specification of the partially free order of NPs in the Mittelfeld. In [44], he extends his previous work by considering the topic of not only constituents in the Mittelfeld but also constituents in so-called “complex fronting” examples. He proposes a lexicalist solution to order based on the empirical evidence. In what

follows. I will review his proposals, pointing out points of convergence with our approach and empirical problems introduced by the theoretical treatments.

2-5.1 Grammar: Basics and Scope

It is indisputable that the VP exists as a constituent in German, at least minimally in peripheral positions, i.e., in topic position and extraposed position. (U98) presents four ID rules for intransitive, accusative transitive, dative transitive and ditransitive verbs respectively. (Here and in the sequel, we use the notation of GKPS which is slightly different than Uszkoreit's.)

- (U98)
- a $VP \rightarrow H[5]$
 - b $VP \rightarrow H[6], NP[+ACC]$
 - c $VP \rightarrow H[7], NP[+DAT]$
 - d $VP \rightarrow H[8], NP[+ACC], NP[+DAT]$

In order to explain the presence of the verb in clause-final position in subordinate clauses and clause initial position in V1 and V2 clauses, Uszkoreit considers a "flat" structure to be necessary, i.e., one in which the finite verb is a sister of both the subject and its other complements. The problem for a GPSG grammar is to derive a set of rules for these flat structures without listing them outright in the grammar. Uszkoreit uses the metarule (U99) to generate the appropriate "flat" ID rules.

- (U99) $VP \rightarrow X \Rightarrow V^3 \rightarrow NP[+NOM], X$

V^3 is the equivalent of S . Metarule (U99) takes a rule which rewrites a VP and produces a rule which rewrites a V^3 into the same list of head and complements plus a nominative NP subject. Thus, there is no NP-VP structure and finite clauses have the verb, the subject and any other complements as sisters.

(U124) introduces two feature cooccurrence restrictions. The first says that an "assertion clause" is a main clause. The second says that a main clause is finite.

- (U124)
- a $+AC \rightarrow +MC$
 - b $+MC \rightarrow +FIN$

Uszkoreit then introduces the two LP rules in (U125) to explain the position of the verb in finite clauses. (U125a) covers the clause-initial case (where the clause

must be +MC) and (U125b) covers the clause-final case (where the clause must be -MC).

- (U125) a $V[+MC] < X$
 b $X < V[-MC]$

(U126) shows the two phrase structure rules which are licensed by the metarule (U99), the feature cooccurrence restrictions in (U124), the LP rules in (U125) and the ID rule (U98b).

- (U126) a $V^3[+MC,+FIN] \rightarrow H[6] NP[+NOM] NP[+ACC]$
 b $V^3[-MC,+FIN] \rightarrow NP[+NOM] NP[+ACC] H[6]$

These phrase structure rules license the grammatical examples in (U127). (U127a) is a verb-initial clause as is found in constituents of V2 clauses and in yes-no V1 questions. (U127b) is a verb-final clause as is found as a constituent of subordinate clauses as in (U128).

- (U127) a kennt der Mann das Buch
 knows [+FIN] the man [+NOM] the book [+DAT]
 'the man knows the book'
 b dem Kind ein Mann hilft
 the child [+DAT] a man [+NOM] helps [+FIN]
 'a man helps the child'

- (U128) weil dem Kind ein Mann hilft
 because the child a man helps
 'because a man is helping the child'

Unfortunately, there is nothing to prevent a nonfinite clause from being formed. The ID rule in (U130) is licensed giving rise to the ungrammatical (U131). This is necessary for the treatment of modals and auxiliaries to be discussed below.

- (U130) $V^3[-MC,-FIN] \rightarrow H[5], NP[+NOM]$

- (U131) *Peter kommen

(U139) is one of the ID rules. It allows a perfective auxiliary to subcategorise for a past participial VP.

- (U139) $VP[+PERF,+AUX] \rightarrow H[1], VP[+PSP]$ *haben, sein*

The problem for Uszkoreit is to explain how auxiliaries and modals combine with their complements. He assumes a "main verb" analysis for German auxiliaries and modals similar to that assumed for English in GPSG. However, because of the flat clause structure, Uszkoreit assumes that auxiliaries and modals can subcategorise for clauses to form a clause. Thus, he introduces metarule (U140) to derive clausal versions of the auxiliary and modal VP rules.

(U140) $VP[+AUX] \rightarrow H, VP \Rightarrow V^3 \rightarrow H, V^3$

(U141) is the clausal rule derived from (U139) by metarule (U140).

(U141) $V^3[+PERF,+AUX] \rightarrow H[1], V^3[+PSP]$

Three other ID rules for modals and auxiliaries are given in (U143)–(U145).

(U143) $VP[+AUX] \rightarrow H[2], VP[+BSE]$ *müssen, können, dürfen...*

(U144) $VP[+AUX] \rightarrow H[3], VP[+PAS]$ *werden*

(U145) $VP[+AUX,+FIN] \rightarrow H[4], VP[+BSE]$ *werden*

These allow (U147) to be generated. Its derivation tree is given in Fig. 2–6.

(U147) *wird Peter gesehen werden können*
 will Peter seen be can
 'will Peter be able to be seen?'

Uszkoreit's fragment covers the data it sets out to but is inadequate if it were to be considered for other data. First, there is no consideration of cases where it looks like clause union or verb raising has applied. For example, it is impossible to generate *es ihm jemand zu lesen versprochen hat*. Second, the treatment of auxiliaries cannot be extended to handle anything but canonical verb order. His treatment relies on the fact that the verbs will be clause-final (except perhaps one). He thus produces a hierarchical structure which makes it impossible to generate the *Ersatzinfinitiv* 1–3–2 order for example. Furthermore, the treatment of modals and auxiliaries can be criticised on theoretical grounds. Although the modals and auxiliaries can be considered sentential operators at the level of semantics, it is doubtful that they take sentential complements at the level of syntax. Third, the account could not be extended to take account of VP extraposition (largely because of the treatment of modals and auxiliaries).

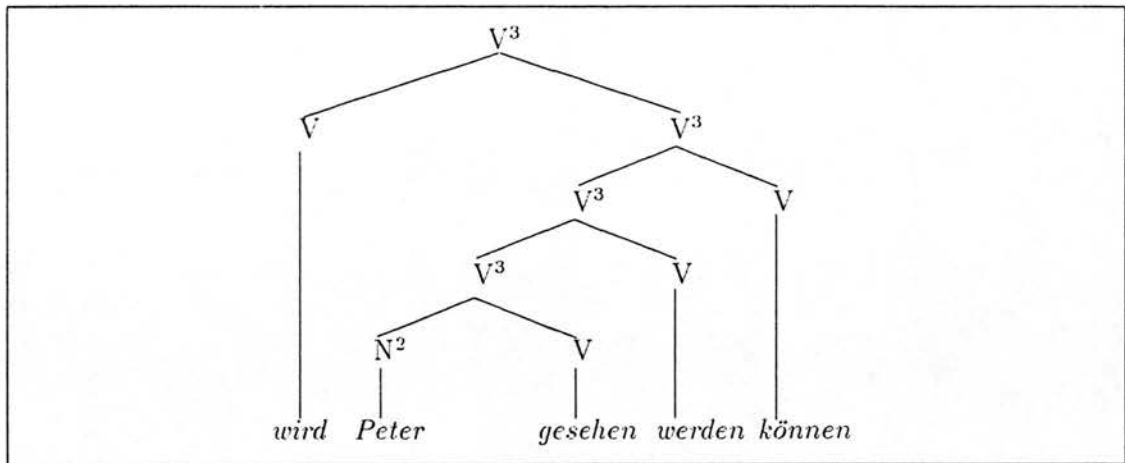


Figure 2-6: Phrase structure tree for *wird Peter gesehen werden können*

2-5.2 The Modified Ip Framework

Consider example (U281).

- (U281) Dann wird der Doktor dem Patienten die Pille geben
 then will the doctor the patient the pill give
 'then the doctor will give the patient the pill'

Uszkoreit points out that all six permutations of the subject, object and indirect object are acceptable. However, if no LP statements are stated then (U282) can be generated.

- (U282) ??Dann hatte einem großen Spielzeuglaster ihm es gegeben
 then had a big toy truck him it given
 'then it had given him a big toy truck'

Uszkoreit introduces the linear precedence statements stated in English in (U283) to partially characterise the order of German constituents within the framework of an extended fragment of German formalised in **GPSG**.

- (U283) a Focus follows nonfocus
 b The unmarked order is SUBJ. IOBJ. DOBJ
 c Personal pronouns precede other NPs

In addition, he adds the LP statements in (U284).

- (U284) a $V[+MC] < X$
 b $X < V[-MC]$
 c $X^2 < \text{SEPPREF}$
 d $+TOP < X$

(U284a) states that the verb in a main clause is clause-initial. (U284b) states that the verb in a subordinate clause is clause-final. (U284c) states that any phrasal constituent precedes a separable prefix in a clause. (U284d) states that the topic is clause initial.

To these LP constraints, Uszkoreit proposes adding the constraints in (U285) as a formalisation of the principles in (U283).

- (U285) a $+NOM < +DAT$
 b $+NOM < +ACC$
 c $+DAT < +ACC$
 d $-FOCUS < +FOCUS$
 e $+PRONOUN < -PRONOUN$

Uszkoreit views the LP component as a function from pairs of symbols to truth values. Given the LP relation $\{\langle \alpha_1, \beta_1 \rangle, \langle \alpha_2, \beta_2 \rangle, \dots, \langle \alpha_n, \beta_n \rangle\}$ and pair of complex symbols $\langle \gamma, \delta \rangle$, the function is defined on $\langle \gamma, \delta \rangle$ as in (U288).

- (U288) $c_1 \wedge c_2 \wedge \dots \wedge c_n$ where
 $c_i = \neg(\alpha_i \subseteq \delta \wedge \beta_i \subseteq \gamma)$
 for $1 \leq i \leq n$

This condition prevents a superset of β_i from preceding a superset of α_i for all i . However, nothing prevents a fictitious rule of the form (U289).

- (U289) $+PRONOUN < -ACC$

Since German has verbs with two accusative objects (e.g., *lehren*), the rule would rule out any occurrence of two pronominalised sister objects. Therefore, Uszkoreit redefines the LP function as in (U290).

- (U290) $c_1 \wedge c_2 \wedge \dots \wedge c_n$ where
 $c_i = (\alpha_i \subseteq \delta \wedge \beta_i \subseteq \gamma) \rightarrow (\alpha_i \subseteq \gamma \wedge \beta_i \subseteq \delta)$
 for $1 \leq i \leq n$

This means that γ can precede δ if all the LP conditions are true.

Now we come to the real heart of Uszkoreit's proposed modification to the LP component for the treatment of German. He introduces the notion of a *disjunctive*

LP constraint. This is notated as in (U291) where each member of the set is called an *atomic LP rule*. The interpretation is that the disjunctive constraint is satisfied if any one of the atomic LP rules are satisfied.

$$(U291) \quad \left\{ \begin{array}{l} +NOM < +DAT \\ +NOM < +ACC \\ +DAT < +ACC \\ -FOC < +FOC \\ +PRO < -PRO \end{array} \right\}$$

If a complex LP rule $\{\langle\alpha_1, \beta_1\rangle, \langle\alpha_2, \beta_2\rangle, \dots, \langle\alpha_m, \beta_m\rangle\}$ is applied to an order pair of categories $\langle\gamma, \delta\rangle$ then the complex LP rule is interpreted as an LP condition of the following form:

$$(2.4) \quad ((\alpha_1 \subseteq \delta \wedge \beta_1 \subseteq \gamma) \vee (\alpha_2 \subseteq \delta \wedge \beta_2 \subseteq \gamma) \vee \dots \vee (\alpha_m \subseteq \delta \wedge \beta_m \subseteq \gamma)) \rightarrow ((\alpha_1 \subseteq \gamma \wedge \beta_1 \subseteq \delta) \vee (\alpha_2 \subseteq \gamma \wedge \beta_2 \subseteq \delta) \vee \dots \vee (\alpha_m \subseteq \gamma \wedge \beta_m \subseteq \delta))$$

Then (U292a) and (U292b) are grammatical and (U292c) is ungrammatical.

- (U292)
- a Dann will der Doktor
then wants the doctor [-FOCUS, +NOM]
dem Mann die Pille [-FOCUS, +ACC] geben
the man [+FOCUS, +DAT] the pill give
'Then the doctor wants to give the man the pill'
- b Dann will der Doktor die Pille
then wants the doctor [-FOCUS, +NOM] the pill [-FOCUS, +ACC]
dem Mann geben
the man [+FOCUS, +DAT] give
- c?? Dann will der Doktor die Pille
then wants the doctor [-FOCUS, +NOM] the pill [+FOCUS, +ACC]
dem Mann geben
the man [-FOCUS, +DAT] give

(U292a) is grammatical since the atomic rules +NOM < +DAT, +DAT < +ACC and +NOM < +ACC are satisfied (even though there is a violation of the atomic LP rule -FOCUS < +FOCUS). (U292b) is grammatical since the atomic rules +NOM < +ACC and -FOCUS < +FOCUS (twice) are satisfied (even though there is a violation of the atomic rule +DAT < +ACC). (U292c) is ungrammatical since both the atomic rules +DAT < +ACC and -FOCUS < +FOCUS are violated with respect to the pair of NPs *die Pille* and *dem Mann*. (Notice that (U292b) and (U292c) are string equivalent. They differ only in the values assigned to the FOCUS feature.)

Thus, the problem of conflicting ordering statements is partially solved. If each pair of categories in a constituent that stand in the ordering relation satisfy one atomic clause of each complex disjunctive LP rule, then the constituent is considered to be well-ordered. However, there is a severe drawback to this approach. It is possible that a very weak ordering constraint can be used to generate an ordering which should be ruled out by stronger ordering constraints which could apply. The problem is that the complex LP rules simply create a binary disjunctive condition when it is clear that what is required is a systematic approach to preferences over LP constraints. Let us consider an example. In addition to the other atomic LP rules in (U291), assume also the atomic LP rule $+DEF < -DEF$. That is, definite NPs precede indefinite NPs. Then we can consider the following permutation of *es ihm jemand zu lesen versprochen hat*.

- (2.5) ??jemand ihm es zu lesen versprochen hat
 someone him it to read promised has
 ‘someone has promised him to read it’

This example is considered grammatical by (U291) since *jemand* is +NOM and *ihm* is +DAT and *es* is +ACC. However, the definiteness constraint is far stronger than the case rules so *es ihm* should precede *jemand*. Secondly, *es* almost always precedes *ihm*. It almost has the status of an ad hoc rule. Therefore, we need an approach which is not just disjunctive but takes the relative strength of the LP constraints into consideration. We shall present an approach which addresses this problem in §2-10.

2-5.3 *Linear Precedence in Discontinuous Constituents: Complex Fronting in German*

In his paper, *Linear Precedence in Discontinuous Constituents: Complex Fronting in German* ([44]), Uszkoreit addresses the problem of linear order amongst “complex” verb phrase topics. “Complex fronting” as it is called involves the fronting of a verb with none or some of its complements and adjuncts (but never all of them). Uszkoreit’s basic concerns are first, that the order of elements in topic position is the same as the order they would exhibit in the Mittelfeld and second, that binary branching structures are inadequate to describe complex fronting since the binary branching structures makes the application of LP rules impossible. The paper further develops the work in his thesis, [45]. An example of the kind of topicalisation that the fragment in [45] can cover is example (5) whose phrase structure tree is shown in Fig. 2-7.

- (5) Den Brief sollte der Kurier nachher einem Spion zustecken
 the letter should the courier later a spy slip
 ‘The courier was later supposed to slip a spy the note’

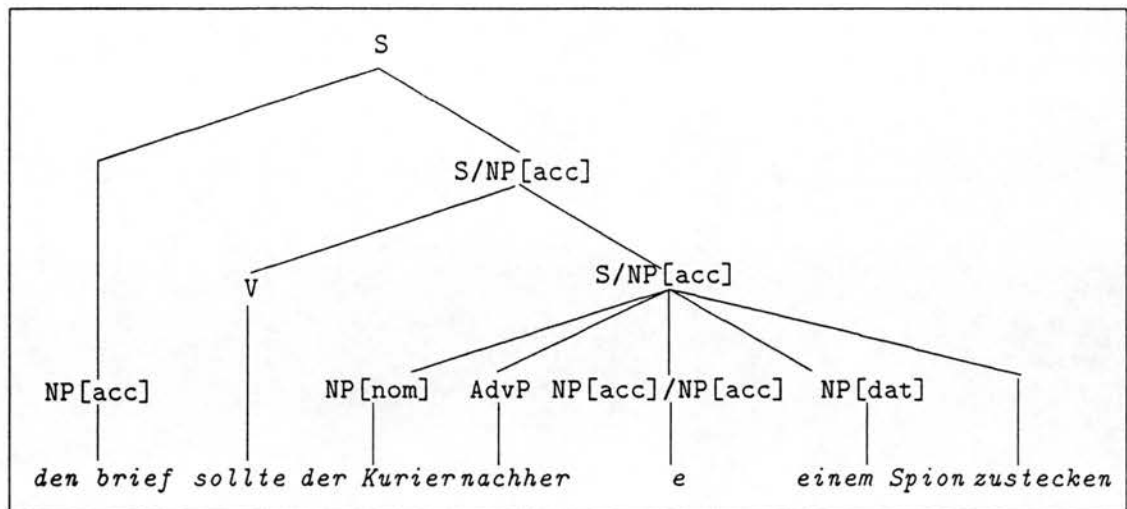


Figure 2-7: Phrase structure tree for *Den Brief sollte der Kurier nachher einem Spion zustecken*

In addition, [45] can cover all of the examples in (11) since the head verb, complements and any adjuncts are treated as sisters.

- (11)
- a **den Brief** sollte der Kurier nachher einem Spion zustecken
 - b **der Kurier** sollte den Brief nachher einem Spion zustecken
 - c **einem Spion** sollte der Kurier den Brief nachher zustecken
 - d **nachher** sollte der Kurier den Brief einem Spion zustecken
 - e **zustecken** sollte der Kurier den Brief nachher einem Spion

(12) indicates that a verb can be fronted with some of its arguments or adjuncts. In (12a), the verb *zustecken* is fronted with its object *den Brief*, in (12b), with its indirect object *einem Spion* and in (12c), with its indirect object *einem Spion* and with the adjunct *nachher*. These are examples of complex fronting.

- (12)
- a **den Brief zustecken** sollte der Kurier nachher einem Spion
 - b **einem Spion zustecken** sollte der Kurier nachher den Brief
 - c **nachher einem Spion zustecken** sollte der Kurier den Brief

At this point, Uszkoreit argues that a binary branching approach is impossible because the phrase structure trees will not have the appropriate typology to allow the correct application of the LP rules.

(18) and (19) are meant to indicate that the order of elements in a subsequence of verb, complements and adjuncts in topic position is exactly the same as if the same subsequence appeared in the *Mittelfeld*. (19a) is a topicalised variant of (18a) (with the topic indicated in bold font) and (19b) is a topicalised variant of

(18b).

- (18) a Der Kurier sollte ihm einen geheimen Brief zustecken
 the courier should him a secret note slip
 'The courier was supposed to slip him a secret note'
- b ?Der Kurier sollte einen geheimen Brief ihm zustecken

- (19) a **Ihm einen geheimen Brief zustecken** sollte der Kurier
- b ?**Einen geheimen Brief ihm zustecken** sollte der Kurier

In both cases, the topic in the (b) example is a suffix of the corresponding (a) example.

(20) and (21) are meant to show that the order in the fronted variants is the same as the order in the Mittelfeld variants even if the entire VP is not fronted.

- (20) a Darum hatte der Spion ihm dem Kurier aus der Tasche
 therefore had the spy him the courier out the pocket
 gezogen
 slipped
 'Therefore the spy had slipped it out of the courier's pocket'

b ?Darum hatte der Spion dem Kurier ihn aus der Tasche gezogen

- (21) a **Dem Kurier aus der Tasche gezogen** hatte der Spion ihn
- b ?**Ihn aus der Tasche gezogen** hatte der Spion dem Kurier

To overcome these problems, Uszkoreit proposes to allow the valency list of lexical verbs to specify all of its complements and free adjuncts. In general, this will define a regular language. Then LP rules will apply to these "uninstantiated" verbs to provide a fully instantiated valency list (subcategorisation list) which is completely ordered. This list is then used with a binary branching rule scheme to achieve the order specified in the list. (2.6) is the uninstantiated lexical entry for *erzählt*.

- (2.6)
$$\left[\begin{array}{ll} \text{V} & \\ \text{MC} & + \\ \text{VFORM} & \text{FIN} \\ \text{PER} & 3 \\ \text{PLU} & - \\ \text{FOC} & - \\ \text{LEX} & \text{erzählt} \\ \text{SEM} & \text{erzählen}' \end{array} \right]$$

The subcategorisation sequence for *erzählt* might then be (23).

$$(23) \quad \left\langle \begin{array}{l} \text{NP} \\ \text{CASE} \quad \text{ACC} \\ \text{TR} \quad \text{THEME} \\ \text{PERSON} \quad 3 \\ \text{PLURAL} \quad - \\ \text{FOCUS} \quad - \\ \text{PRO} \quad + \end{array} \right\rangle, \left\langle \begin{array}{l} \text{NP} \\ \text{CASE} \quad \text{NOM} \\ \text{TR} \quad \text{AGENT} \\ \text{PERSON} \quad 3 \\ \text{PLURAL} \quad - \\ \text{FOCUS} \quad - \\ \text{PRO} \quad - \end{array} \right\rangle, \left\langle \begin{array}{l} \text{NP} \\ \text{CASE} \quad \text{DAT} \\ \text{TR} \quad \text{GOAL} \\ \text{PERSON} \quad 3 \\ \text{PLURAL} \quad - \\ \text{FOCUS} \quad + \\ \text{PRO} \quad - \end{array} \right\rangle$$

The rule schema is simply

$$s[\langle \alpha_1, \dots, \alpha_n \rangle] \rightarrow \alpha_0, s[\langle \alpha_0, \alpha_1, \dots, \alpha_n \rangle]$$

(24) and (25) are given as further evidence in support of a binary branching analysis. We shall see below however, that both examples are actually counterevidence to a simple binary style analysis without functional composition.

- (24) a weil er sie hätte sehen können
 since he her had see could
 'since he could have seen her'

b ?weil er sie sehen gekonnt hätte

- (25) a weil er sie hätte aus den Augen verlieren können
 since he her had out the eyes loose could
 'since he could have lost sight of her'
- b weil er sie hätte nach Hause begleiten dürfen
 since he her he had home accompany may
 'since he could have been allowed to accompany her home'
- c weil er in diesem Buch hätte eine Antwort finden können
 since he in this book had an answer find could
 'since he could have found an answer in this book'

The problem with (24) is that the complement *sie* 'her' of *sehen* is separated from *sehen* by the finite auxiliary *hätte*. Presumably, the binary branching account will create a structure for the verbs that is schematically [_S *hätte* [_{VP} [_{VP} *sehen*] *können*]]. But then some explanation is necessary for how *sie* combines with *sehen*.

There are similar problems for each of the examples in (25). In (25a), the complement *sie* of *verlieren* is separated from *verlieren* by the string *hätte aus den*

Augen, in (25b), *sie* is separated from *begleiten* by *hätte nach Hause* and in (25c), the complement *diesem Buch* is separated from *finden* by the sequence *hätte eine Antwort*. Therefore, a simple binary analysis is impossible without functional composition.

Uszkoreit's proposals seem to be motivated by the recognition of the need for a binary branching approach which still allows the LP rules to be applied effectively. That is, the binary branching analysis seems to be assumed as a given. As we will argue below, the binary branching approach is empirically inadequate. One other comment needs to be made however. Whatever syntactic structure one assumes, LP rules will apply to constituents of the topic position just as they would apply to the constituents of the Mittelfeld. That is, if the same elements are sisters in topic position and in the Mittelfeld then their order will be the same. Thus we needn't do anything to guarantee that they are ordered in the same way as Uszkoreit endeavours to ensure. This throws quite some doubt on the need to consider lexical valency specifications as regular expressions over complements and possible free adjuncts.

There is one further serious complication for this approach. Complex fronting is not limited to a verb and none or some of its arguments. As Netter points out ([28]) complex fronting can also involve an object control verb and the verb it governs as in (KN102) and (KN103).

(KN102) *zu lesen gebeten hat er ihn es nicht*
 to read asked has he him it not
 'he didn't ask him to read it'

(KN103) *zu lesen empfohlen hat er ihm alle die Bücher, die ...*
 to read recommended has he him all those books which ...
 'He recommended to him to read all those books which ...'

In (KN102), the verb sequence *zu lesen gebeten* is fronted leaving behind the complement *es* of *zu lesen* and the complement *ihn* of *gebeten* in the Mittelfeld. In (KN103), the verb sequence *zu lesen empfohlen* is fronted leaving behind the complement *die Bücher* of *zu lesen* and *ihm* of *empfohlen*.

The problem for Uszkoreit's account (and binary branching accounts in general) is that there is no binary branching structure possible which will allow both verbs to be fronted without the complements of either. Rather the only possible way out seems to be an analysis which "moves" the material "left behind" to the finite clause followed by topicalisation of the VP dominating *zu lesen gebeten* in (KN102) and *zu lesen empfohlen* in (KN103).

There is one last further empirical comment to make which is related to the previous comment. Uszkoreit's approach fails entirely to account for cases where we would say that domain union is evident. Consider yet again *es ihm jemand zu lesen versprochen hat*. There is simply no way to generate such a clause in Uszkoreit's framework. Uszkoreit's proposals are all stated in terms of a single verb.

its complements and possible adjuncts. No consideration is given for complement verbs which may also take complements.

2-6 Head-Driven Phrase Structure Grammar

Pollard and Sag present a unification-based theory of syntax and semantics in their book *Information-based Syntax and Semantics* ([31], henceforth P&S1). It is not my purpose to present an overview of their theory or of the formalism which they employ. I will have to assume familiarity with P&S1. Rather, I will review the treatment of inverted clauses and second, the treatment of constituent order. Both of these will figure heavily in Ch. 3.

2-6.1 Inverted Clauses

(PS1:296) presents an inventory of inverted clause structures in English. (PS1:296a) is a yes-no question, (PS1:296b) is a *wh*-question, (PS1:296c) is a negative adverb preposing clause and (PS1:296d) is an exclamatory clause.

- (PS1:296) a Is Dana walking to the store?
 b Whose brother will Lou visit?
 c Never have I seen a taller tree.
 d Did she ever ace the test!

As in GKPS, P&S1 assume that subject-auxiliary clauses in English have a syntactic structure where the clause immediately dominates the head verb, the subject and the complement of the verb. LP constraints of the type used in GPSG then order the constituents into the correct order.

Rather than use the integer-valued **SUBCAT** feature of GPSG, HPSG uses a **SUBCAT** feature which takes a sequence of signs as its value. Syntactic head-complement rules remove some of the elements from the **SUBCAT** list and add them to the value of the **DTRS** attribute. Therefore, HPSG can treat inverted clauses by requiring that an inverted lexical head verb combine with all of its complements at once. This is precisely what Rule 3 (PS1:297) accomplishes. In schematic form, this is (PS1:298).

(PS1:297) Rule 3

$$\left[\begin{array}{l} \text{SYN|LOC|SUBCAT } \langle \rangle \\ \text{DTRS|HEAD-DTR|SYN|LOC} \left[\begin{array}{l} \text{HEAD|INV } + \\ \text{LEX } + \end{array} \right] \end{array} \right]$$

(PS1:298) $[\text{SUBCAT } \langle \rangle] \rightarrow \text{H}[\text{INV } +, \text{LEX } +], \text{C}^*$

Since every lexical sign specified as [INV +] is a finite auxiliary verb and the fact that auxiliary verbs always have **SUBCAT** sequences of length two, it follows that inverted phrasal signs will always be of the form (PS1:299).

$$(2.7) \quad \left[\begin{array}{l} \text{PHON } \boxed{4} \\ \text{SYN|LOC } \left[\begin{array}{l} \text{HEAD } \boxed{1} \\ \text{SUBCAT } \langle \boxed{2}, \boxed{3} \rangle \\ \text{LEX } + \end{array} \right] \end{array} \right]$$

$$\left[\begin{array}{l} \text{MAJ } \quad \text{V} \\ \text{VFORM } \quad \text{FIN} \\ \text{INV } \quad + \\ \text{AUX } \quad + \end{array} \right]$$

$$(PS1:299) \quad \left[\begin{array}{l} \text{PHON } \langle \boxed{4}, \boxed{5}, \boxed{6} \rangle \\ \text{SYN|LOC } \left[\begin{array}{l} \text{HEAD } \boxed{1} \\ \text{SUBCAT } \langle \rangle \\ \text{LEX } - \end{array} \right] \\ \text{DTRS } \left[\begin{array}{l} \text{HEAD-DTR } \boxed{0} \\ \text{COMP-DTRS } \langle \boxed{2}[\text{PHON } \boxed{6}], \boxed{3}[\text{PHON } \boxed{5}] \rangle \end{array} \right] \end{array} \right]$$

2-6.2 Principles of Constituent Order

Constituent order is determined by the *Constituent Order Principle* (PS1:320). It makes use of a functional dependency order-constituents. Order-constituents' value is a disjunction of the possible permutations of the daughter constituents which are consistent with the LP statements.

(PS1:320) Constituent Order Principle

$$\text{phrasal-sign } [] \Rightarrow \left[\begin{array}{l} \text{PHON } \text{order-constituents}(\boxed{1}) \\ \text{DTRS } \boxed{1} \end{array} \right]$$

For our purposes, we are only interested in two LP statements. The first, LP1. (PS1:326) requires that a lexical head precede all other sister constituents. This covers the case of inverted clauses since the verb, as lexical head, will be forced to precede the subject (since it is considered a complement) and the complement.

(PS1:326) Linear Precedence Constraint 1 (LP1)

$$\text{HEAD}[\text{LEX } +] < []$$

An example of the proper ordering of these structures is given in Fig. 2-8.

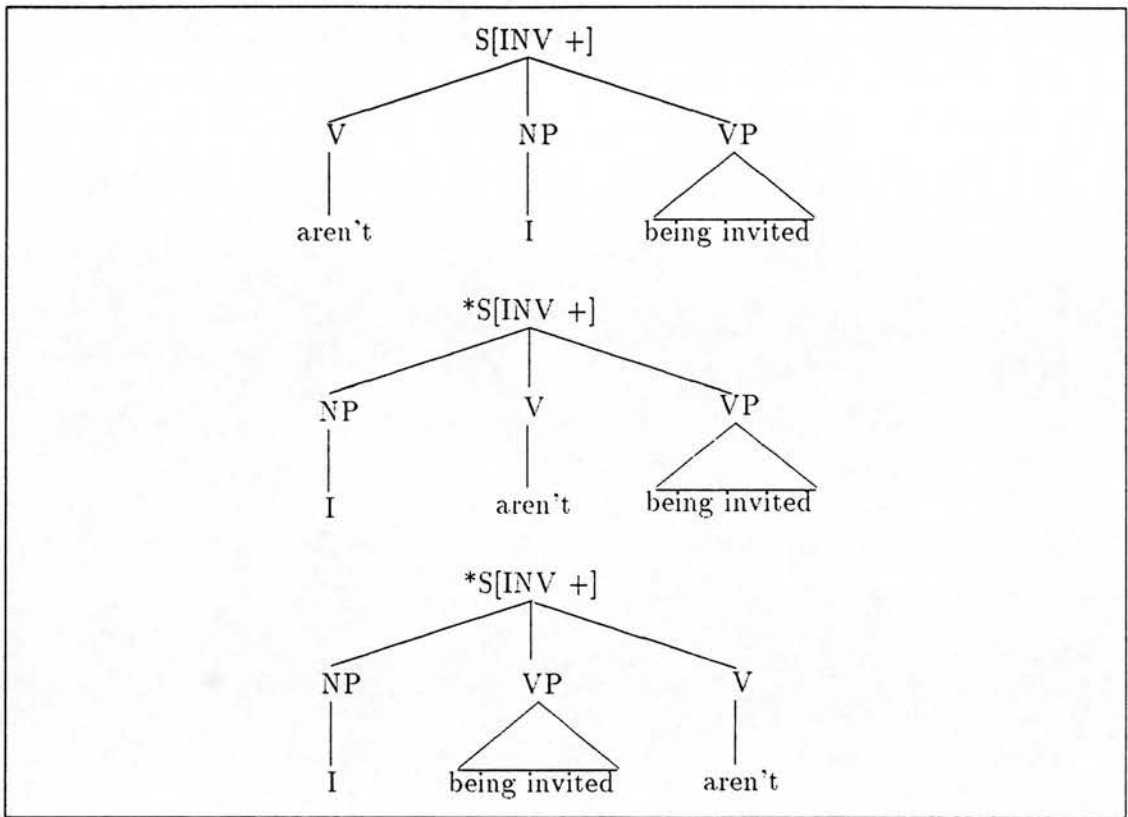


Figure 2-8: P&S1 Fig. 325

Finally, we are interested in specifying that the filler in filler-gap constructions (such as German V2 clauses) precedes the gapped constituent, (PS1:371) is required. It states that a filler must precede a nonlexical head (the constituent with the gap).

(PS1:371) Filler-Head Rule (LP4):
 FILLER < HEAD[LEX -]

2-7 Bech (1955)

Gunnar Bech's *Studien über das Deutsche Verbum Infinitum* ([3]) is the classic study of infinitival constructions and verb raising and extraposition in Standard German although he does not use these terms and presents his work in what would now be considered a pre-theoretical and pre-modern grammar framework. In this section, I will show that in many respects the distinctions I make parallel those in Bech's work and may even be considered an implementation of his ideas. Of course, I extend his ideas in various directions. The exposition quotes Evers ([13, pp52-55]) liberally as he gives a very good overview of the salient points of Bech's work. Evers makes the claim that "... Bech (1955) analysed the V-raising phenomena in approximately the way proposed in the preceding sections."

2-7.1 *status*

Bech introduces twenty major distinctions in his discussion of German infinitival constructions. The first of these is what he calls *status*.

"Bech (1955, 12, 19) starts out drawing a morphological distinction between finite and nonfinite verb forms. Subsequently he divides the nonfinite forms into those with the distribution and morphological variation of adjectives and the supinal forms. These are:

	∅	lieb	en	(infinitive without <i>zu</i>)
(157)	zu	lieb	en	(infinitive with <i>zu</i>)
	ge	lieb	t	(uninflected past participle)

The morphological variation of the supinal forms is indicated by Bech as 'status variation'. The choice of 'status' in verb complements is dependent on the matrix verb. Therefore, the choice of 'status' may be said to correspond to the choice of a complementizer element c_2 , cf. section 1.3."

Status obviously corresponds to the verb form that the head verb subcategorises for, whether it is a bare infinitival (INF), a *zu*-infinitival (ZU) or a past participle (PSP).

2-7.2 *oriëntation*

Bech also makes a distinction which he calls *orientation*. In modern terms, this corresponds to the coindexing between the NP which is controlled by a control verb and the subject of the verb that it is the functional subject of.

“Bech (o.c., 32) further observes that the supinal forms of a verb complement is oriented to a constituent in the matrix clause. The relevant constituent that orients the complement may be the subject, the object, or the indirect object of the matrix construction, depending on the kind of matrix verb. Therefore Bech (o.c., 32) assigns to each matrix verb a coefficient to indicate which of the matrix constituents is relevant to the interpretation of the complement, especially the reflexives and reciprocals in the complement. This is clearly a parallel to the control problem of Equi-NP-Deletion, cf. Rosenbaum (1967) and Jackendoff (1972, 210).”

A verb is *oriented* to a constituent NP if the NP is the logical subject of the controlled verb.

2-7.3 *Verbalfeld*

Bech also introduces the notion of a *Verbalfeld* ‘verb field’. Roughly speaking, a *Verbalfeld* is a verb with all of its syntactically realised subcategorised elements except complement VPs or Ss.

“Each verb in finite or supinal form belongs to a set of sentence parts that is dependent on that verb. The verb together with its dependents constitute the *Verbalfeld* (o.c., 43).

In transformational terms one would say that the dependents of a verb are the constituents that subcategorize it. The constituents that subcategorize the verb are, nevertheless, explicitly excluded by Bech from the *Verbalfeld* of that verb. These are (i) the constituent that orients a supinal form – this constituent does not belong to the *Verbalfeld* of the supinal form; (ii) the verb that derives its *status* from a matrix verb – this verb does not belong to the *Verbalfeld* of the matrix verb.

Considering this, one can say that Bech’s *Verbalfeld* coincides with the set of deep structure clausemates.”

Evers quote is slightly confusing but it amounts to the following. The NP that orients a verb is excluded from the verb’s *Verbalfeld*. It is as if Bech assumes an NP-VP analysis for verbs of perception and control verbs. The NP objects of these verbs are not part of the *Verbalfeld* of the head of the VP. Furthermore.

a verb that derives its status from a verb is not part of the verb's Verbalfeld. In other words, the head of a VP subcategorised by a verb is not part of that verb's Verbalfeld. Then, if we consider the example *es ihm jemand zu lesen versprochen hat*, *es zu lesen* is a Verbalfeld, *ihm versprochen* is a Verbalfeld and *jemand hat* is a Verbalfeld. The comparison with our approach is obvious. The Verbalfeld is a verb plus all the constituents it subcategorises for minus any VP subcategorisation.

2-7.4 *kohärenzfeld*

Bech also introduces the notion of a *kohärenzfeld* 'coherence field' which is similar to our domains.

"A *kohärenzfeld* 'coherence field' consists of two or more *Verbalfelder* 'verb fields'. All verbs that belong to the same *kohärenzfeld* are placed at the end of the *kohärenzfeld* in a distributional unit called the *verbale schlussfeld* 'verbal end field'. This unit corresponds to the V-cluster that results from V-raising.

The *kohärenzfeld* minus the *schlussfeld* is the *restfeld* 'rest field'.

Bech presents (158) as an examples of a *kohärenzfeld* that consists of two *Verbalfelder*, F_1 and F_2 .

- (158) daß ich ihm nicht helfen konnte
 that I ihm not help could
 'that I couldn't help him'

schlussfeld : (helfen konnte)

F_1 : daß ich — nicht — konnte

F_2 : — — ihm — helfen —

It is obvious from (158) that a *Verbalfeld* is not necessarily a unit in the surface structure. In Bech's own words ([3, p61]):

"(das kohärenzfeld) bildet in topologischer hinsicht eine geschlossene einheit, ... aber innerhalb eines kohärenzfeld stehen die glieder der verschiedenen Verbalfelder zwischen einander."

the coherence field forms a closed unit topologically, ... but within a coherence field members of the different verbal fields can occur between each other.

The coherence field can be defined in transformational terms as the set of surface structure clausemates."

Clearly the coherence field is the equivalent of one of our domains and the verb fields are the equivalent of functor-argument structure sisters (minus a possible VP subcategorisation). Bech's decomposition of a coherence field into two or more verb fields corresponds to the decomposition of domains into daughter domains. For example if we consider *es ihm jemand zu lesen versprochen hat*, the clausal domain minus the coherence field (domain) of its VP complement is the verb field *jemand hat*. Similarly, the VP domain of *es ihm zu lesen versprochen* minus the coherence field (domain) of its VP complement is *ihm versprochen*. Finally, the coherence field of the VP *es zu lesen* is just *es zu lesen*. Thus there is a clear correspondence between the decomposition of coherence fields into verb fields and our decomposition of domains into constituent domains.

Bech also presents some interesting data concerning possible verb cluster order in Standard German which will be relevant in Ch. 3. Therefore, I present the following quote.

“Although this is not immediately relevant to the transformational reinterpretation of Bech (1955), I would like to draw attention to a highly interesting set of order variations within the German verb cluster, observed by Bech (o.c. 63) and presented by him in the following scheme:

(159)

	1	2	3	4	5
0	V ₁	V ₂ V ₁	V ₃ V ₂ V ₁	V ₄ V ₃ V ₂ V ₁	V ₅ V ₄ V ₃ V ₂ V ₁
1			V ₁ V ₃ V ₂	V ₁ V ₄ V ₃ V ₂	V ₁ V ₅ V ₄ V ₃ V ₂
2				V ₁ V ₂ V ₄ V ₃	V ₁ V ₂ V ₅ V ₄ V ₃
3					V ₁ V ₂ V ₃ V ₅ V ₄

The columns in (159) represent verb clusters of respectively one, two, three, four and five verbs. The index under each V indicates the degree of subordination. The V₁ can be finite or nonfinite (o.c. 64); the rows in (159) show types in the left-right order of verbs. The top row, indexed as 0, presents the verbs in the order predicted by the German variant of (125). Each cluster in the lower rows of (159) preserves its two most embedded verbs in this order. The other verbs may appear in the ‘Dutch’ order. Bech calls the part in the ‘Dutch’ order (1 2 3 ...) *oberfeld* and the part in ‘German’ order (5 4 3 ...) *unterfeld*. The first part of the *unterfeld* is stressed, whereas the *oberfeld* is unstressed. The *oberfeld* can only contain verbs that have ‘<tense>’ or ‘∅’ as second complementizer element (o.c., 63). Bech exemplifies column 3 and 4 with quotes from contemporary German literature – his standard way of providing examples – and constructed the following example for column 5 (o.c., 64):

(160)

dass man ihn hier	$\left\{ \begin{array}{l} \text{liegen bleiben lassen können wird} \\ \text{wird liegen bleiben lassen können} \\ \text{wird können liegen bleiben lassen} \\ \text{wird können lassen liegen bleiben} \end{array} \right\}$
-------------------	--

The order of verbal forms in Dutch is a fully rigid $V_1V_2V_3\dots$ etc. The only exception is that for some speakers of Dutch a cluster of two verbs V_1V_2 allows a variant V_2V_1 , if the V_1 is a finite auxiliary or modal verb."

Finally, Bech explicitly defends the view that modal and auxiliary verbs act as main verbs as is now widely assumed. Specifically, he defends the point of view that modals and auxiliaries head their own verb fields just as we assume that modals and auxiliaries head their own domains.

"Finally, it may be noted that Bech (o.c. 87–88) explicitly defends the view that modal and auxiliary constructions are composed of two *Verbalfelder* – in transformational terms: the view that they are derived from a (1a) structure. The two *Verbalfelder* are united in a *kohärenzkonstruktion* – in transformational terms: they form a (1b) structure. The reason for the *kohärenzkonstruktion* is, according to Bech, the fact that the verb that is constructed with an auxiliary or a modal verb invariably takes the supinal form without *zu*. Bech (o.c. Chs. XI and XV) points out that other verbs that have a complement based on a supinal form without *zu*, also also invariably end up in the *kohärenzkonstruktion*, e.g., *sehen*, *hören*, *helfen* and *lehren*. This corresponds with the regular cases of obligatory V-raising."

2–8 Evers (1975)

Evers ([13]) argues on the basis of gapping, nominalisation, extraposition, relative clause shift, extraposition from NPs, PP shift, clitic placement, quantifier float, clause negation, S-pronominalisation, reflexivisation and passive that Dutch and German clauses require a (1a) structure (Fig. 2–9) at deep structure and a (1b) structure (Fig. 2–10) at surface structure. This is formulated as a rule of *V-raising* which deletes the embedded S node, merges the complements of the embedded S with the matrix S and "raises" the embedded verb to be sister to the matrix verb. This rule applies equally to both German and Dutch clauses, even in those cases where the German derivation results in string vacuous movement. Below, I will give examples of V-raising for the five classes of verbs which Evers distinguishes.

Class I consists of verbs with a sentential object in deep structure. The verb takes a bare infinitival complement verb and Equi-NP-Deletion does not apply.

Members of Class I are:

Dutch	German	gloss
<i>horen</i>	<i>hören</i>	'hear'
<i>zien</i>	<i>sehen</i>	'see'
<i>laten</i>	<i>lassen</i>	'let, make'
<i>vinden</i>	<i>finden</i>	'find'
<i>voelen</i>	<i>fühlen</i>	'touch, feel'

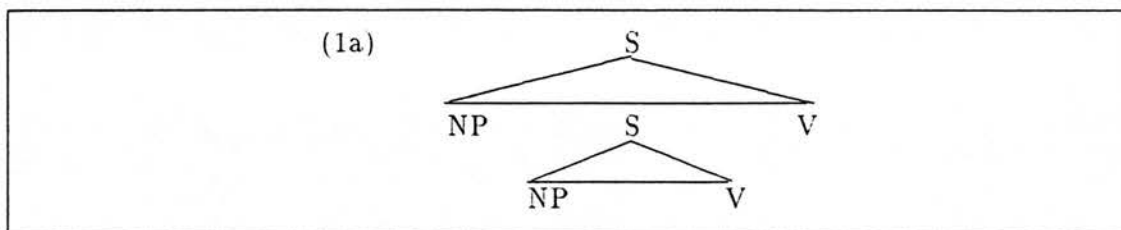


Figure 2-9: Schematic deep structure for V-raising

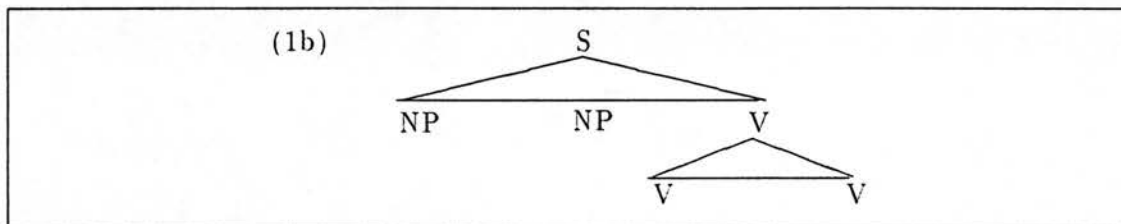


Figure 2-10: Schematic surface structure for V-raising

The trees in Fig. 2-11 show the deep and surface structures for the example *omdat Cecilia de kraanvogels zag vliegen*.

- (2.8)
- a omdat Cecilia de kraanvogels zag vliegen
 because Cecilia the cranes saw fly
 'because Cecilia saw the cranes fly'
- b weil Cecilia die Kraniche fliegen sah
 because Cecilia the cranes fly saw

Class II consists of verbs with a sentential object in deep structure. The matrix verb takes a bare infinitival complement. Equi-NP-Deletion and V-Raising apply obligatorily.

Members of Class II are:

Dutch	German	gloss
<i>leren</i>	<i>lernen</i>	'learn'
<i>leren</i>	<i>lehren</i>	'teach'
<i>helpen</i>	<i>helfen</i>	'help'
<i>willen</i>	<i>wollen</i>	'want'
<i>kunnen</i>	<i>können</i>	'can'
<i>moeten</i>	<i>sollen</i>	'shall, ought to'
<i>mogen</i>	<i>durfen</i>	'may'

The trees in Fig. 2-12 show the deep and surface structures for the example *omdat*

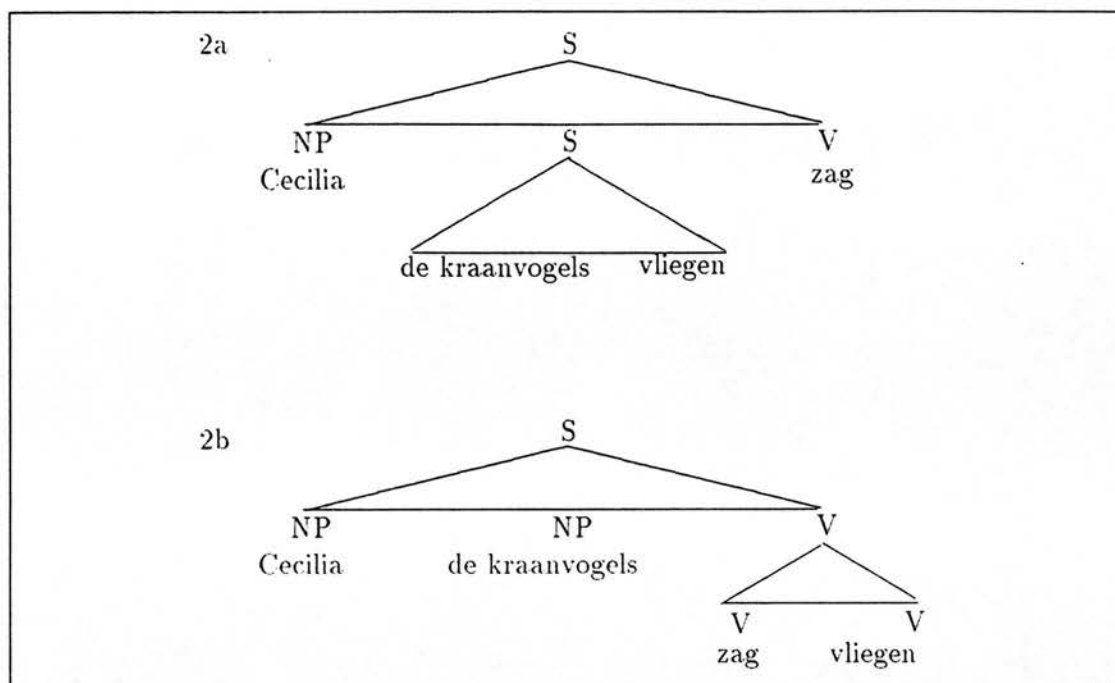


Figure 2-11: Deep and surface structure trees for *Cecilia de kraanvogels zag vliegen*

Cecilia ons de kraanvogels hielp fotograferen.

- (2.9)
- a omdat Cecilia ons de kraanvogels hielp fotograferen
because Cecilia us the cranes help photograph
'because Cecilia helps us photograh the cranes'
- b weil Cecilia uns die Kraniche fotograferen half
because Cecilia us the cranes photograph help

Class III consists of verbs with a sentential object in deep structure. The complement is either a *te* or *zu* infinitival, Equi-NP-Deletion is obligatory and V-raising is obligatory for the members of class IIIa but optional for IIIb.

The members of IIIa are:

Dutch	German	gloss
<i>believen</i>	<i>belieben</i>	'please'
<i>dienen</i>		
<i>plegen</i>	<i>pfliegen</i>	'care for'
<i>weten</i>	<i>wissen</i>	'know'
<i>durven</i>		

The members of IIIb are (among others):

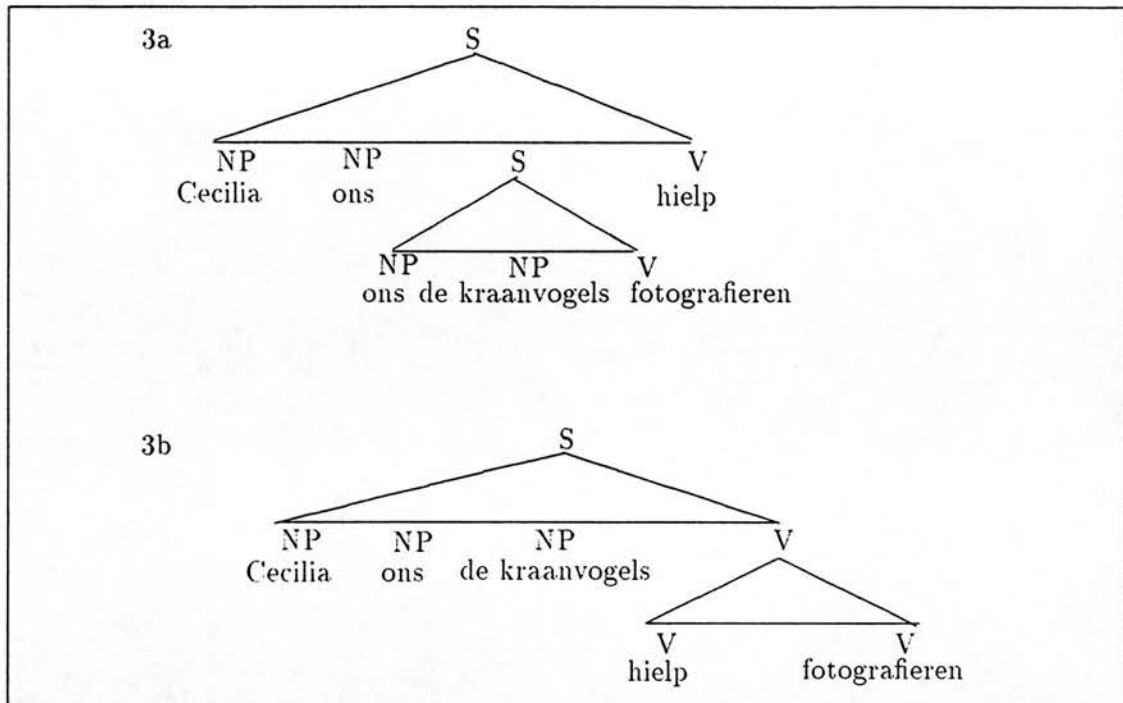


Figure 2-12: Deep and surface structure trees for *Cecilia ons de kraanvogels hielp fotograferen*

Dutch	German	gloss
<i>beogen</i>	<i>beabsichtigen</i>	'intend'
<i>besluiten</i>	<i>beschliessen</i>	'decide'
<i>beweren</i>	<i>behaupten</i>	'claim'
<i>eisen</i>	<i>fordern</i>	'demand'
<i>hopen</i>	<i>hoffen</i>	'hope'
<i>menen</i>	<i>meinen</i>	'mean, think, intend'
<i>verbieden</i>	<i>verbieten</i>	'forbid'
<i>vergeten</i>	<i>vergessen</i>	'forget'
<i>vermijden</i>	<i>vermeiden</i>	'avoid'
<i>verplichten</i>	<i>verpflichten</i>	'oblige'
<i>verzuimen</i>	<i>versäumen</i>	'miss, neglect, fail'
<i>vrezen</i>	<i>befürchten</i>	'fear'
<i>wagen</i>	<i>wagen</i>	'risk, dare'
<i>weigeren</i>	<i>verweigern</i>	'refuse'
<i>wensen</i>	<i>wünschen</i>	'wish'

The trees in Fig. 2-13 show the deep and surface structures for the example *omdat*

Cecilia de kraanvogels beweerde te filmen.

- (2.10) a omdat Cecilia de kraanvogels beweerde te filmen
 because Cecilia the cranes claimed to film
 'because Cecilia claimed to film the cranes'
- b weil Cecilia die Kraniche zu filmen behauptete
 because Cecilia the cranes to film claimed

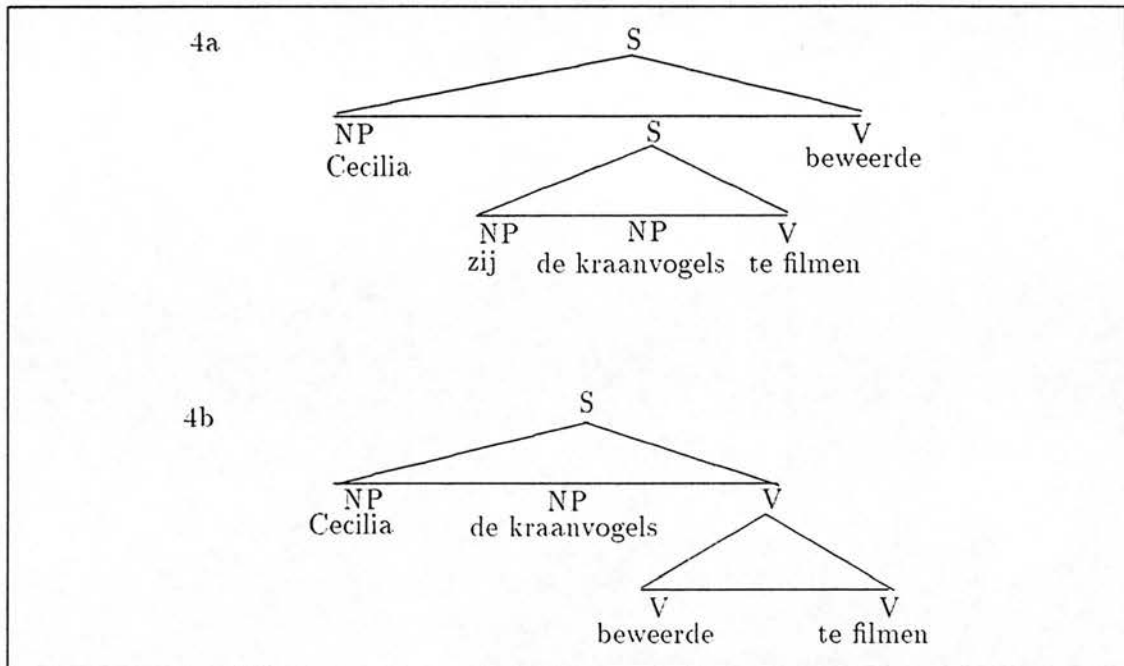


Figure 2-13: Deep and surface structure trees for *Cecilia de kraanvogels beweerde te filmen*

Class IV consists of verbs with a sentential subject in deep structure. The verb takes a bare infinitival complement. Equi-NP-Deletion does not apply. V-raising is obligatory.

The members of Class IV are:

Dutch	German	gloss
<i>kunnen</i>	<i>können</i>	'can'
<i>zullen</i>	<i>werden</i>	'become, shall'
<i>moeten</i>	<i>müssen</i>	'must'

The trees in Fig. 2-14 show the deep and surface structures for the example *omdat*

Cecilia de kraanvogels kan vergiftigen.

- (2.11) a omdat Cecilia de kraanvogels kan vergiftigen
 because Cecilia the cranes can poison
 'because Cecilia can poison the cranes'
- b weil Cecilia die Kraniche vergiften kan
 because Cecilia the cranes poison can

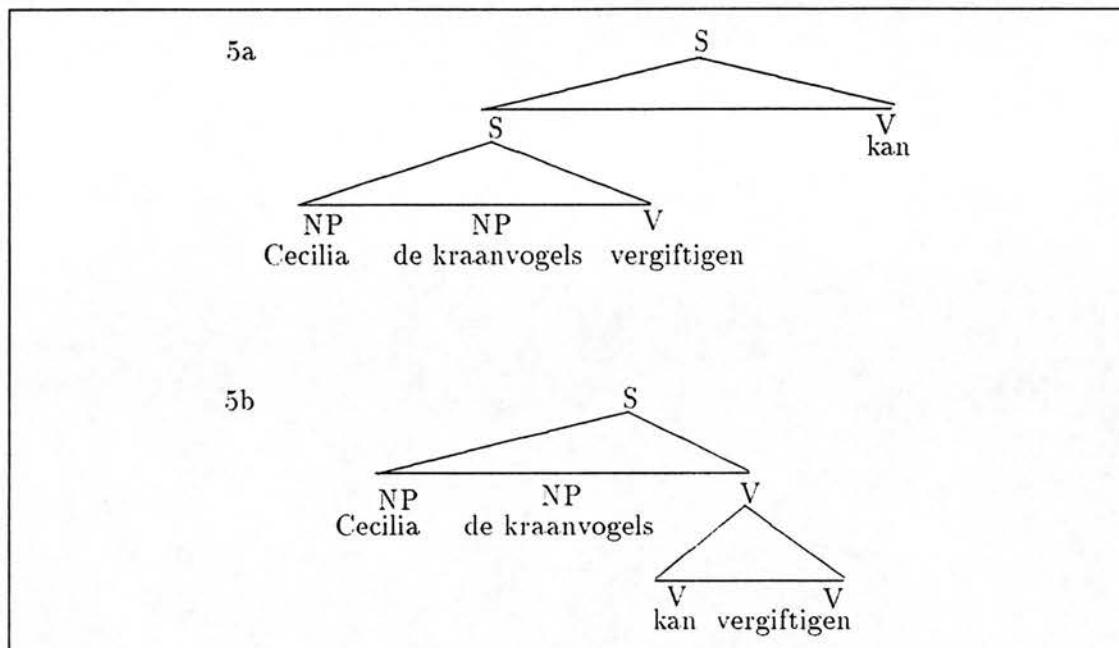


Figure 2-14: Deep and surface structure trees for *Cecilia de kraanvogels kan vergiftigen*

Class V consists of verbs with a sentential subject in deep structure. The verb takes a *te* or *zu* infinitival complement. Equi-NP-Deletion does not apply. V-raising is obligatory.

The members of Class V are:

Dutch	German	gloss
<i>beginnen</i>	<i>beginnen</i>	'begin'
<i>blijven</i>	<i>bleiven</i>	
<i>dreigen</i>	<i>drohen</i>	'threaten'
<i>schijnen</i>	<i>scheinen</i>	'seem'
<i>hebben</i>	<i>haben</i>	'have'
<i>zijn</i>	<i>sein</i>	'be'
<i>hoeven</i>	<i>brauchen</i>	'use, want, need'
<i>blijken</i>		

The trees in Fig. 2-15 show the deep and surface structures for the example *omdat het water de kraanvogels schijnt te vergiftigen*.

- (2.12) a omdat het water de kraanvogels schijnt te vergiftigen
 because the water the cranes seems to poison
 'because the water seems to poison the cranes'
- b weil das Wasser die Kraniche zu vergiften scheint
 because the water the cranes to poison seems

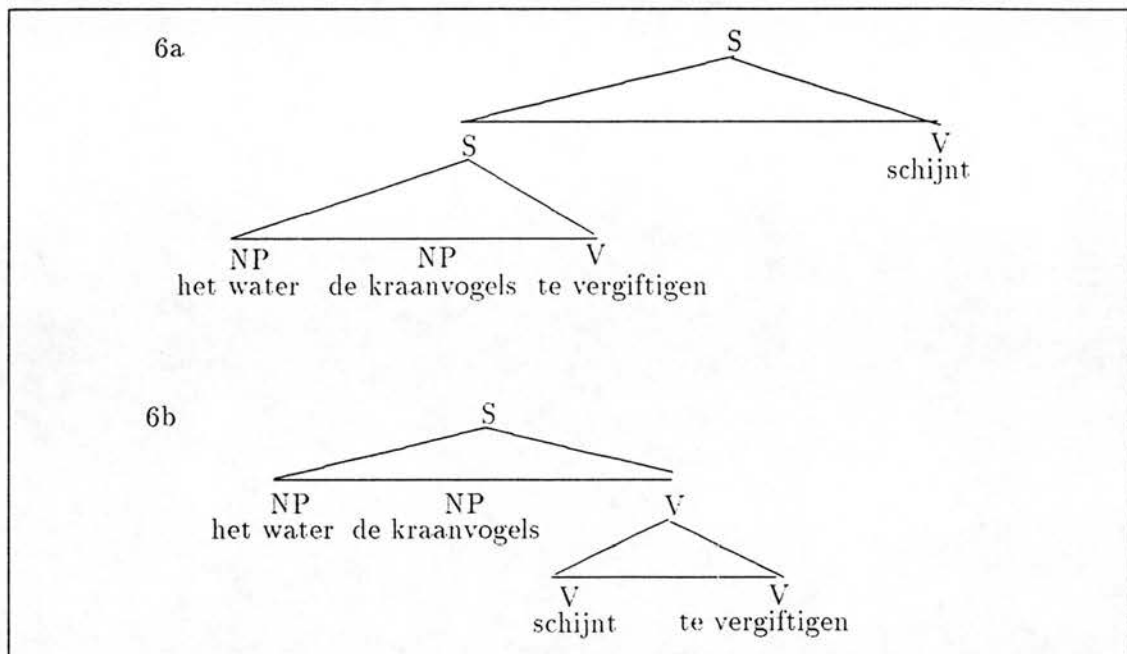


Figure 2-15: Deep and surface structure trees for *omdat het water de kraanvogels schijnt te vergiftigen*

In all such cases, the matrix verb takes a sentential complement. However, in every case a (1b) structure is created. The (1b) structures are flat except for the V node which dominates the verbs in the verb cluster. Evers hypothesises this node since the verb cluster is not interruptible by other material. He provides some theory internal argumentation to support this structure based largely on an account of gapping and nominalisation. This takes the form of arguing that the transformations involved refer to the dominating V node. In all other respects, the (1b) structures are equivalent to the domains that domain union creates for the same classes of verbs.

There are several issues which Evers does not address. One is variation among NP order. The NPs of the embedded S are always spliced into the matrix clause at the position at which the embedded S node occurred. Thus, the linear order of the NPs is fixed. Second, he provides no discussion of verb order variation

and fails to provide any sort of explanation of why Dutch governs to the right and German to the left. Third, he provides no account of VP extraposition, especially of examples like *dass er versucht hat, das Buch zu lesen* in which the extraposed VP is separated from the verb that it depends on. Fourth, Evers doesn't consider (1b) structures in the context of extraposed VPs or topicalised VPs. Presumably, he would have no trouble extending his account to include (1b) structures in peripheral position, however.

To summarise, Evers derives (1b) structures from (1a) structures in the cases of Class I to Class V verbs. In our terminology, (1a) structures are similar to the functor-argument structures which we hypothesise. (1b) structures are similar to the "final" domain structures assigned to matrix clauses. The difference in the two approaches is that Evers' account relies on the transformational rule of V-raising while ours relies on domain construction. Because of the similarities in the structures between the two accounts, most of Evers' empirical arguments for his approach apply to our approach as well. We justify our departure from Evers' approach by the fact that our theory covers more data and is a theory of discontinuous constituency in general, whereas the rule of V-raising is very specific to German and Dutch.

2-9 Domain Union

Domain union can be formalised as a ternary relation \bigcirc over sequences called *sequence union*, which is the *shuffle* operator of formal language theory.² (Cf. [21]). Let ϵ be the *empty sequence*, σ_1 , σ_2 and σ_3 be sequences and \circ the string concatenation operator. Then

$$(2.13) \quad \begin{aligned} & \bigcirc(\epsilon, \epsilon, \epsilon) \\ & \bigcirc(x \circ \sigma_1, \sigma_2, x \circ \sigma_3) \leftrightarrow \bigcirc(\sigma_1, \sigma_2, \sigma_3) \\ & \bigcirc(\sigma_1, x \circ \sigma_2, x \circ \sigma_3) \leftrightarrow \bigcirc(\sigma_1, \sigma_2, \sigma_3) \end{aligned}$$

(Note that these formulas are implicitly disjoined and that each disjunct is universally quantified. For further details, cf. App. A. Other axioms presented in this fashion are also disjoined and universally quantified.)

\bigcirc is clearly very similar to \circ . Let σ also be a sequence. Then

$$(2.14) \quad \begin{aligned} & \circ(\epsilon, \sigma, \sigma) \\ & \circ(x \circ \sigma_1, \sigma_2, x \circ \sigma_3) \leftrightarrow \circ(\sigma_1, \sigma_2, \sigma_3) \end{aligned}$$

Note however that while \circ is an operator (i.e., a *function*), \bigcirc is not. It is a relation. The definition of \circ should be very familiar to Prolog programmers. It is just the definition of `append/3`.

²In the past I have used the symbol $\cup_{()}$ for sequence union. I am abandoning it here for typographical reasons.

```

append([],L,L).
append([X|L1],L2,[X|L3]) :-
    append(L1,L2,L3).

```

Similarly, \circ can be defined in Prolog and its definition is very similar to `append/3`. It can be defined as the pure Prolog predicate `shuffle/3` which is exactly analogous to its abstract definition.

```

shuffle([],[],[]).
shuffle([X|L1],L2,[X|L3]) :-
    shuffle(L1,L2,L3).
shuffle(L1,[X|L2],[X|L3]) :-
    shuffle(L1,L2,L3).

```

Notice that the second clause is the same as the second clause of `append/3`. The third clause allows elements to be removed from the second “input” list and be added to the output list `L3` as well as be removed from the first list. The base case first clause is an instance of the base case first clause of `append/3`. It can be this specific since the predicate can recur on both the first and second input lists.

E.g., Let $A = \langle a,b \rangle$ and $B = \langle c,d \rangle$. Then $\circ(A,B,C)$ for all sequences C in (2.15).

```

(2.15)   $\langle a, b, c, d \rangle$ 
         $\langle a, c, b, d \rangle$ 
         $\langle a, c, d, b \rangle$ 
         $\langle c, d, a, b \rangle$ 
         $\langle c, a, d, b \rangle$ 
         $\langle c, a, b, d \rangle$ 

```

Informally, C contains all and only the elements of A and B and the order of the elements of A and B are preserved in C . That is, for each value of C in (2.15), $a \prec b$ as in A and $c \prec d$ as in B . Furthermore, if $C = \langle a,b,c,d \rangle$ then $\circ(A,B,C)$

is true for all A and B in (2.16).

	A	B
(2.16)	$\langle a,b,c,d \rangle$	$\langle \rangle$
	$\langle a,b,c \rangle$	$\langle d \rangle$
	$\langle a,b,d \rangle$	$\langle c \rangle$
	$\langle a,b \rangle$	$\langle c,d \rangle$
	$\langle a,c,d \rangle$	$\langle b \rangle$
	$\langle a,c \rangle$	$\langle b,d \rangle$
	$\langle a,d \rangle$	$\langle b,c \rangle$
	$\langle a \rangle$	$\langle b,c,d \rangle$
	$\langle b,c,d \rangle$	$\langle a \rangle$
	$\langle b,c \rangle$	$\langle a,d \rangle$
	$\langle b,d \rangle$	$\langle a,c \rangle$
	$\langle b \rangle$	$\langle a,c,d \rangle$
	$\langle c,d \rangle$	$\langle a,b \rangle$
	$\langle c \rangle$	$\langle a,b,d \rangle$
	$\langle d \rangle$	$\langle a,b,c \rangle$
	$\langle \rangle$	$\langle a,b,c,d \rangle$

\circ is usually written as a binary infix operator (e.g., $A \circ B$). In this case, its abstract definition is (2.17).

$$(2.17) \quad \begin{aligned} \circ(\epsilon, \sigma) &\leftrightarrow \sigma \\ (x \circ \sigma_1) \circ \sigma_2 &\leftrightarrow x \circ (\sigma_1 \circ \sigma_2) \end{aligned}$$

Notice that the second formula in the definition of \circ of arity three collapses into the associativity axiom in the definition of \circ of arity two.

\bigcirc can also be written as a binary infix relation (e.g., $A \bigcirc B$). In this case, its abstract definition is (2.18).

$$(2.18) \quad \begin{aligned} \epsilon \bigcirc \epsilon &\leftrightarrow \epsilon \\ (x \circ \sigma_1) \bigcirc \sigma_2 &\leftrightarrow x \circ (\sigma_1 \bigcirc \sigma_2) \\ \sigma_1 \bigcirc (x \circ \sigma_2) &\leftrightarrow x \circ (\sigma_1 \bigcirc \sigma_2) \end{aligned}$$

This is the form of \bigcirc which we will use in this work. Let $A = \langle a,b \rangle$ and $B = \langle c,d \rangle$ once again. Then we can derive the following biconditional.

$$(2.19) \quad \begin{aligned} A \bigcirc B &\leftrightarrow \\ &\langle a, b, c, d \rangle \vee \\ &\langle a, c, b, d \rangle \vee \\ &\langle a, c, d, b \rangle \vee \\ &\langle c, d, a, b \rangle \vee \\ &\langle c, a, d, b \rangle \vee \\ &\langle c, a, b, d \rangle \end{aligned}$$

In what follows, we assume a level of *syntactic-semantic functor-argument structure*. This functor-argument structure encodes (nonsurface) syntactic structure. However, since word order domains determine word order, this level of syntactic representation is unordered. Therefore, if we say that the phrase VP_1 is of the form $[_{VP_1} NP_1 VP_2 V_1]$ for example, the order of the daughter elements NP_1 , VP_2 and V_1 is undefined. When we refer to syntactic constituents in the sequel, it is to this level of representation that we refer.

A convenient way to think about domain union informally is in terms of bracket erasure of labelled bracketed strings. Although domains are just sequences of constituents we will usually subscript the opening bracket of the labelled bracketed string denoting the domain of constituent X with X itself for clarity. However, the subscript is purely to make the bracketed strings more legible. This notation runs the risk of confusing syntactic structures with domains since they are both represented by labelled bracketed strings, but we will be careful to indicate the difference in what follows.

If VP_1 is a phrase of the form $[_{VP_1} NP_1 VP_2 V_1]$ and we assume for the sake of argument that NP_1 precedes V_1 in the domain of VP_1 (written $D(VP_1)$) then $D(VP_1)$ can be any of the domains in (2.20) by virtue of the fact that a daughter can be an element of its mothers domain.

$$(2.20) \quad \begin{array}{l} [_{VP_1} \quad VP_2 \quad NP_1 \quad V_1 \quad] \\ [_{VP_1} \quad NP_1 \quad VP_2 \quad V_1 \quad] \\ [_{VP_1} \quad NP_1 \quad V_1 \quad VP_2 \quad] \end{array}$$

Now assume that $D(VP_2)$ is $[_{VP_2} NP_2 V_2]$. This means that NP_2 precedes V_2 in any domain which $D(VP_2)$ is domain unioned into. Informally, we can think of domain unioning $D(VP_2)$ into $D(VP_1)$ by placing $D(VP_2)$ anywhere in $D(VP_1)$, erasing the brackets $[_{VP_2}]$ and then allowing NP_2 and V_2 to "float" arbitrarily far to the left or right within $D(VP_1)$ so long as NP_2 precedes V_2 . This means that any of the labelled bracketed strings in (2.21) can be derived (which are all possible domains of VP_1 in addition to those shown in (2.20)).

$$(2.21) \quad \begin{array}{l} (a) [_{VP_1} \quad NP_1 \quad V_1 \quad NP_2 \quad V_2 \quad] \\ (b) [_{VP_1} \quad NP_1 \quad NP_2 \quad V_1 \quad V_2 \quad] \\ (c) [_{VP_1} \quad NP_1 \quad NP_2 \quad V_2 \quad V_1 \quad] \\ (d) [_{VP_1} \quad NP_2 \quad V_2 \quad NP_1 \quad V_1 \quad] \\ (e) [_{VP_1} \quad NP_2 \quad NP_1 \quad V_2 \quad V_1 \quad] \\ (f) [_{VP_1} \quad NP_2 \quad NP_1 \quad V_1 \quad V_2 \quad] \end{array}$$

Now substitute $[_{VP_2} NP_2 V_2]$ for $D(VP_2)$ in 2.20.

$$(2.22) \quad \begin{array}{l} (a) [_{VP_1} \quad [_{VP_2} \quad NP_2 \quad V_2] \quad NP_1 \quad V_1 \quad] \\ (b) [_{VP_1} \quad NP_1 \quad [_{VP_2} \quad NP_2 \quad V_2] \quad V_1 \quad] \\ (c) [_{VP_1} \quad NP_1 \quad V_1 \quad [_{VP_2} \quad NP_2 \quad V_2] \quad] \end{array}$$

Then (2.21a) will produce the same word order as in (2.22c), (2.21c) will produce the same word order as in (2.22b) and (2.21d) will produce the same word order as in (2.22a), although in the (2.21) domains, the $[VP_2]$ brackets have been eliminated while in the (2.22) domains they are preserved. This means that if a (2.22) domain is unioned into another domain then $D(VP_2)$ will necessarily be continuous in that domain whereas if a (2.21) domain is unioned into another domain, $D(VP_2)$ might appear discontinuously in it.

2-10 A theory of linear precedence constraints

2-10.1 Linear precedence constraints over word order domains

The order of elements in a domain which has been domain unioned is partially determined by the domain union relation but we have said nothing so far which determines the order of elements within a non-unioned domain. Domain order is also partially determined by *linear precedence* (LP) constraints of a form similar to those in GPSG. However, LP constraints are defined here as wellformedness conditions on word order domains (sequences) as opposed to GPSG where they are defined as wellformedness conditions on local trees. There are two types of LP constraints. A sequence σ satisfies an LP constraint of the form $\phi_1 \prec \phi_2$ iff every element of σ which satisfies ϕ_1 precedes every other element of σ which satisfies ϕ_2 . A sequence σ satisfies an LP constraint of the form $\phi_1 \preceq \phi_2$ iff every element of σ which satisfies ϕ_1 precedes (or is equal to) every element of σ which satisfies ϕ_2 . The \preceq form will become important in §3.

For example, let $A = \langle NP[DAT] v_1 \rangle$ and $B = \langle NP[ACC] v_2 \rangle$ and assume the LP constraints $NP[DAT] \prec NP[ACC]$ and $NP \prec v$. Then

$$(2.23) A \circ B \leftrightarrow \langle NP[DAT] NP[ACC] v_1 v_2 \rangle \vee \langle NP[DAT] NP[ACC] v_2 v_1 \rangle$$

LP constraints apply to *every* domain. So, although $NP[DAT] \prec NP[ACC]$ has no affect on A or B , it requires that the $NP[DAT]$ from A precedes the $NP[ACC]$ from B .

2-10.2 Partially Ordered LP constraints

As mentioned above, German (and probably all of the other West Germanic languages including English) do not have absolute LP constraints but rather "weak word order constraints" which can be overridden by other "stronger" LP constraints. For example, we know that the "weak" constraints $[NOM] \prec [DAT]$ and $[DAT] \prec [ACC]$ hold for German. However, we also know that there is a strong preference for pronouns to precede nonpronominals (i.e., $[PRO +] \prec [PRO -]$).

For example, if we consider the example *daß es ihm jemand zu lesen versprochen hat* we see that there are several “weak” word constraints in play, some of which are overridden by other stronger constraints. Some of the relevant constraints to this example are the following.

- (2.24) [PHON *es*] \prec NP
 (2.25) [PRO +] \prec [PRO -]
 (2.26) [DEF +] \prec [DEF -]
 (2.27) [NOM] \prec [DAT]
 (2.28) [NOM] \prec [ACC]
 (2.29) [DAT] \prec [ACC]

According to my intuitions, these are listed from strongest to weakest. Then we can see that the position of *jemand* in the example violates both (2.28) and (2.29) but satisfies (2.24) and (2.26) and vacuously satisfies (2.25). *ihm* violates (2.28) and (2.29) but satisfies (2.24) and vacuously satisfies (2.25) and (2.26). Finally, *es* precedes the two other NPs and satisfies (2.24) and (2.26) with respect to the indefinite *jemand*, vacuously satisfies (2.25) and violates both (2.28) and (2.29). In other words, *es* precedes the other two NPs because the LP constraint that requires that *es* precedes other NPs is the strongest constraint. It can “override” the requirements that it follow the dative indirect object and the nominative subject. This ordering is “aided” by the fact that *jemand* is indefinite while *es* is prototypically definite. Similarly, *ihm* precedes *jemand* since it is definite and *jemand* is indefinite even though this overrides (2.28).

This suggests that we should be able to order LP constraints on the basis of their relevant strengths, or to put it another way, on the basis of preferences. Uszkoreit ([45]) offers a list of LP constraints “ordered by weight” listed in (2.30)–(2.34).

- (2.30) [PRO +] \prec [PRO -]
 (2.31) [ROLE|AGENT] \prec [ROLE|THEME]
 (2.32) [ROLE|AGENT] \prec [ROLE|GOAL]
 (2.33) [ROLE|GOAL] \prec [ROLE|THEME]
 (2.34) [FOCUS -] \prec [FOCUS +]

We can reconstruct (2.30)–(2.34) as the “complex” LP constraint AGENT \prec GOAL \prec THEME. Although the constraints are ordered by “weight”, Uszkoreit does not try to assign a mathematically “nondiscrete” or “continuous” probabilistic or statistical weight. Rather, he just orders them according to their relative “strength”. This suggests that the constraints constitute a *partially ordered set*

or *poset*. A poset is a structure $\langle P, < \rangle$ where P is some set and $<$ is a binary relation on P such that the following three axioms hold.

$$\begin{aligned}
 &\text{Reflexivity:} && x \preceq x \\
 (2.35) \text{ Antisymmetry:} && x \prec y \wedge y \prec x \rightarrow x = y \\
 &\text{Transitivity:} && x \prec y \wedge y \prec z \rightarrow x \prec z
 \end{aligned}$$

Then the LP constraints presented by Uszkoreit are a partial order. In fact, they are a *total order* since each constraint is stronger than all of the constraints lower in the list. Let the constraints be labelled L_1 through L_5 . Then they are the totally ordered set $\langle P, < \rangle$ such that

$$\begin{aligned}
 P &= \{L_1, L_2, L_3, L_4, L_5\} \\
 < &= \{(L_1, L_2), (L_1, L_3), (L_1, L_4), (L_1, L_5), \\
 (2.36) \quad &(L_2, L_3), (L_2, L_4), (L_2, L_5), \\
 &(L_3, L_4), (L_3, L_5), \\
 &(L_4, L_5)\}
 \end{aligned}$$

By definition (although informally only) it follows that such ranked LP constraints constitute a poset. A constraint that is high in the order will be preferred much more highly than one which is very low in the order. The only difficulty that might arise would be a situation where $\alpha < \beta$ (possibly by transitivity) but that $\beta < \alpha$ is also true with the same ranking. There are two ways out of such a dilemma. First, we can simply stipulate that such a ranking of constraints is not valid. That is, to make it a definition of a welldefined grammar. The second way is to make the point that this will never arise for natural languages. The latter is I believe the correct response. Furthermore, we will show below that such cases can be catered for by slightly complicating our definitions.

The ability to state that one constraint is stronger than another without assigning numerical “weights” or values or probabilities has many advantages. First, the approach can continue to be expressed in terms of *discrete mathematics*, (i.e., the theory of ordering relations and not *continuous mathematics* like probability or statistics) which fits in better with the inherently discrete nature of the rest of grammar. Second, assigning numerical weights would be a very difficult task since large corpora-based statistical analysis would be needed to determine correct numerical weights or probabilities. Third, the implications of this approach are that there might be cases where there is essentially a choice of ordering possibilities arising from incomparable (in the theory of partial orders sense) constraints in the partial order. However, we will see below that this is actually a fruitful implication of the casting of weighted LP constraints explicitly as a partial order.

Statistical accounts have an advantage over the partial order account since there may be very slight differences in preferences between two poset elements while there may also be very great differences in preferences between two other elements. However, the difference is lost since the poset ordering is all that counts under this

approach. In practice, I don't think this arises as a problem given the additional definitions to follow. In any event, I don't have the statistical data needed across a large number of languages to see whether this "problem" ever arises.

2-10.3 Partially Ordered Sets of LP Constraints

From the partial order $\langle P, < \rangle$, a secondary partial order can be defined over sets of LP constraints. One basic possible definition is that a set A of constraints is "stronger" than a set B of constraints iff for every constraint in B there is one in A that is "stronger" than or equal to it. Here "stronger" means "is greater than". Let us call this order, \ll , the order *induced* by $<$. Formally, \ll is defined as follows.³

$$(2.37) B \ll A \text{ iff } \forall b \in B. \exists a \in A. (b < a).$$

The carrier set defined by \ll is the biggest subset of 2^P such that there are no *inconsistent* elements (LP constraints). An element of 2^P is *consistent* if it contains no two elements which are comparable with respect to the original ordering $<$ and are the inverses of each other. Then in general, the induced poset will be a subset of 2^P and \ll will be a subset of $2^P \times 2^P$. We also note that for all $A, B \in 2^P$, if $B \subset A$, then $B \ll A$ since every element of B is an element of A but not vice versa.

Informally, $B \ll A$ iff A is a stronger set of LP constraints than B . In other words, the ordering statements in A give rise to more acceptable word orderings than B does. To use an abused piece of terminology, if $B \ll A$ then the orderings which are licensed by B are more *marked* than those which are licensed by A . Sets of LP constraints which are "highest" in the partial order \ll are more acceptable and less marked than those which are less than it. Sets of LP constraints which are "lowest" in the partial order \ll are less acceptable and more marked than those which are greater than it.

Let \square_1 be the category (2.38) and \square_2 be the category (2.39).

$$(2.38) A = \square_1 \left[\begin{array}{cc} \text{ROLE} & \text{THEME} \\ \text{FOCUS} & - \\ \text{PRO} & + \end{array} \right]$$

$$(2.39) B = \square_2 \left[\begin{array}{cc} \text{ROLE} & \text{AGENT} \\ \text{FOCUS} & - \\ \text{PRO} & - \end{array} \right]$$

³For the technically-minded, the relation of \ll is a *Hoare simulation* from A to B .

Then by definition of \ll , $\boxed{1} \prec \boxed{2}$. Let us consider each attribute-value pair of $\boxed{2}$ in turn. $\boxed{1}$ has the thematic role **THEME**. However, $\boxed{2}$ has the thematic role **AGENT** and so the LP constraint $[\text{ROLE}|\text{AGENT}] \prec [\text{ROLE}|\text{THEME}]$ in isolation would require that an NP which satisfies *A* precede any NP which satisfies $\boxed{2}$. However, the LP constraint $[\text{PRO} +] \prec [\text{PRO} -]$ is stronger than the other axiom so the thematic role axiom need not apply. Furthermore, both values of **FOCUS** are $-$ so there is no ordering preference induced by topic-focus structure. Finally, $\boxed{1}$ is $[\text{PRO} +]$ while $\boxed{2}$ is $[\text{PRO} -]$, therefore $\boxed{1} \prec \boxed{2}$.

In general then, given two constituents, e.g., two NPs, we take the greatest consistent set of LP constraints which are *satisfied* by each NP and then use the partial order \ll to order the two NPs. In general, the greatest consistent sets of constraints for the two NPs can be incomparable. In this case, we predict that either order is acceptable.

Consider the sentence *Johan Hans Marie zu helfen₁ versprochen₂ hat₃*. Then *Johan*, *Hans* and *Marie* are all proper names, are all definite and are all $[\text{FOCUS} -]$. Then we know that in the absence of any overt morphological, prosodic, topic-focus, grammatical functional or syntactic information which distinguishes the three NPs, then the case LP constraint (2.27) implies that *Johan* is in the nominative case and that *Hans* is in either the dative or the accusative case. Subcategorisation requires that both *Hans* and *Marie* be in the dative case. Furthermore, (2.40)–(2.42) require that agents precede themes precede patients. Therefore, *Hans* must be the theme and *Marie* the patient. But then (2.40) and (2.42) require that *Hans* be the theme (i.e., the dative object of *versprochen*) and *Marie* be the patient (i.e., the dative object of *zu helfen*). Therefore, we arrive at the coindexation (2.43).

(2.40) $[\text{ROLE}|\text{AGENT}] \prec [\text{ROLE}|\text{THEME}]$

(2.41) $[\text{ROLE}|\text{AGENT}] \prec [\text{ROLE}|\text{PATIENT}]$

(2.42) $[\text{ROLE}|\text{THEME}] \prec [\text{ROLE}|\text{PATIENT}]$

(2.43) *Johan₁ Hans₂ Marie₃ zu helfen₁ versprochen₂ hat₃*

This means that no matter what the government relation of the three verbs, the “default” interpretation is for the three NPs to occur in the order that they do if they are to be assigned the coindexation in (2.43). In the simple unproblematic case where *Johan*, *Hans* and *Marie* are all complements of the same verb, then the LP constraints (2.44)–(2.46) will guarantee the correct interpretation *because every other factor is constant*.

(2.44) $[\text{GR-FUN}|\text{SUBJ}] \prec [\text{GR-FUN}|\text{IOBJ}]$

(2.45) $[\text{GR-FUN}|\text{SUBJ}] \prec [\text{GR-FUN}|\text{OBJ}]$

(2.46) $[\text{GR-FUN}|\text{IOBJ}] \prec [\text{GR-FUN}|\text{OBJ}]$

The definition of \ll is too general however. In fact, it is not a partial order. Given a poset P , it is possible to have two elements of 2^P , A and B , such that $A = \{\lambda_1\}$ and $B = \{\lambda_1, \lambda_2\}$ where λ_1 and λ_2 are LP constraints with exactly the same "weight". To make this concrete, let $P = \{\text{NP}[\text{NOM}] \prec \text{NP}[\text{DAT}], \text{NP}[\text{NOM}] \prec \text{NP}[\text{ACC}]\}$. Let $\lambda_1 = (\text{NP}[\text{NOM}] \prec \text{NP}[\text{DAT}])$ and $\lambda_2 = (\text{NP}[\text{NOM}] \prec \text{NP}[\text{ACC}])$ and let $\lambda_2 < \lambda_1$.

Then $B \ll A$ and $A \ll B$. But then by the antisymmetry axiom, $A = B$ and we have a contradiction. Therefore, let us revise the definition of \ll in the terms of a modified powerset construction.

Let P be the set of ordering constraints. (For the moment we will assume that P is a partial order.) Let P^\pm be the set

$$(2.47) \{ x|y \subseteq P \text{ and } z = P \setminus y \text{ and } x = y \cup \text{inverse}(z) \}$$

where $\text{inverse}(x)$ is a function defined on sets as follows.

$$(2.48) \quad \text{inverse}(\{ \}) = \{ \}$$

$$(2.49) \text{inverse}(\{ \phi \prec \psi \} \cup x) = \{ \psi \prec \phi \} \cup \text{inverse}(x)$$

We will call P^\pm the *complemented powerset of P* . Let's consider a simple example. Let

$$P = \left\{ \begin{array}{l} (\text{NOM} \prec \text{DAT}) \\ (\text{NOM} \prec \text{ACC}) \end{array} \right\}$$

where $(\text{NOM} \prec \text{DAT}) < (\text{NOM} \prec \text{ACC})$.

Then

$$2^P = \left\{ \begin{array}{l} \{(\text{NOM} \prec \text{DAT}), (\text{NOM} \prec \text{ACC})\} \\ \{(\text{NOM} \prec \text{DAT})\} \\ \{(\text{NOM} \prec \text{ACC})\} \\ \{ \} \end{array} \right\}$$

and

$$P^\pm = \left\{ \begin{array}{l} \{(\text{NOM} \prec \text{DAT}), (\text{NOM} \prec \text{ACC})\} \\ \{(\text{NOM} \prec \text{DAT}), (\text{ACC} \prec \text{NOM})\} \\ \{(\text{DAT} \prec \text{NOM}), (\text{NOM} \prec \text{ACC})\} \\ \{(\text{DAT} \prec \text{NOM}), (\text{ACC} \prec \text{NOM})\} \end{array} \right\}$$

We can see that P^\pm is a *completion* of P in the sense that either λ or its inverse is an element of every set in P^\pm . The completion of A is

$\{(\text{NOM} \prec \text{DAT}), (\text{ACC} \prec \text{NOM})\}$ and that B is the set $\{(\text{NOM} \prec \text{DAT}), (\text{NOM} \prec \text{ACC})\}$. Informally, at this point we can see that $A \ll B$ but $B \not\ll A$ because B contains only elements of P but A contains the inverse of $(\text{NOM} \prec \text{ACC})$. Let λ^{-1} be the inverse of λ . Then in order to maintain the definition of \ll , we need to augment P with the inverses of each element of P such that if $\lambda_1 < \lambda_2$ in P $\lambda_2^{-1} < \lambda_1^{-1}$. Now $A \ll B$ and $B \ll A$.

Notice that the complemented powerset of P always has the set P as a greatest element (highest acceptability, i.e., "grammatical") and the set of inverses of the elements of P as its least element (highest unacceptability, i.e., ungrammatical). Informally, we augment an element A of P with the inverses of $P \setminus A$ since the inverses of the set complement $P \setminus A$ are implicit in their absence in A . In this case,

$$P' = \left\{ \begin{array}{l} (\text{NOM} \prec \text{DAT}) \\ (\text{NOM} \prec \text{ACC}) \\ (\text{DAT} \prec \text{NOM}) \\ (\text{ACC} \prec \text{NOM}) \end{array} \right\}$$

with the partial order indicated by the list. Then $A \ll B$.

2-10.4 Partially Ordered Sets of Partially Ordered Sets of LP constraints

We have discussed how two constituents can be ordered with respect to the order relation \ll over the completion of the powerset of P . We will now discuss how to order n constituents (for $n > 0$) with respect to P^\pm .

One possible solution is to compare each pair of n constituents with respect to P^\pm . However, for each element of P^\pm used to compare two constituents, we would like each such element to be as "great" as possible. This may be hard to do because the choice of one element for a constituent pair may influence the choice of another element for yet another pair. One way to avoid this is to consider the cross product of P^\pm times itself $(n-1)!$ times. The resulting set then induces a third poset $\langle (P^\pm)^n, \ll \rangle$. The relation \ll is defined as follows.

$$(2.50) \quad \langle s_1, \dots, s_{(n-1)!} \rangle \ll \langle t_1, \dots, t_{(n-1)!} \rangle \\ \text{iff} \\ (s_1 \ll t_1) \text{ and } \dots \text{ and } (s_{(n-1)!} \ll t_{(n-1)!})$$

Then the tuples of $(P^\pm)^n$ can be used to either determine the most likely interpretation or best possible ordering for a sequence of n phrases.

2-10.5 Markedness

With respect to word order theory, this approach makes sense of the ubiquitous but usually undefined (either formally or informally) concept of *markedness*. Elements near the top of the partial order are considered less marked than elements near the bottom. The greatest elements of the poset can be considered to be grammatical and acceptable. Similarly, the least elements of the poset can be considered to be ungrammatical and unacceptable. Elements in the domain which are not least or greatest elements are considered to be marked and to be less acceptable according to the LP constraints. Thus if presented with a sentence with marked order, P^\pm will indicate which ordering possibilities are less marked and more acceptable. Conversely, the ordering P can be analytically determined by examining enough data and permutations of that data and ranking them for acceptability. It is to be expected that different speakers will attach different preferences to a given constraint relative to the others and so we can expect there to be quite a bit of dialectical and idiolectal variation. It is also possible that a set of constraints A is greater than a set of constraints B but that corresponding acceptability judgements do not follow the partial order \ll . This is most likely to be the case when the number of LP constraints in A and B is high, when the constraints are complicated, when the number of "weak" constraints is high or when the strength of some constraints are very close. Furthermore, constraints which refer to pragmatics or topic-focus structure will be hard to place in the partial order since pragmatics refers to the current state of the world and topic-focus structure refers to the state of the speaker's mind and the dialogue that (s)he is participating in.

So, first, are LP constraints grammatical or extragrammatical and second, can we equate grammaticality with acceptability? I think the answer to the first question is that LP constraints can be either grammatical or extragrammatical and even both. I would claim that a purely grammatical LP constraint is the one which places the finite verb in second position in V2 clauses. Purely extragrammatical constraints are those which refer to pragmatic information and topic-focus structure. An example of a constraint that is both grammatical and extragrammatical is the LP constraint which requires that object pronoun clitic *es* in German must precede any other NPs in the word order domain that it occurs in. It is grammatical in the sense that it has become grammaticalised and is entirely robust. However, it also occurs in its position because it is definite, a pronoun and prosodically "very light". This last extragrammatical constraint is a general one, which also holds of Dutch and German, which places "heavy" elements to the right and "light" elements to the left. *es* is one of the "lightest" words in German.

This leads to the answer to the second question. The approach that we will take in this thesis is that if a schematic word order pattern can be instantiated to some set of words of the appropriate categories, then that schema is grammatical in the language even if there is a large number of instantiations which are unacceptable. Further, if there is no instantiation of the schematic pattern to be found in the

language then that pattern is ungrammatical. The implication of this position is that grammaticality and extragrammaticality cannot be equated (at least for the theory of word order). There may be many instantiations of a grammatical pattern that are not acceptable but they are still grammatical. This also leads to the conclusion that grammaticality can only be binary but that acceptability is graded. We have seen how the approach of modelling word order by partial orders of partially ordered LP constraints explains the grading of acceptability.

This means that there is no such thing as a “marked” grammatical *rule*. “Rules” (in the HPSG sense) are either in the grammar or they are not. Similarly, we cannot say that there is a single unmarked word order for a particular schematic rule in the grammar unless it happens to be the case that the partial order of constraints and its derived (induced) partial orders are total. To summarize, *canonical* order or *unmarked* order makes no sense in the context of the approach to LP constraints put forward here.

2-10.6 Statistics and a Discovery Procedure for Linear Precedence Constraints

The question which is implied by the previous sections is how to determine what the constraints are and what the partial order on them is. To date, the constraints and their relative strength has largely been postulated by linguists on the basis of their intuitions. In this section I want to propose that the statistical analysis of large corpora of annotated domains can serve as a *discovery procedure* which will help linguists determine the constraints and their partial order. By “discovery procedure” I do not mean a purely mechanical procedure in the spirit of American structuralism but rather an iterative analysis of large corpora which involves iterations of manual analysis by the linguist followed by mechanical statistical analysis on a computer to find correlations between LP constraints and to determine what the constraints are and what their relative strength is.

I’ll now give an example of how the procedure works. We’ll first consider the subordinate clause *es ihm jemand zu lesen versprochen hat*. We start with an annotated description of the domain of the clause. The clause’s domain consists of three pronominal NPs followed by a series of verbs. We make the initial assumption that the phonology (PHON), case (CASE), thematic role (ROLE), focus (FOCUS), grammatical function (GRAM-FN), pronominal (PRO) and definiteness

(DEF) features are referenced in the LP constraints.

(2.51)	a	③	<table style="border-collapse: collapse; width: 100%;"> <tr><td style="padding: 2px 5px;">PHON</td><td style="padding: 2px 5px;">⟨es⟩</td></tr> <tr><td style="padding: 2px 5px;">CASE</td><td style="padding: 2px 5px;">ACC</td></tr> <tr><td style="padding: 2px 5px;">ROLE</td><td style="padding: 2px 5px;">THEME</td></tr> <tr><td style="padding: 2px 5px;">FOCUS</td><td style="padding: 2px 5px;">-</td></tr> <tr><td style="padding: 2px 5px;">GRAM-FN</td><td style="padding: 2px 5px;">OBJ</td></tr> <tr><td style="padding: 2px 5px;">PRO</td><td style="padding: 2px 5px;">+</td></tr> <tr><td style="padding: 2px 5px;">DEF</td><td style="padding: 2px 5px;">+</td></tr> </table>	PHON	⟨es⟩	CASE	ACC	ROLE	THEME	FOCUS	-	GRAM-FN	OBJ	PRO	+	DEF	+
PHON	⟨es⟩																
CASE	ACC																
ROLE	THEME																
FOCUS	-																
GRAM-FN	OBJ																
PRO	+																
DEF	+																
	b	②	<table style="border-collapse: collapse; width: 100%;"> <tr><td style="padding: 2px 5px;">PHON</td><td style="padding: 2px 5px;">⟨ihm⟩</td></tr> <tr><td style="padding: 2px 5px;">CASE</td><td style="padding: 2px 5px;">DAT</td></tr> <tr><td style="padding: 2px 5px;">ROLE</td><td style="padding: 2px 5px;">PATIENT</td></tr> <tr><td style="padding: 2px 5px;">FOCUS</td><td style="padding: 2px 5px;">-</td></tr> <tr><td style="padding: 2px 5px;">GRAM-FN</td><td style="padding: 2px 5px;">IOBJ</td></tr> <tr><td style="padding: 2px 5px;">PRO</td><td style="padding: 2px 5px;">+</td></tr> <tr><td style="padding: 2px 5px;">DEF</td><td style="padding: 2px 5px;">+</td></tr> </table>	PHON	⟨ihm⟩	CASE	DAT	ROLE	PATIENT	FOCUS	-	GRAM-FN	IOBJ	PRO	+	DEF	+
PHON	⟨ihm⟩																
CASE	DAT																
ROLE	PATIENT																
FOCUS	-																
GRAM-FN	IOBJ																
PRO	+																
DEF	+																
	c	①	<table style="border-collapse: collapse; width: 100%;"> <tr><td style="padding: 2px 5px;">PHON</td><td style="padding: 2px 5px;">⟨jemand⟩</td></tr> <tr><td style="padding: 2px 5px;">CASE</td><td style="padding: 2px 5px;">NOM</td></tr> <tr><td style="padding: 2px 5px;">ROLE</td><td style="padding: 2px 5px;">AGENT</td></tr> <tr><td style="padding: 2px 5px;">FOCUS</td><td style="padding: 2px 5px;">+</td></tr> <tr><td style="padding: 2px 5px;">GRAM-FN</td><td style="padding: 2px 5px;">SUBJ</td></tr> <tr><td style="padding: 2px 5px;">PRO</td><td style="padding: 2px 5px;">+</td></tr> <tr><td style="padding: 2px 5px;">DEF</td><td style="padding: 2px 5px;">-</td></tr> </table>	PHON	⟨jemand⟩	CASE	NOM	ROLE	AGENT	FOCUS	+	GRAM-FN	SUBJ	PRO	+	DEF	-
PHON	⟨jemand⟩																
CASE	NOM																
ROLE	AGENT																
FOCUS	+																
GRAM-FN	SUBJ																
PRO	+																
DEF	-																

(2.52) ⟨ ③NP[ACC], ②NP[DAT], ①NP[NOM], V₃, V₂, V₁ ⟩

On the basis of this data, we might hypothesise the following LP constraints:

(2.53) [PHON ⟨es⟩] < [PHON ⟨ihm⟩]

(2.54) [PHON ⟨es⟩] < [PHON ⟨jemand⟩]

(2.55) [PHON ⟨ihm⟩] < [PHON ⟨jemand⟩]

(2.56) [CASE ACC] < [CASE DAT]

(2.57) [CASE ACC] < [CASE NOM]

(2.58) [CASE DAT] < [CASE NOM]

(2.59) [ROLE THEME] < [ROLE PATIENT]

(2.60) [ROLE THEME] < [ROLE AGENT]

(2.61) [ROLE PATIENT] < [ROLE AGENT]

- (2.62) [FOCUS -] \prec [FOCUS +]
 (2.63) [GRAM-FN OBJ] \prec [GRAM-FN IOBJ]
 (2.64) [GRAM-FN OBJ] \prec [GRAM-FN SUBJ]
 (2.65) [GRAM-FN IOBJ] \prec [GRAM-FN SUBJ]
 (2.66) [PRO +] \prec [PRO -]
 (2.67) [DEF +] \prec [DEF -]

Some of these are correct and some are not. (2.53)-(2.55) are probably correct but weaker than some of the other LP constraints such as (2.67). (2.56)-(2.58) are correct but again weaker than some of the others. (2.59)-(2.59) are probably incorrect but we can make use of them as very weak constraints (or as part of the completion of the robust constraints). (2.62) is probably robustly correct. (2.63)-(2.65) are probably incorrect but can serve as very weak constraints in the completion of the partial order. (2.66) and (2.67) are correct and robust.

With this set of hypothesised constraints in mind, we can then examine other data and do a statistical analysis to determine which constraints hold the most often. With this statistical data we can determine whether a given LP feature or factor is stronger than or weaker than some other constraint. If a constraint and its inverse occur roughly the same number of times, then we can eliminate it from the set of LP constraints since the pattern and its inverse are essentially in free variation. For example, this is likely to be the case with derived constraints concerning phonology. The only times when there are likely phonological constraints which hold are those of the type found in (2.53)-(2.55). However, in these cases it seems likely that these constraints are derivable from the fact that *es* is "lighter" than either *ihm* or *jemand* and that *es* and *ihm* are definite while *jemand* is indefinite.

We then check these assumptions against subsets of the data in large corpora where we try to control for the different parameters of variation. We then examine that data, updating the LP constraints accordingly and reanalyse the data again. We perform this procedure repeatedly until the statistics stabilise and use the final set of LP constraints as the basis for the grammar. Eventually you will converge to a stable state where the LP constraints are infallible modulo possible free variation or you will converge to a state where the LP factors are deemed to be inadequate. You then have to hypothesise other constraints on the basis of a manual examination of the evidence or take extragrammatical factors into consideration.

I believe that using a computer on large machine-readable corpora is the only way that we will ever fully understand the factors involved in German word order. In my opinion, this is probably one of the few places in the study of syntax and word order that can be facilitated by statistical analysis on a computer with large corpora of data. Furthermore, we can check for correlations between the different factors using statistical analysis to see if certain constraints are actually derivable from other constraints. Constraints like (2.53)-(2.55) are of this type.

2-11 Dowty (1991)

2-11.1 Introduction

In his paper *Towards a Minimalist Theory of Syntactic Structure* ([12]), David Dowty presents a theory of syntax and word order that is very similar in many respects to the one presented in this thesis. Here I will briefly review his proposals and point out where they agree and where they deviate substantially from my approach.

Dowty draws the same distinction between *which phrases* combine to form larger phrases and the *way* they combine or the *form* that those combinations take. This is the *tectogrammatic-phenogrammatic* distinction introduced elsewhere.⁴ His methodological assumptions are very similar to my own but he assumes slightly different grammatical machinery. I will outline his approach below and an example of it applied to Finnish. I will then consider whether the approach as stated in the paper is of the adequate for a simple German and a simple Dutch example.

2-11.2 The Theoretical Framework

The components of Dowty's theory are as follows:⁵

1. "A **Categorial Grammar** with compositional semantics; syntactic operations combine words and phrases to form larger phrases and sentences."
2. "The **default operation** for combining two expressions syntactically is to merge their words into a single unordered list."
3. **Linear Precedence Principles** of the kind familiar from **GPSG** and **HPSG** order constituents. Rules of the form "must appear in second position" are allowed. Furthermore, all **LP** principles are default rules which can be overridden by rule-specific syntactic operations.
4. "For each language, there is a list of **Bounding Categories**: parts of expressions of these categories cannot mingle with expressions outside the bounding category expression and vice-versa; these are "constituents" in the traditional sense. The list of bounding categories is regarded as a language-specific parameter. . . ."
5. "Constituent-formation *can* be specified as a rule-specific syntactic operation, but a marked one: there are two other kinds of possible syntactic operations besides the default one: (i) **ordering** one expression to the left or right of the head of another, and (ii) **attaching** one expression to the

⁴In fact, my first exposure to the terminology was in Dowty's paper.

⁵The following exposition is based very closely on [12]. In places, I simply quote his definitions. This is indicated by double quotes "".

left or right of the head of another. The difference is that two expressions connected by attachment cannot be subsequently separated by other expressions, while ordering allows this.”

6. Some treatment of agreement and government morphology that satisfies “Keenan’s principle” that agreement morphology is copied from the argument to the functor, while the functor determines what government morphology appears on the argument.

Before we proceed, item (5) needs some further comment. In a footnote, Dowty adds the following.

“To order one expression before (after) the head of another is only meaningful in this framework if the first is bounded. Thus, options the theory could take at this point are (i) that such ordering is only to be well-defined for bounded expressions, and (ii) that if the first expression is not bounded, this operation has the effect of ordering the head of the first before (after) the head of the second. The latter will be adopted below.”

We will see that this second option is probably necessary to cover both German and Dutch.

Dowty uses set notation to indicate unordered sets of constituents. If a constituent is bounded, then its elements are surrounded by set brackets. Therefore, the notation $\{a, b, \{c, d\}\}$ is an unordered set that contains elements a , b and $\{c, d\}$. $\{c, d\}$ is a *bounded constituent* that contains the unordered elements c and d . Assume the LP principles $a \prec b$ and $c \prec d$. Then $\{a, b, \{c, d\}\}$ can be linearised only as one of the following three sequences.

$$(2.68) \begin{array}{cccc} & a & b & c & d \\ a & & c & d & b \\ c & & d & a & b \end{array}$$

“For grammatical rules, I will adopt a Montague-style notation (since phrase-structure rules will obviously not be suitable), and the default combination operation will be represented by set union:

(10) (Default Syntactic Operation)

S1. If $\alpha \in A/B$, $\beta \in B$, then $F_1(\alpha, \beta) \in A$, where $F_1(\alpha, \beta) = \alpha \cup \beta$.

Lexical items themselves will be singleton sets; therefore all expressions, lexical and complex will be sets, so set union is always well-defined on them.

The two "marked" operations will be symbolized as in (11).

- (11) a $F_2(\alpha, \beta) = \alpha \ll \beta$ ("α ordered to the left of β")
 b $F_3(\alpha, \beta) = \alpha + \beta$ ("β attached to the head of α")"

2-11.3 An Example from Finnish

Dowty provides an example analysis of the Finnish sentence (19).

- (19) En minä ole aikonut ruveta pelaamaan näissä tennistä
 not I have intend start play these in tennis
 'I did not intend to start to play tennis in these'

The relevant factors to know about Finnish to appreciate this example are the following (quoted from [23, p48]).

- (12) a In declarative sentences, subjects and objects may occur in any order with respect to the main verb and one another.
 c The negative auxiliary (*e-*) precedes the temporal one (*ole-*) and both precede the main verb, but the three types of verbs need not be adjacent.
 d Elements of participial and infinitival clauses can be interspersed among the constituents of a superordinate clause.

Dowty assumes the following rules.

- (13) a If $\alpha \in S/NP$, $\beta \in NP[NOM]$, then $F_1(\alpha, \beta) \in S$ where $F_1(\alpha, \beta) = \alpha \cup \beta$.
 b If $\alpha \in TV$, $\beta \in NP[ACC]$, then $F_1(\alpha, \beta) \in VP$ where $F_1(\alpha, \beta) = \alpha \cup \beta$.

To implement the ordering restrictions over auxiliaries Dowty proposes the LP principle $v[+AUX, +NEG] \prec v[+AUX, +TEM] \prec v[-AUX]$.

Then just one more rule is needed to analyse (19).

- (21) If $\alpha \in VP/VP$, $\beta \in VP[-FIN]$, then $F_2(\alpha, \beta) \in VP$ where $F_2(\alpha, \beta) = \alpha \cup \beta$.

Then the analysis in Fig. 2-16 follows immediately by assuming that VP is not a bounded category.

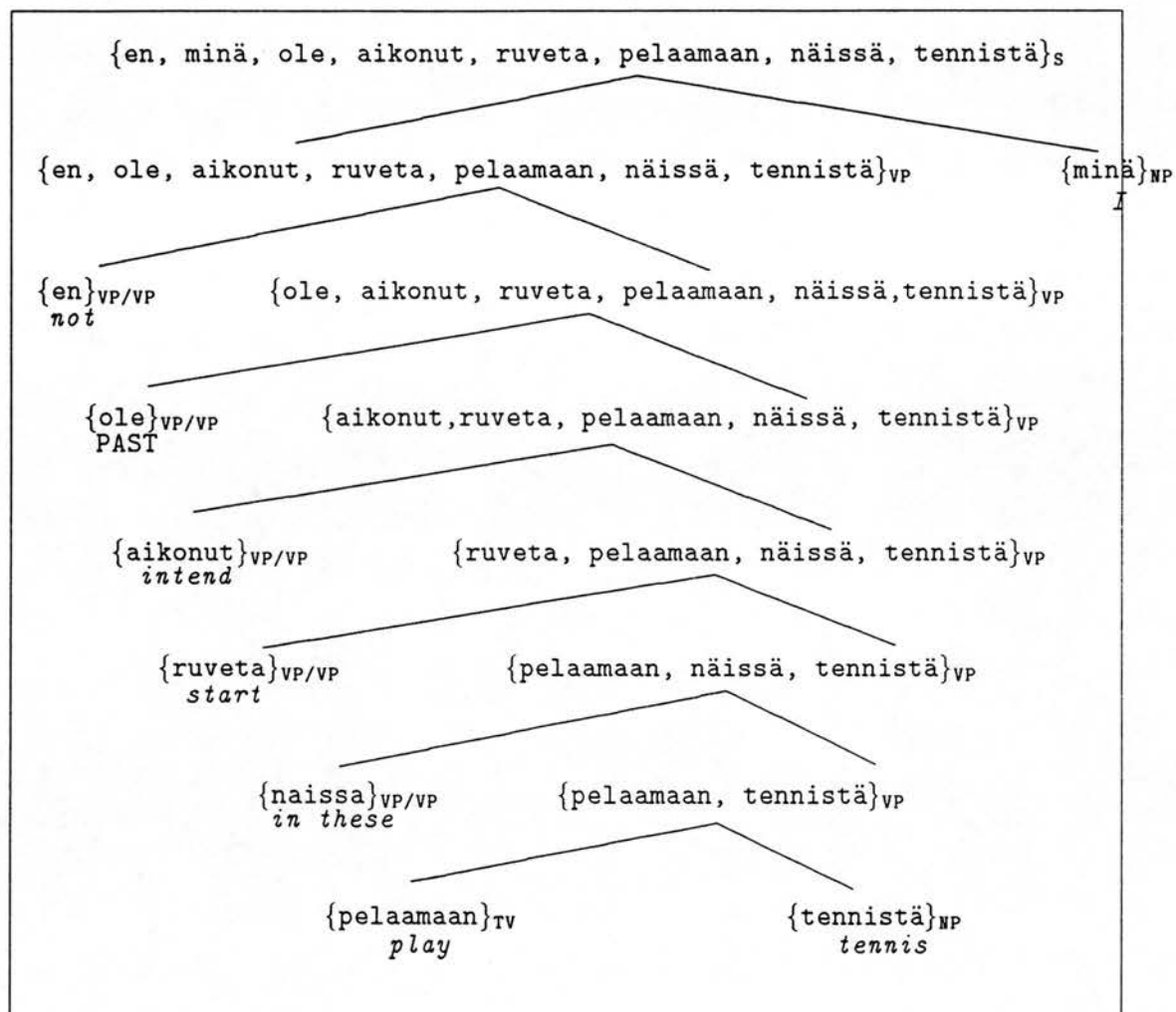


Figure 2-16: Dowty style analysis tree for *en minä ole aikunut ruveta pelaamaan näissä tennistä*

In addition, the examples in (20) can be generated (in addition to 38 more).

- (20)
- | | | | | | |
|----------------|-----|----------|---------|--------|-------------------|
| En minä näissä | ole | tennistä | aikonut | ruveta | pelaamaan |
| not I | | these in | have | tennis | intend start play |
-
- | | | | | |
|-------------------------|--------|----------|--------|-------------------|
| En minä tennistä näissä | ole | aikonut | ruveta | pelaamaan |
| not I | tennis | these in | have | intend start play |
-
- | | | | | | |
|------------------|--------|---------|--------|----------|------------|
| En minä tennistä | ole | aikonut | näissä | ruveta | pelaamaan |
| not I | tennis | have | intend | these in | start play |

2-11.4 Examples from German and Dutch

We'll now consider an example from German and an example from Dutch along the lines of Dowty's proposals. Both analyses depend crucially on allowing one head to be attached to another while treating both categories as unbounded. In the derivation tree in Fig. 2-17, the head of the argument of the functor verb is attached to the left of the functor. Thus at the top of the tree, the sequence *zu lesen versprochen hat* is attached. The analysis predicts that the NPs are all unordered even with respect to the verb sequence. This can easily be taken care of by an LP principle of the form $NP \prec V$. As NP ordering in the *Mittelfeld* is notoriously difficult, we assume a standard appeal to an account of LP principle preferences.

The same style of analysis applies to the derivation tree in Fig. 2-18. This time however attachment is to the right instead of to the left. Ignoring the problem of the order of the two objects, we must still explain how the subject ends up in clause initial position. Barring some ad hoc LP principle, I can't see how this is explained.

In either case, some explanation must be given for the clause initial position of the finite verb in V2 and V1 clauses. One simple explanation is that there is two versions of each finite verb, one marked +INV and the other marked -INV and there are two additional LP principles of the form $V[+INV] \prec X$ and $X \prec V[-INV]$ as in Uszkoreit's treatment ([45]).

We see then that the use of attachment enables a very elegant account of the basic facts of German and Dutch subordinate clause order. We will discuss this further below.

2-11.5 Discussion

Obviously, Dowty's characterisation of the "merging" operation as set union over unordered sets is not sufficient. A set of words does not contain the category and other information which LP principles are sensitive to. Rather, something like domain structure which maintains categorial information (at a minimum)

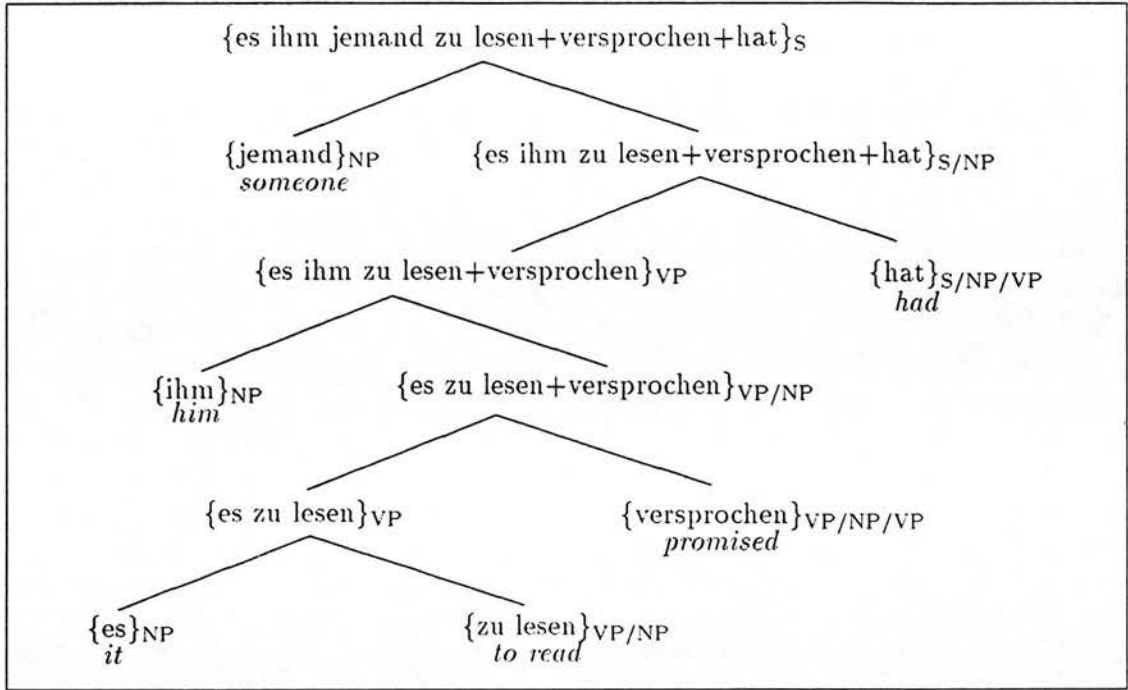


Figure 2-17: Dowty style analysis tree for *es ihm jemand zu lesen versprochen hat*

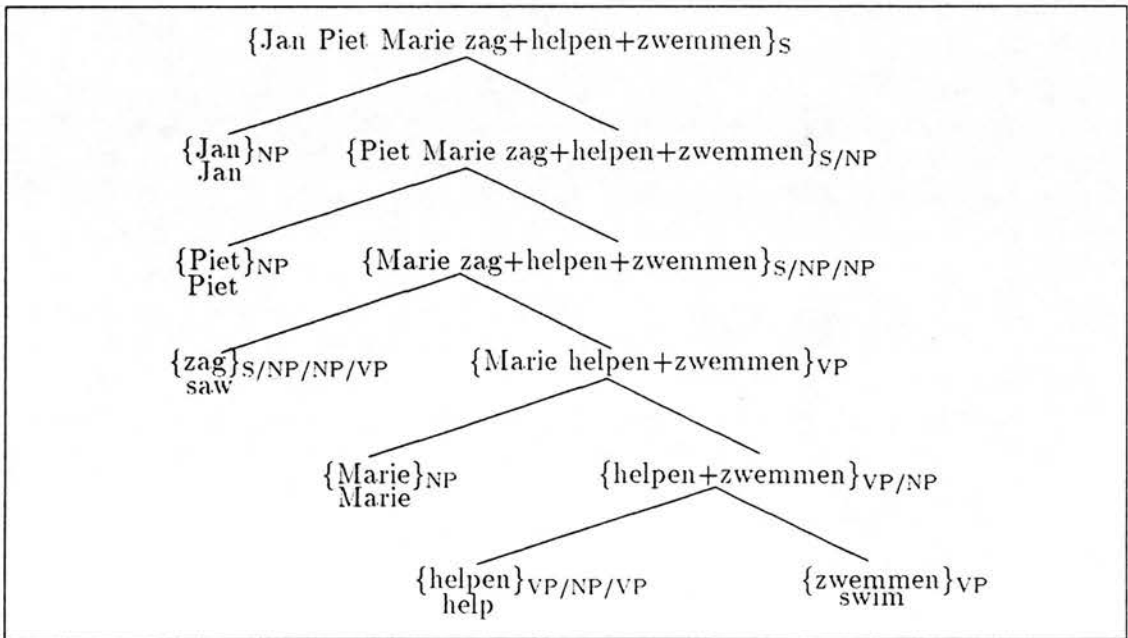


Figure 2-18: Dowty style analysis tree for *Jan Piet Marie zag helpen zwemmen*

is required. Otherwise, Dowty's merging operation is the same as our domain union operation except for two factors. First, Dowty assumes a binary branching categorial functor-argument structure so his operation is simpler to formalise. Second, Dowty makes no explicit statement whether two merged structures are ordered by the LP statements or not and further makes no explicit statement about whether the order of the constituents of the merged structures must be preserved in the merged structure. The analysis that I will give for German and Dutch crucially relies on this property holding.

Although the analysis in this thesis does not make use of it, I also assume that functors can lexically specify that their arguments must occur to the left or right of the functor. If I had built this into the framework explicitly, the analysis of Swiss German would be much more elegant. This functor specification of direction is of course a direct counterpart of standard categorial grammars and is well worth maintaining.

One possible difficulty for Dowty's approach is his assumption of bounding categories being specified for an entire language. If we consider German for example, it seems very clear that VPs are not bounded (to use Dowty's terminology) when they enter into verb cluster constructions but are bound when they occur in peripheral position, i.e., in either topicalised or extraposed position. For example, if Finnish contains a topic position that can be filled by VPs then the topicalised VP must be bound otherwise elements of the topic could "float" arbitrarily far to the right. I believe a better approach is to allow functors to specify whether their arguments are bounded and in some cases to allow language specific principles to determine whether a category is bounded or not, e.g., German VPs.

The major difference between Dowty's and my approach is his use of attachment operations. As shown in the German and Dutch examples above, this very nicely provides an analysis of canonical German and Dutch subordinate clauses. Furthermore, there is phonological evidence to suggest that verb clusters in German and Dutch are in fact "attached" since they form a single intonational phrase and cannot be interrupted by other material. One question that arises however is whether Dowty's assumptions are sufficient to explain alternative verb orderings such as 1-3-2. V_3 can be attached to the left of V_2 without problem. But then the question is what does it mean for V_1 to attach to the left of V_3 especially since V_3 and V_2 are attached. In a footnote, Dowty makes a lengthy comment which is relevant to this issue.

"I assume head is defined as more less customary in categorial grammar: in a complex expression formed from one expression of category A/B and a second category B , where $A \neq B$, the first is the head: if $A = B$, then the second is the head (i.e., the functor A/A is a modifier). Below, we will have occasion to refer to the lexical head of an expression, the simplest definition of which is that it is both a head and a lexical expression (i.e. is not a syntactically derived phrase). Some expressions would not have a lexical head by this definition, of course. If it turns out that we need to assume lexical heads for all expressions.

the intuitive way to proceed is by a recursive definition: one seeks the head of the functor, and if that functor is not lexical, then one seeks the head of that, and so on. However, the problem with this kind of definition is that it indirectly makes reference to the derivational history of an expression and therefore to its tectogrammatical "constituent structure", thereby defeating my goal of avoiding reference to constituency in syntactic operations. But the data treated in this paper do not suffice to determine what sort of notion of "head" is needed; the experience of Pollard (1984) and Pollard and Sag (1987) should be relevant, but I am not sure how to apply it here. Cases of attachment in this paper involve attachment to a (first-order) lexical head, or attachment to a group of attached words, of which the group is the (first-order) head and which also includes the (ultimate) lexical head."

Then if the "attached group of words" V_2+V_3 is considered the head of VP_2 , then V_1 can attach to it and the derivation goes through. In fact, in that case, any verb cluster order of the form that a verb is either to the left or the right of all the verbs it governs can be analysed by means of attachment, and elegantly at that. In this thesis, I have had no cause to use attachment but its use is certainly worthy of further investigation.

Chapter 3

An HPSG Formalisation of German Word Order and Syntax

In this chapter, I will present an HPSG formalisation of the fragment of German given in the previous chapter as well as discussions of semantic interpretation, semantic representations, the syntax and semantics of adjuncts, specifiers and quantifiers, a treatment of quantifier scope and the scope of modification, and a formalisation of raising.

I assume the treatment of unbounded dependencies presented in [32, Ch. 4].

3-1 Introduction

I'll now present an interpretation of this account of word order in a variant of HPSG as described in [31] (P&S1).¹ The account presented here is reminiscent of certain categorial grammar analyses. As in standard HPSG, the basic linguistic object is the *sign*. All signs are specified for the attributes **PHON** (**PHONOLOGY**), **SYN** (**SYNTAX**) and **SEM** (**SEMANTICS**) encoding the phonology, syntax and semantics respectively of a sign. **PHON** takes a sequence of atomic elements representing orthographic words as its value. **SYN** and **SEM** are to be interpreted precisely as in P&S1. Phrasal signs are also specified for the **DOM** (**DOMAIN**) attribute. It encodes the word order domains discussed in Ch. 1 and takes a sequence of signs as its value. Lexical signs are not defined for the **DOM** attribute. Whereas domains as labelled bracketed strings encode category information in the label, here domains are the value of the **DOM** attribute within a sign. That is,

$$(3.1) \quad [_{VP} X_1, \dots, X_n] \equiv \left[\begin{array}{l} \text{SYN} | \text{LOC VP} \\ \text{DOM } \langle X_1, \dots, X_n \rangle \end{array} \right]$$

¹In this work, we will have to assume familiarity with the development of HPSG in [31]. For formal details of the semantics of a notational variant of the formalism used here, cf. [34].

Phrasal signs are specified for the **DTRS** (**DAUGHTERS**) attribute. Its value can be of type *functor-argument-structure*. Functor-argument structures are defined for the attributes **FUN-DTR** (**FUNCTOR-DAUGHTER**), **ARG-DTRS** (**ARGUMENT-DAUGHTERS**) and **HEAD-DTR** (**HEAD-DAUGHTER**). **HEAD-DTR** is the same attribute as in P&S1. **FUN-DTR** is the syntactic functor in a phrasal sign. It is either the head daughter (**HEAD-DTR**) in a *head-complement-structure* (cf. (3.2)) or the adjunct daughter in a *head-adjunct-structure* (cf. (3.3)). Both *head-complement-structure* and *head-adjunct-structure* are subtypes of *functor-argument-structure*.

$$(3.2) \quad [\text{DTRS } \textit{head-complement-structure} \ [\]] \Rightarrow \left[\text{DTRS} \begin{array}{l} \text{HEAD-DTR } \boxed{1} \\ \text{FUN-DTR } \boxed{1} \end{array} \right]$$

$$(3.3) \quad [\text{DTRS } \textit{head-adjunct-structure} \ [\]] \Rightarrow \left[\text{DTRS} \begin{array}{l} \text{HEAD-DTR } \boxed{1} \\ \text{FUN-DTR|SYN|LOC|ARGS } \langle \boxed{1} \rangle \end{array} \right]$$

ARGS (**ARGUMENTS**) takes a sequence of signs as its value representing the arguments of the functor. (It replaces the **SUBCAT** attribute of P&S1.) **ARG-DTRS** is the subsequence of **ARGS** that have been syntactically saturated. (It replaces the **COMP-DTRS** attribute of P&S1.) The *Head Feature Principle* guarantees that the **HEAD** features of the mother and head daughter are coindexed properly. (This is exactly as in P&S1.)

$$(3.4) \quad [\text{DTRS } \textit{headed-structure} \ [\]] \Rightarrow \left[\begin{array}{l} \text{SYN|LOC|HEAD } \boxed{1} \\ \text{DTRS|HEAD-DTR|SYN|LOC|HEAD } \boxed{1} \end{array} \right]$$

The *Functor-Argument Principle* replaces the *Subcategorisation Principle* of P&S1 and amends it in accordance with the changes to the feature system described above.²

$$(3.5) \quad [\text{DTRS } \textit{functor-argument-structure} \ [\]] \Rightarrow \left[\begin{array}{l} \text{SYN|LOC|ARGS } \boxed{1} \\ \text{DTRS} \begin{array}{l} \text{FUN-DTR|SYN|LOC|ARGS } \boxed{1} \circ \boxed{2} \\ \text{ARG-DTRS } \boxed{2} \end{array} \end{array} \right]$$

In addition to the *Functor-Argument Principle*, there is also the *Domain Principle* (which doesn't occur in P&S1).³ Although it looks formidable, the intuition

²Unlike Volume 1 of P&S1 and like Volume 2, less oblique arguments appear to the left and more oblique arguments appear to the right in the argument list.

³For the technically inclined, the *Domain Principle* is really a schema.

underlying it is quite simple. It states the relation between the **DTRS** attribute and the **DOM** attribute according to the domain construction rules given in §1-2. Informally, it “maps” the syntactic structure onto the domain structure “non-deterministically”. It references the **UNIONED** attribute which specifies whether a daughter can be sequence unioned into the word order domain **DOM**. **UNIONED** is an attribute of signs. (The value of **UNIONED** can either be specified by lexical functors of their arguments or by language specific principles. In German, nonverb projections are [**UNIONED** -] while verb projections are unspecified for **UNIONED**. Therefore they can either appear continuously or discontinuously in a domain.) Basically, the Domain Principle states that the functor (**FUN-DTR**) is an element of the domain **DOM** and that every argument daughter (element of **ARG-DTRS**) is either an element of **DOM** or its domain is domain unioned into **DOM**.

$$(3.6) \quad \boxed{0} \langle \boxed{2}[\text{UNIONED } -], \dots, \boxed{n}[\text{UNIONED } -] \rangle \circ \langle [\text{UNIONED } +, \text{DOM } \boxed{i+1}], \dots, [\text{UNIONED } +, \text{DOM } \boxed{n}] \rangle$$

$$(3.7) \quad \left[\text{DTRS } \textit{functor-argument-structure} \left[\begin{array}{l} \boxed{1} \\ \boxed{2} \\ \dots \\ \boxed{n} \end{array} \right] \right] \Rightarrow \left[\begin{array}{l} \text{DTRS} \quad \left[\begin{array}{l} \text{FUN-DTR} \quad \boxed{1} \\ \text{ARG-DTRS} \quad \boxed{0} \end{array} \right] \\ \text{DOM} \quad \langle \boxed{1} \rangle \circ \langle \boxed{2} \rangle \circ \dots \circ \langle \boxed{n} \rangle \circ \langle \boxed{i+1} \rangle \circ \dots \circ \langle \boxed{n} \rangle \end{array} \right]$$

Given the definition of \circ . \circ is deterministic on a sequence C if every element of C is marked either $\mathbf{F} +$ or $\mathbf{F} -$ for some feature \mathbf{F} . That is, there will exist unique sequences A and B such that $\circ(A, B, C)$ and every element of B is marked $\mathbf{F} +$ and every element of A is marked $\mathbf{F} -$. Therefore, \circ is deterministic in the definition of the Domain Principle since every element is either **UNIONED** + or **UNIONED** -. Therefore, there will be only one way of dividing a sequence into the set of unioned and nonunioned elements.

The *HPSG Constituent Ordering Principle* is amended so that it requires that the value of the **PHON** attribute be the concatenation of the values of the **PHON** attributes of the elements of the **DOM** sequence if it is specified. Lexical signs specify their phonologies lexically.

$$(3.8) \quad \textit{phrasal-sign} \left[\begin{array}{l} \boxed{1} \\ \boxed{2} \\ \dots \\ \boxed{n} \end{array} \right] \Rightarrow \left[\begin{array}{l} \text{PHON} \quad \boxed{1} \circ \dots \circ \boxed{n} \\ \text{DOM} \quad \langle [\text{PHON } \boxed{1}], \dots, [\text{PHON } \boxed{n}] \rangle \end{array} \right]$$

This concludes the modifications to the universal principles of P&S1 (except for the *Semantics Principle* which is discussed in §3-3). We'll now look at the three phrase structure rules and the language specific principles required for the fragment of German considered in Ch. 3.

(3.9)-(3.11) replace Rules 1, 2 and 3 of P&S1. They are all amended for the **FUN-DTR**, **ARGS** and **ARG-DTRS** attributes. Rule 4 is eliminated. Instead, Rules

1 and 2 generalise the syntax of head-complement signs and head-adjunct signs. Rule 1 lets nonlexical heads combine with a single remaining argument. It disallows VPs from combining with subjects ($\text{NP}[\text{NOM}]$) since the analysis assumes that there is no VP in a clause. This enforces the prohibition of $[_s \text{ NP}[\text{NOM}] \text{ VP}]$ syntactic structures in German.

(3.9) Rule 1

$$\begin{array}{l} \text{a} \quad \left[\begin{array}{l} \text{SYN|LOC|ARGS} \langle \rangle \\ \text{DTRS} \quad \left[\begin{array}{l} \text{FUN-DTR|SYN|LOC|LEX} - \\ \text{ARG-DTRS} \langle \neg \text{NP}[\text{NOM}] \rangle \end{array} \right] \end{array} \right] \\ \text{b} \quad [\text{ARGS} \langle \rangle] \rightarrow \text{F}[\text{LEX} -], \text{A}(\neg \text{NP}[\text{NOM}]) \end{array}$$

((3.9b), (3.10b) and (3.11b) are the “schematic” form of the rules. **F** indicates the functor and **A** a complement in the schematic rules.)

Rule 2 is just like the corresponding rule for English and Dutch. It lets a lexical head combine with all but one of its complements. It specifies $\text{INV} -$ to guarantee that VPs are head-final.

(3.10)Rule 2

$$\begin{array}{l} \text{a} \quad \left[\begin{array}{l} \text{SYN|LOC|ARGS} \langle [] \rangle \\ \text{DTRS|FUN-DTR|SYN|LOC} \quad \left[\begin{array}{l} \text{HEAD|INV} - \\ \text{LEX} + \end{array} \right] \end{array} \right] \\ \text{b} \quad [\text{ARGS} \langle [] \rangle] \rightarrow \text{F}[\text{INV} -, \text{LEX} +], \text{A}^* \end{array}$$

Rule 3 drops the $[\text{INV} -]$ specification on the head and requires that the functor (the head verb) be a finite verb. This rule allows all of the complements of a verb to be combined with a finite verb at one time, thus allowing the subject to appear in the same domain as the nonsubject complements.

(3.11)Rule 3

$$\begin{array}{l} \text{a} \quad \left[\begin{array}{l} \text{SYN|LOC|ARGS} \langle \rangle \\ \text{DTRS|FUN-DTR|SYN|LOC} \text{ V}[\text{FIN.LEX} +] \end{array} \right] \\ \text{b} \quad [\text{ARGS} \langle \rangle] \rightarrow \text{V}[\text{FIN.LEX} +], \text{A}^* \end{array}$$

Rules 1, 2 and 3, the Functor-Argument Principle and the Domain Principle determine the elements of a **DOM** sequence but only partially determine the *order* of the elements in a **DOM** sequence.

Assume the LP statements in (3.12)-(3.15).

$$(3.12) \quad [\text{DOM}] \Rightarrow [\text{DOM NP} \prec \text{V}[\text{INV} -]]$$

$$(3.13)[\text{DTRS}|\text{HEAD-DTR} \sqcap \text{V}[\text{INV} -]] \Rightarrow [\text{DOM V} \preceq \sqcap]$$

$$(3.14)[\text{DTRS}|\text{HEAD-DTR} \sqcap \text{V}[\text{INV} +]] \Rightarrow [\text{DOM} \sqcap \preceq []]$$

$$(3.15) \quad [\text{DOM}] \Rightarrow [\text{DOM} [\text{EXTRA} -] \prec [\text{EXTRA} +]]$$

(3.12) requires that NPs precede verbs, (3.13) that noninverted (i.e., non-clause-initial) head daughter verbs are preceded by the verbs they govern, (3.14) that inverted (clause-initial) verbs precede everything and (3.15) that non-extrapolated constituents precede extrapolated constituents. (3.13) has the effect it does because it requires the head verb of a sign to be preceded by (or be equal to) all the verbs in the sign's domain. Since all the verbs in the sign's domain are governed by the head verb, they will remain in that order in any domain the domain is unioned into. So, in any domain, each verb will precede any verb that governs it. Inverted clauses occur in V1 and V2 clauses.

Let **VP** be notational shorthand for $\text{V}[\text{ARGS} \langle \text{NP}[\text{NOM}] \rangle]$ and **S** notational shorthand for $\text{V}[\text{ARGS} \langle \rangle]$. Then also assume the language specific principles (3.16) and (3.17).

$$(3.16) \quad [\text{EXTRA} +] \Rightarrow \text{VP}$$

$$(3.17)[\text{UNIONED} +] \Rightarrow \text{VP} \vee \text{S}$$

(3.16) encodes the fact that only VPs can be extrapolated. (3.17) encodes the fact that only VPs and Ss can be sequence unioned. The contrapositives (3.18) and (3.19) can also be derived since every sign is specified for **EXTRA** and **UNIONED**.

$$(3.18) \quad \neg \text{VP} \Rightarrow [\text{EXTRA} -]$$

$$(3.19)\neg \text{VP} \wedge \neg \text{S} \Rightarrow [\text{UNIONED} -]$$

With these formal preliminaries in hand, we can now consider an example. Let the features **NOM**, **DAT**, **ACC**, **FIN** and **PSP** be exactly as in P&S1. We also assume the **VFORM** features **ZU** for *zu*-infinitivals and **INF** for bare infinitivals. Then assume the following (schematic) German lexical entries.

$$(3.20) \quad \textit{jemand} : \text{NP}[\text{NOM}]$$

$$(3.21) \quad \textit{ihm} : \text{NP}[\text{DAT}]$$

$$(3.22) \quad \textit{es} : \text{NP}[\text{ACC}]$$

$$(3.23) \quad \textit{hat} : \text{V}[\text{FIN}, \text{ARGS} \langle \text{NP}[\text{NOM}], \text{VP}[\text{PSP}] \rangle]$$

$$(3.24)\textit{versprochen} : \text{V}[\text{PSP}, \text{ARGS} \langle \text{NP}[\text{NOM}], \text{NP}[\text{DAT}], \text{VP}[\text{ZU}] \rangle]$$

$$(3.25) \quad \textit{zu lesen} : \text{V}[\text{ZU}, \text{ARGS} \langle \text{NP}[\text{NOM}], \text{NP}[\text{ACC}] \rangle]$$

Given these assumptions, we can give an analysis of (1.15), *es ihm jemand zu lesen versprochen hat* 'someone has promised him to read it'.

In (3.26), the sign labelled $\boxed{8}$, the NP *es* ($\boxed{1}$) has combined by Rule 2 with the *zu*-infinitival verb *zu lesen* ($\boxed{4}$) forming the VP[ZU] *es zu lesen* whose DOM sequence is $\langle \boxed{1}, \boxed{4} \rangle$. $\boxed{1}$ precedes $\boxed{4}$ as required by (3.12).

$$(3.26) \quad \boxed{8} \left[\begin{array}{l} \text{PHON } \langle \text{es zu lesen} \rangle \\ \text{SYN|LOC VP[ZU]} \\ \text{DTRS} \left[\begin{array}{l} \text{HEAD-DTR } \boxed{4} \left[\begin{array}{l} \text{PHON } \langle \text{zu lesen} \rangle \\ \text{SYN|LOC V[ZU, ARGS(NP[NOM], \boxed{1})] \end{array} \right] \\ \text{FUN-DTR } \boxed{4} \\ \text{ARG-DTRS } \left\langle \boxed{1} \left[\begin{array}{l} \text{PHON } \langle \text{es} \rangle \\ \text{SYN|LOC|HEAD NP[ACC] \end{array} \right] \right\rangle \end{array} \right] \\ \text{DOM } \langle \boxed{1}, \boxed{4} \rangle \end{array} \right]$$

In (3.27), the sign labelled $\boxed{7}$, the NP *ihm* ($\boxed{2}$) and the VP *es zu lesen* ($\boxed{8}$) have combined by Rule 2 with the past participle *versprochen* ($\boxed{5}$) forming the VP[PSP] *es ihm zu lesen versprochen* whose DOM sequence is $\langle \boxed{1}, \boxed{2}, \boxed{4}, \boxed{5} \rangle$. This time, the domain of the VP *es zu lesen* has been unioned into the domain of $\boxed{7}$. Both NPs ($\boxed{1}$ and $\boxed{2}$) precede both verbs ($\boxed{4}$ and $\boxed{5}$) in the domain as required by (3.12). Furthermore, the governing verb *versprochen* ($\boxed{5}$) is preceded by the governed verb *zu lesen* ($\boxed{4}$) as required by (3.13). *es* and *ihm* are not restricted in order with respect to each other so they can come in either order.

$$(3.27) \quad \boxed{7} \left[\begin{array}{l} \text{PHON } \langle \text{es ihm zu lesen versprochen} \rangle \\ \text{SYN|LOC VP[PSP]} \\ \text{DTRS} \left[\begin{array}{l} \text{HEAD-DTR } \boxed{5} \left[\begin{array}{l} \text{PHON } \langle \text{versprochen} \rangle \\ \text{SYN|LOC V[PSP, ARGS(NP[NOM], \boxed{2}, \boxed{8})] \end{array} \right] \\ \text{FUN-DTR } \boxed{5} \\ \text{ARG-DTRS } \left\langle \boxed{2} \left[\begin{array}{l} \text{PHON } \langle \text{ihm} \rangle \\ \text{SYN|LOC|HEAD NP[DAT] \end{array} \right] \right\rangle \cdot \boxed{8} \end{array} \right] \\ \text{DOM } \langle \boxed{1}, \boxed{2}, \boxed{4}, \boxed{5} \rangle \end{array} \right]$$

Finally, in (3.28), the NP *jemand* ($\boxed{3}$) and the VP *es ihm zu lesen versprochen* ($\boxed{7}$) have combined by Rule 3 with the auxiliary *hat* ($\boxed{6}$) forming the S *es ihm jemand zu lesen versprochen hat* whose DOM sequence is $\langle \boxed{1}, \boxed{2}, \boxed{3}, \boxed{4}, \boxed{5}, \boxed{6} \rangle$. The domain of the VP *es ihm zu lesen versprochen* has been unioned into the domain of the S.

Again, all NPs precede all verbs and governed verbs precede governing verbs as required by (3.12) and (3.13).

$$(3.28) \left[\begin{array}{l} \text{PHON } \langle \text{es ihm jemand zu lesen versprochen hat} \rangle \\ \text{SYN|LOC } \text{S} \\ \text{DTRS } \left[\begin{array}{l} \text{HEAD-DTR } \boxed{6} \left[\begin{array}{l} \text{PHON } \langle \text{hat} \rangle \\ \text{SYN|LOC } \text{V}[\text{ARGS}(\boxed{3}, \boxed{7})] \end{array} \right] \\ \text{FUN-DTR } \boxed{6} \\ \text{ARG-DTRS } \left\langle \boxed{3} \left[\begin{array}{l} \text{PHON } \langle \text{jemand} \rangle \\ \text{SYN|LOC|HEAD } \text{NP}[\text{NOM}] \end{array} \right] \boxed{7} \right\rangle \end{array} \right] \\ \text{DOM } \langle \boxed{1}, \boxed{2}, \boxed{3}, \boxed{4}, \boxed{5}, \boxed{6} \rangle \end{array} \right]$$

There are a few things which have not been dealt with in either the informal treatment in Ch. 3 or in the formal treatment in this section. The first of these is the account of the complementiser *daß*. (3.29) is its lexical entry.

$$(3.29) \left[\begin{array}{l} \text{PHON } \langle \text{daß} \rangle \\ \text{SYN|LOC } \left[\begin{array}{l} \text{COMP } + \\ \text{ARGS } \left\langle \text{S} \left[\begin{array}{l} \text{SYN|LOC } \left[\begin{array}{l} \text{HEAD } \left[\begin{array}{l} \text{VFORM } \text{FIN} \\ \text{INV } - \end{array} \right] \\ \text{COMP } - \end{array} \right] \right] \right] \right] \right\rangle \end{array} \right] \end{array} \right]$$

daß is a functor which takes a noninverted, finite clause as its argument.⁴ The feature **COMP** (which is a **SYN|LOC** feature) keeps track of whether or not a clause has combined with a complementiser to form a subordinate clause (which we abbreviate as **s[COMP +]**). Since *daß* requires an **s[COMP -]** argument but is itself an **s[COMP +]**, it effectively allows only one complementizer to combine with a clause. (This is essentially the same as the **GPSG** analysis.) There is a small problem here in that the **+** value of **COMP** will not automatically be identified with the **COMP** value of the subordinate clause since it is not a head feature and the complementiser is not the head anyway. A purely technical solution is to add the implication (3.30) to the grammar.

$$(3.30) \quad \underline{[\text{DTRS|FUN-DTR|SYN|LOC|COMP } +]} \Rightarrow [\text{SYN|LOC|COMP } +]$$

⁴This is too strict as inverted subordinate clauses are sometimes found in German and in many dialects.

We will also see momentarily that we could state the antecedent in terms of the type *head-complementiser-structure*.

$$(3.31) \quad [\text{DTRS } \textit{head-complementiser-structure} \ [\]] \Rightarrow [\text{SYN|LOC|COMP } +]$$

Both (3.30) and (3.31) require that if the sign is a subordinate clause then the value of **SYN|LOC|COMP** is +. This is reminiscent of the Head Feature Principle. An analysis which takes the complementiser to be the head would not suffer from this problem. An even more categorial style of analysis would solve this problem as well. An additional problem is that the **COMP** value needs to be percolated from the head to the mother when the functor is not a complementiser. I'll leave the best way to handle this as an open question.

Another issue is the distribution of **INV** and **COMP**. Since both **INV** and **COMP** are binary valued, there are four possibilities. All four combinations are realised. They are summarised in (3.32).

$$(3.32) \quad \begin{array}{ll} [\text{COMP } +, \text{INV } +] & \text{dialect subordinate clauses} \\ [\text{COMP } +, \text{INV } -] & \text{subordinate clauses} \\ [\text{COMP } -, \text{INV } +] & \text{conditional clauses, V1,} \\ & \text{constituent of V2 clauses} \\ [\text{COMP } -, \text{INV } -] & \text{constituent of subordinate clauses} \end{array}$$

In addition to head-adjunct structures, we need to allow other headed structures which are not head-complement structures. For example, specifiers are not adjuncts but enter into headed structures where the Head Feature Principle should apply. We will discuss this further in §3-3. For the present purpose, we will assume that the value of the **DTRS** feature of a subordinate clause is a headed structure of type *head-complementiser-structure*. Since the complementiser specifies a head clause argument, it will combine with its argument by Rule 1. The Head Feature Principle means that the head features of the head clause will be identified with the head features of the **s[COMP +]**.

We also need to describe V2 clauses. As stated in §4-1, V2 clauses consist of a topic followed by an inverted clause with a gap in it corresponding to the topic. This means that we will need a filler-gap rule which requires that the gapped clause be **[INV +]**. For the link between the filler and the gap we assume some trace-based account of unbounded dependencies. A trace has a null phonology. That is, the value of the **PHON** feature is ϵ , the empty sequence, so it will contribute nothing to the phonology of any sign whose domain contains it since the Constituent Ordering Principle will just concatenate ϵ with the other **PHON** values.

The filler-gap rule will create a headed structure with a **FILLER-DTR** and a **HEAD-DTR**. This structure will be of type *head-filler-structure*. Schematically, the *head-filler-structure* type is of the form $[\text{CP } \square, \text{s}[\text{SLASH } \{\square\}, \text{INV } +]]$. In detail,

it is (3.33). Notice that the **FILLER-DTR** (the topic) is the only element in the **SLASH** set.

$$(3.33) \left[\begin{array}{l} \text{DTRS} \\ \left[\begin{array}{l} \text{FILLER-DTR } \boxed{1} \\ \text{HEAD-DTR} \\ \left[\text{SYN} \left[\begin{array}{l} \text{LOC|HEAD S[INV +]} \\ \text{BIND|SLASH } \{\boxed{1}\} \end{array} \right] \right] \end{array} \right] \end{array} \right]$$

The Head-Filler Rule (3.34) is the filler-gap rule.

(3.34) Head-Filler Rule

$$\left[\begin{array}{l} \text{SYN|BIND|SLASH } \{ \} \\ \text{DTRS} \\ \left[\begin{array}{l} \text{FILLER-DTR } \boxed{1} \\ \text{HEAD-DTR } \boxed{2} \\ \left[\text{SYN} \left[\begin{array}{l} \text{LOC|HEAD S[FIN,INV +]} \\ \text{BIND|SLASH } \{\boxed{1}\} \end{array} \right] \right] \end{array} \right] \\ \text{DOM } \langle \boxed{1}, \boxed{2} \rangle \end{array} \right]$$

Root wh-questions are just a subset of the V2 clauses. There is a [WH +] constituent in topic position which has been moved from the inverted, finite head clause. This is exactly the same in both Dutch and English. Inverted clauses with no topic, i.e., V1 clauses, also appear in German, Dutch and English as yes-no questions. In German, they are also used as conjunction-less conditional clauses and the verb must be in the subjunctive. The same is true to a lesser degree of English. E.g., *Were I to win a million pounds, I would quit work tomorrow.* Of course in English, a subjunctive, inverted verb must be an auxiliary or a modal.

A further issue is the distribution of the specifications [INV +] and [INV -] with respect to VPs and Ss. The implication $\text{VP} \Rightarrow [\text{INV -}]$ is sufficient. For verb projections, this is equivalent to the implication $[\text{INV +}] \Rightarrow \neg \text{VP}$ since **INV** is defined for all verb projections (although its value may be unspecified). Rule 3 (and its equivalent for Dutch and English) produce [INV +] for German, Dutch and English clauses. In German, the heads of subordinate clauses (without a subject NP-VP configuration) are [INV -]. In Dutch and English we will see that a clause can be [INV -] by virtue of the fact that there is a [s NP VP[INV -]] configuration available by the Dutch equivalent of Rule 2 which is unlicensed in German. An s[INV -] which is not of the form [s NP VP[INV -]] is impossible in Dutch and English because the Dutch equivalent of Rule 3 licenses VP-less clauses only if they are [INV +].

With respect to VPs, [s NP VP] is not licensed in German, so we need not consider the distribution of the values of **INV**. We already said above that Rule 2 licenses

[INV -] VPs in Dutch and English. But furthermore, only Rule 2 licenses VPs so VP[INV +] is out.

One point which needs some elaboration is the scheme used to order verbs in a verb cluster. Earlier we saw that the LP constraint (3.13) (repeated here as (3.35)) was responsible for the characteristic 3 – 2 – 1 German verb government order.

$$(3.35) [\text{DTRS} | \text{HEAD-DTR} \sqsupset \text{V} [\text{INV} -]] \Rightarrow [\text{DOM } v \preceq \sqsupset]$$

Of course, there are many other word orders available in German. For example, in the so-called *Ersatzinfinitiv* construction, the government order for verbs is 1 – 3 – 2 where V_1 is a finite auxiliary or finite modal verb and both V_2 and V_3 are bare infinitives as in (3.36).

$$(3.36) \quad \begin{array}{l} \text{weil} \quad \text{er} \text{ hat}_1 \text{ kommen}_3 \text{ dürfen}_3 \\ \text{because he has come may} \\ \text{'because he was allowed to come'} \end{array}$$

Bech ([3]) cites many other possible orders including 1 – 2 – 3 – 5 – 4 and 1 – 2 – 5 – 4 – 3. I won't bother to give a characterisation of the possible orders here. For a very good overview, see [13]. However, all of these verb cluster orders have the property that, for a cluster of length n , V_{n-1} (where the subscript denotes depth of government) is (trivially) adjacent either to the left or right of V_n , V_{n-2} is adjacent to the left or right of the V_{n-1} sequence, V_{n-3} is adjacent to the left or right of the V_{n-2} sequence and so on for V_{n-m} for $m < n$.⁵ That is, the possible orders are a subset of (3.37).

$$(3.37) \quad \begin{array}{l} V_1 \\ V_1 V_2 \quad V_2 V_1 \\ V_1 V_2 V_3 \quad V_2 V_3 V_1 \quad V_3 V_2 V_1 \quad V_1 V_3 V_2 \\ \dots \end{array}$$

That is, all of these orders can be explained in terms of a verb V_i appearing to the left or right of every verb V_{i+j} for $j > 0$ it governs. This suggests that direction of verb government should be specified in terms of a feature **DIR** (**DIRECTION**) for direction of "status government" or "verbal case" which takes only the values **LEFT** and **RIGHT** and a general ordering principle that says that if the value of **DIR** is **LEFT** then all verbs governed by the head verb precede it and if the value of **DIR** is **RIGHT** then all verbs governed by the head verb follow it. **DIR** should be an attribute of **SYN|LOC**. Then (3.38) is the *Direction of Status Government Principle* (in two parts).

$$(3.38) \quad (a) \quad [\text{DTRS} | \text{HEAD-DTR} \sqsupset [\text{SYN} | \text{LOC} | \text{DIR} | \text{LEFT}]] \Rightarrow [\text{DOM } v \preceq \sqsupset]$$

$$(b) \quad [\text{DTRS} | \text{HEAD-DTR} \sqsupset [\text{SYN} | \text{LOC} | \text{DIR} | \text{RIGHT}]] \Rightarrow [\text{DOM } \sqsupset \preceq v]$$

⁵The V_{n-i} sequence is the sequence of verbs containing $V_{n-i}, V_{n-i+1}, \dots, V_n$.

The Ersatzinfinitiv (3.36) can then be explained as follows. The bare infinitive *dürfen* subcategorises for a bare infinitival VP and lexically specifies the value **LEFT** for the value of its **DIR** attribute. *kommen* is such a VP so the verb cluster *kommen dürfen* is wellformed since *kommen* is to the left of *dürfen* as required by the value of **DIR**. Next, the finite auxiliary *hat* subcategorises for a bare infinitive VP rather than a past participle as expected and lexically specifies the value **RIGHT** for the value of its **DIR** attribute. *kommen dürfen* is such a VP so the verb cluster *hat kommen dürfen* is wellformed since *kommen dürfen* is to the right of *hat* as required by the value of **DIR**. The more complicated patterns are just extensions of this scheme. Although every lexical verb is defined for the attribute **DIR**, its value need not be specified. This would allow other principles of the grammar to determine the direction of verbal government, for example. Another possibility is that a given verb might be capable of governing in either direction.

The types of government order that this treatment cannot account for are those like 4 – 1 – 3 – 2 which violate the left-right adjacency pattern. In that case, there are two ways out available. Either V_4 is considered to be raised directly to a higher domain where its governing verb has no effect on its position and it is positioned by a different type of ordering constraint for raised lexical heads or we can look for evidence that some or all of the verbs are unordered for direction of government. Then it might be the case that V_4 is unordered by V_3 and so is not ordered by any of V_1 , V_2 or V_3 by transitivity. As mentioned before, the class of verbs which subcategorise for bare infinitival VPs in Zurich German seem to be of this type.

3–2 Semantic Interpretation

One of the many things that this account hopes to achieve is to make the problem of semantic compositionality straightforward, both for discontinuous constituents and in general. Here we will address the issue of semantic compositionality in head-complement structures, head-adjunct (-modifier) structures and head-specifier structures.

If we wanted to follow the semantics given in P&S1, the *Semantics Principle* (3.39) would be sufficient.

$$(3.39) \quad [\text{DTRS } \textit{headed-structure} []] \Rightarrow$$

$$\left[\begin{array}{l} \text{SEM} \quad \left[\begin{array}{l} \text{CONT} \quad \textit{successively-combine-semantics}(\boxed{1}, \boxed{2}) \\ \text{INDICES} \quad \textit{collect-indices}(\boxed{3}) \end{array} \right] \\ \text{DTRS } \boxed{3} \quad \left[\begin{array}{l} \text{HEAD-DTR} | \text{SEM} | \text{CONT } \boxed{1} \\ \text{COMP-DTRS } \boxed{2} \end{array} \right] \end{array} \right]$$

(3.40) is modified for use with the **FUN** and **ARG-DTRS** features (instead of the

HEAD-DTR and COMP-DTRS features). It also deletes the INDICES feature since it will not be used in the development to follow.

$$(3.40) \quad [\text{DTRS } \textit{headed-structure} []] \Rightarrow \left[\begin{array}{l} \text{SEM|CONT successively-combine-semantic}(\boxed{1}, \boxed{2}) \\ \text{DTRS } \left[\begin{array}{l} \text{FUN-DTR|SEM|CONT } \boxed{1} \\ \text{ARG-DTRS } \boxed{2} \end{array} \right] \end{array} \right]$$

The function successively-combine-semantic is defined semi-formally in P&S1 as (3.41).

$$(3.41) \quad \begin{array}{l} \text{successively-combine-semantic}(A, L) = \\ \text{if } \text{length}(L) = 0 \\ \text{then } A \\ \text{else} \\ \quad \text{successively-combine-semantic} \\ \quad (\text{combine-semantic}(A.\text{SEM|CONT of first}(L)), \text{rest}(L)). \end{array}$$

It is defined formally within the HPSG formalism in (3.42).

$$(3.42) \quad \begin{array}{l} \text{successively-combine-semantic}(\boxed{1}, \langle \rangle) \Leftrightarrow \boxed{1} \\ \text{successively-combine-semantic}(\boxed{1}, \langle [\text{SEM|CONT } \boxed{2}] \circ \boxed{3} \rangle) \Leftrightarrow \\ \text{successively-combine-semantic}(\text{combine-semantic}(\boxed{1}, \boxed{2}), \boxed{3}) \end{array}$$

P&S1 define the function combine-semantic semi-formally as (3.43).

$$(3.43) \quad \begin{array}{l} \text{combine-semantic}(A.B) = \\ \text{if } \begin{array}{l} A \text{ has type } \textit{circumstance} \text{ and} \\ B \text{ has type } \textit{quantifier} \end{array} \\ \text{then } \left[\begin{array}{l} \text{QUANT } B \\ \text{SCOPE } A \end{array} \right] \\ \text{else } A \end{array}$$

It is defined formally within the HPSG formalism in (3.44).

$$(3.44) \quad \begin{array}{l} \text{combine-semantic}(\textit{circumstance } \boxed{1}, \textit{quantifier } \boxed{2}) \Leftrightarrow \\ \left[\begin{array}{l} \text{QUANT } \boxed{2} \\ \text{SCOPE } \boxed{1} \end{array} \right] \\ \text{combine-semantic}(\neg \textit{circumstance } \boxed{1}, \textit{quantifier } \boxed{2}) \Rightarrow \boxed{1} \\ \text{combine-semantic}(\textit{circumstance } \boxed{1}, \neg \textit{quantifier } \boxed{2}) \Rightarrow \boxed{1} \\ \text{combine-semantic}(\neg \textit{circumstance } \boxed{1}, \neg \textit{quantifier } \boxed{2}) \Rightarrow \boxed{1} \end{array}$$

Then semantic interpretation would work exactly as in P&S1. That is, semantic interpretation depends on functor-argument structure (subcategorisation requirements). Lexical functor signs identify semantic roles in the functor semantics with the semantics of their arguments directly. (This works equally well for head-complement, head-adjunct and head-specifier signs.) Thus the problems of an interpretive semantics associated with some other treatments of discontinuous constituency are avoided. That is, we don't have to "figure out" how to put the semantic interpretations of discontinuous constituents together since the semantics is given by the functor-argument structure. Therefore, the semantics is inherently compositional.

However, I propose to eliminate the Semantics Principle entirely and let functors determine the semantics directly as in varieties of categorial unification grammar. For example, a simplified version of the sign for *kicks* might look like (3.45) (where the subscript indexes on $\text{NP}[\text{NOM}]_1$ and $\text{NP}[\text{ACC}]_2$ are coindexed with the values of the thematic (argument) roles of $\text{SEM}|\text{CONT}$ of each sign).

$$(3.45) \left[\begin{array}{l} \text{PHON} \quad \text{kicks} \\ \text{SYN}|\text{LOC}|\text{ARGS} \quad \langle \text{NP}[\text{NOM}]_1, \text{NP}[\text{ACC}]_2 \rangle \\ \text{SEM}|\text{CONT} \quad \left[\begin{array}{l} \text{RELN} \quad \text{kick} \\ \text{AGENT} \quad 1 \\ \text{PATIENT} \quad 2 \end{array} \right] \end{array} \right]$$

In this example, the semantics of the subject is coindexed with the semantics of the **AGENT** role and the semantics of the object with the **PATIENT** role so there is no need for a Semantics Principle. Of course this does nothing about quantifier semantics of NPs for example where the semantics of the NP should be "wrapped around" the semantics of the verb. However, determining scope of quantification directly at this level is inappropriate. If the semantics of proper names is treated exactly like the treatment of quantified NPs then this presents no difficulties. In any case, NP complements are the worst case. In most other cases, the functor can combine via coindexing without any difficulties whatsoever. For example, the semantics of a sentential complement will just fill a **PROPOSITION** role in the functor without further manipulation.

If we are willing to accept this approach to NP argument semantics then there are two questions to be answered. First, how do adjuncts select their heads and secondly, what is the treatment of specifiers (since they are not complements as in P&S1)? We will answer these two questions in the following sections.

3-3 Semantic Representations

3-3.1 The HPSG Treatment of Semantic Interpretation

The treatment of semantics in P&S1 is substantially revised in P&S2. Therefore, we will only present a review of semantic representation as presented in P&S2. The smallest basic unit of the semantic structure is the *parameterised state of affairs* or *psoa* for short. It is a feature structure which contains a **RELATION** attribute indicating the basic property or predicate and other features which mark the thematic roles of the arguments of the predicate. For example, the *psoa* for *gives* is (3.46).

$$(3.46) \left[\begin{array}{ll} \mathbf{RELATION} & \mathbf{GIVE} \\ \mathbf{AGENT} & \boxed{1} \\ \mathbf{GOAL} & \boxed{2} \\ \mathbf{THEME} & \boxed{3} \end{array} \right]$$

The **RELATION** is the atom **GIVE** which encodes the *give* predicate. There are three other features, **AGENT**, **GOAL** and **THEME** corresponding to the agent, goal and theme thematic roles respective of the predicate *give*.

Common nouns have a **CONTENTS** value of sort *referential-object* (*ref-obj*). The content value of the common noun *book* is shown in (PS2:7).

$$(PS2:7) \left[\begin{array}{l} \mathbf{PARAMETER} \quad \boxed{1} \\ \mathbf{RESTRICTION} \left\{ \begin{array}{l} \mathbf{RELATION} \quad \mathbf{BOOK} \\ \mathbf{INSTANCE} \quad \boxed{1} \end{array} \right\} \end{array} \right] \left[\begin{array}{l} \mathbf{INDEX} \\ \mathbf{PERSON} \quad \mathbf{3RD} \\ \mathbf{NUMBER} \quad \mathbf{SNG} \\ \mathbf{GENDER} \quad \mathbf{NEU} \end{array} \right] \left[\begin{array}{l} \mathbf{PERSON} \quad \mathbf{3RD} \\ \mathbf{NUMBER} \quad \mathbf{SNG} \\ \mathbf{GENDER} \quad \mathbf{NEU} \end{array} \right]$$

npro *index*

psoa

ref-obj

A *ref-obj* has several parts. First, it contains a **PARAMETER** feature whose value is a feature structure with the feature **INDEX**. The value of the index feature is a feature structure of sort *index*. An *index* contains the three features **PERSON**, **NUMBER** and **GENDER**. Therefore, HPSG is a semantic theory of agreement.⁶ Furthermore, a *ref-obj* can contain a set of *restrictions* which restrict the set of referents that the *index* can refer to. The set of such restrictions is the value of

⁶We also follow HPSG in adopting a semantic theory of agreement but instead use sorts to encode agreement information.

the **RESTRICTION** feature. Restrictions are also *psaos*. Notice that in the **CONTENT** of *book*, the argument of the *book* relation is coindexed with the *index* of the **PARAMETER**. This guarantees that the referent of the **INDEX** of the **CONTENT** value is in fact a book.⁷

The semantic representations of intransitive, transitive and ditransitive verbs is particularly simple. In each case, the **CONTENT** of the verb is a *psoa* which contains one, two or three additional features respectively encoding thematic roles. In each case, the *index* of each subcategorised NP is coindexed (via subscripting of the NP with an index) with the value of the appropriate thematic role feature. The schematic signs for the verbs *walks*, *sees* and *gives* is given in (PS2:10)–(PS2:12).

$$(PS2:10) \left[\begin{array}{l} \text{CAT} \\ \text{CONTENT} \end{array} \left[\begin{array}{l} \text{HEAD VERB[FIN]} \\ \text{SUBCAT } \langle \text{NP[NOM]}_{[1]} \rangle \\ \text{RELATION WALK} \\ \text{AGENT } [1] \end{array} \right] \right]$$

$$(PS2:11) \left[\begin{array}{l} \text{CAT} \\ \text{CONTENT} \end{array} \left[\begin{array}{l} \text{HEAD VERB[FIN]} \\ \text{SUBCAT } \langle \text{NP[NOM]}_{[1]}, \text{NP[ACC]}_{[2]} \rangle \\ \text{RELATION SEE} \\ \text{EXPERIENCER } [1] \\ \text{THEME } [2] \end{array} \right] \right]$$

⁷The basic differences between the treatment of the **CONTENTS** value in P&S1 and P&S2 can be illuminated by considering the following sign which presents the value of the **CONTENTS** attribute as defined in P&S1.

(PS2:fn24i)

$$\begin{array}{l} \text{ind-obj} \left[\begin{array}{l} \text{INDEX} \\ \text{RESTRICTION} \end{array} \left[\begin{array}{l} \text{VARIABLE } [1] \\ \text{RESTRICTION} \end{array} \right] \right] \\ \text{index} \left[\begin{array}{l} \text{variable} \left[\begin{array}{l} \text{PERSON 3RD} \\ \text{NUMBER SNG} \\ \text{GENDER NEU} \end{array} \right] \\ \text{psoa} \left[\begin{array}{l} \text{RELATION BOOK} \\ \text{INSTANCE } [1] \end{array} \right] \end{array} \right] \end{array}$$

$$(PS2:12) \left[\begin{array}{l} \text{CAT} \\ \\ \\ \text{CONTENT} \end{array} \left[\begin{array}{ll} \text{HEAD} & \text{VERB[FIN]} \\ \text{SUBCAT} & \langle \text{NP[NOM]}_{[1]}, \text{NP[ACC]}_{[2]}, \text{NP[ACC]}_{[3]} \rangle \\ \text{RELATION} & \text{GIVE} \\ \text{AGENT} & [1] \\ \text{GOAL} & [2] \\ \text{THEME} & [3] \end{array} \right] \right]$$

3-3.2 The Syntax of Semantic Representations

The semantic representation “language” is based fairly closely on the representation language InL (*Indexed Language*) developed by Henk Zeevat in the context of Unification Categorical Grammar ([47]). In InL formulas are terms of the form $[e]foo(t_1, \dots, t_n)$ where the argument roles of HPSG are specified by position rather than explicitly named. I will represent formulas as feature structures as in HPSG where the argument roles are explicitly named as attributes of a complex feature structure. However, I make several simplifications to achieve uniformity of structure of the representations in keeping with the practice of InL where every formula and subformula has the same basic structure. Every level of semantic description of the value of a CONTENTS feature is based on the pattern in (3.47).

$$(3.47) \left[\begin{array}{ll} \text{INDEX} & [0] \\ \text{RELATION} & \langle \text{atomic symbol} \rangle \\ \langle \text{argument role } 1 \rangle & \langle \text{semantic representation } 1 \rangle \\ \dots & \dots \\ \langle \text{argument role } n \rangle & \langle \text{semantic representation } n \rangle \end{array} \right]$$

The INDEX is a distinguished sorted variable which represents the ontological type of the semantic object. RELATION is the same as the feature of the same name in the CONTENTS value in HPSG. However, unlike HPSG, the relation can be a property, a predicate, a logical connective or a quantifier. $\langle \text{argument role } 1 \rangle$ – $\langle \text{argument role } n \rangle$ are the argument roles of the RELATION. For properties, the INDEX replaces the single argument role INSTANCE (INST) of HPSG. For predicates, the argument roles are *thematic roles* (e.g., AGENT, PATIENT, THEME, etc.), for connectives, conjuncts, antecedents, consequents and the like and for quantifiers, the variable which is quantified (QUANTIFIED-INDEX or abbreviated QIND) and the proposition (PROPOSITION) in the scope of the quantifier.

There may be zero or more argument roles. They are usually similar or identical to the roles which appear in P&S2 though there are exceptions.

A simple property, such as $\text{man}(x)$, is represented in P&S2 as (3.48) whereas we represent it as in (3.49).

$$(3.48) \left[\begin{array}{ll} \text{RELATION} & \text{MAN} \\ \text{INSTANCE} & \boxed{1} \end{array} \right]$$

$$(3.49) \left[\begin{array}{ll} \text{INDEX} & \boxed{1} \\ \text{RELATION} & \text{MAN} \end{array} \right]$$

A predicate, such as $\text{give}(x,y,z)$, is represented in P&S2 as (3.50) whereas we represent it as in (3.51).

$$(3.50) \left[\begin{array}{ll} \text{RELATION} & \text{GIVE} \\ \text{AGENT} & \boxed{1} \\ \text{GOAL} & \boxed{2} \\ \text{THEME} & \boxed{3} \end{array} \right]$$

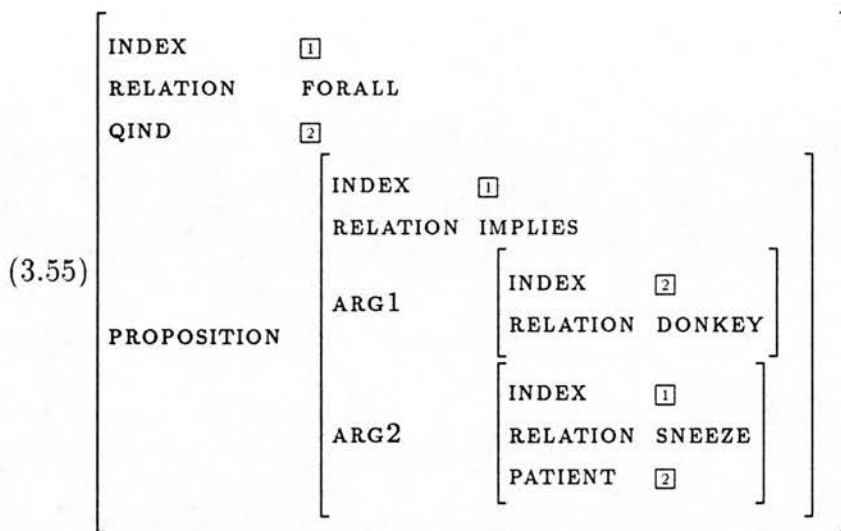
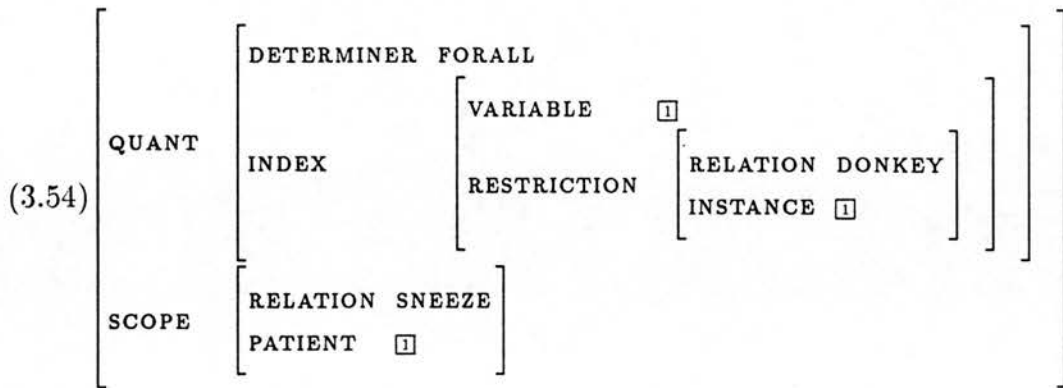
$$(3.51) \left[\begin{array}{ll} \text{INDEX} & \text{event } \boxed{0} \\ \text{RELATION} & \text{GIVE} \\ \text{AGENT} & \boxed{1} \\ \text{GOAL} & \boxed{2} \\ \text{THEME} & \boxed{3} \end{array} \right]$$

A connective, such as $\phi \wedge \psi$, is represented in P&S2 as (3.52) whereas we represent it as in (3.53).

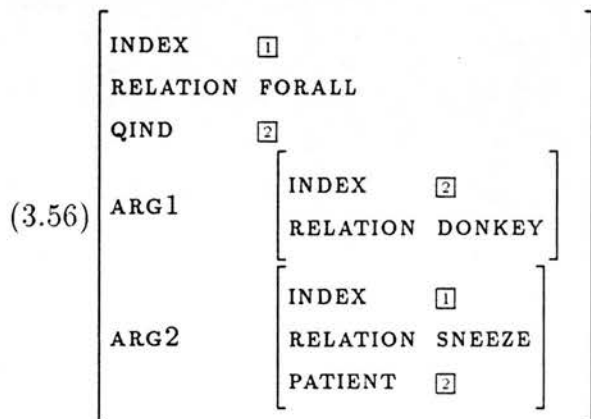
$$(3.52) \left[\begin{array}{ll} \text{CONNECTIVE} & \text{AND} \\ \text{JUNCTS} & \{\boxed{1}, \boxed{2}\} \end{array} \right]$$

$$(3.53) \left[\begin{array}{ll} \text{INDEX} & \boxed{0} \\ \text{RELATION} & \text{AND} \\ \text{ARG1} & \boxed{1} \\ \text{ARG2} & \boxed{2} \end{array} \right]$$

A quantified formula, such as $\forall x.(\text{donkey}(x) \rightarrow \text{sneeze}(x))$, is represented in P&S2 as (3.54) whereas we represent it as in (3.55).



If we wanted to adopt a generalised quantifier treatment of natural language quantifiers, (3.55) would take the form in (3.56) or perhaps (3.57) where $\boxed{2}$ is the index of ARG1 in (3.56) or the index of QUANT in (3.57).



$$(3.57) \left[\begin{array}{ll} \text{INDEX} & \boxed{1} \\ \text{RELATION} & \text{FORALL} \\ \text{QIND} & \boxed{2} \\ \text{QUANT} & \left[\begin{array}{ll} \text{INDEX} & \boxed{2} \\ \text{RELATION} & \text{DONKEY} \end{array} \right] \\ \text{SCOPE} & \left[\begin{array}{ll} \text{INDEX} & \boxed{1} \\ \text{RELATION} & \text{SNEEZE} \\ \text{PATIENT} & \boxed{2} \end{array} \right] \end{array} \right]$$

3-3.3 First Order Translation of Semantic Representations

In this section we show how semantic representations of the schematic form (3.58) can be translated into formulas of sorted first order logic.

$$(3.58) \left[\begin{array}{ll} \text{INDEX} & \boxed{0} \\ \text{RELATION} & \langle \text{atomic symbol} \rangle \\ \langle \text{argument role } 1 \rangle & \langle \text{semantic representation } 1 \rangle \\ \dots & \dots \\ \langle \text{argument role } n \rangle & \langle \text{semantic representation } n \rangle \end{array} \right]$$

Let τ be a translation function from semantic representations to formulas of sorted first order logic. Then, the nonce representation

$$(3.59) \left[\begin{array}{ll} \text{INDEX} & \boxed{0} \\ \text{RELATION} & \text{FOO} \\ \text{ARG1} & \boxed{1} \\ \text{ARG2} & \boxed{2} \\ \text{ARG3} & \boxed{3} \end{array} \right]$$

is translated as

$$(3.60) \exists x_0 x_1 x_2 x_3. \text{foo}(x_0, x_1, x_2, x_3) \wedge \text{arg1}(x_0, x_1) \wedge \text{arg2}(x_0, x_2) \wedge \text{arg3}(x_0, x_3) \wedge \tau(\boxed{1}) \wedge \tau(\boxed{2}) \wedge \tau(\boxed{3})$$

where x_0, x_1, x_2 and x_3 are sorted first order variables.

If the **RELATION** is a predicate symbol (and not a logical connective or a quantifier) then the semantic representation is of the following schematic form.

$$(3.61) \left[\begin{array}{ll} \text{INDEX} & \boxed{0} \\ \text{RELATION} & \langle \text{atomic symbol} \rangle \\ \langle \text{thematic role } 1 \rangle & \langle \text{semantic representation } 1 \rangle \\ \dots & \dots \\ \langle \text{thematic role } n \rangle & \langle \text{semantic representation } n \rangle \end{array} \right]$$

Then, the nonce representation

$$(3.62) \left[\begin{array}{ll} \text{INDEX} & \boxed{0} \\ \text{RELATION} & \text{GIVE} \\ \text{AGENT} & \boxed{1} \\ \text{GOAL} & \boxed{2} \\ \text{THEME} & \boxed{3} \end{array} \right]$$

is translated as in (3.63).

$$(3.63) \exists x_0 x_1 x_2 x_3. \text{give}(x_0, x_1, x_2, x_3) \wedge \\ \text{agent}(x_0, x_1) \wedge \text{goal}(x_0, x_2) \wedge \text{theme}(x_0, x_3) \wedge \\ \tau(\boxed{1}) \wedge \tau(\boxed{2}) \wedge \tau(\boxed{3})$$

If the **RELATION** is a binary connective then the semantic representation is of the following schematic form.

$$(3.64) \left[\begin{array}{ll} \text{INDEX} & \boxed{0} \\ \text{RELATION} & \langle \text{connective} \rangle \\ \text{ARG1} & \langle \text{semantic representation } 1 \rangle \\ \text{ARG2} & \langle \text{semantic representation } 2 \rangle \end{array} \right]$$

Then, the nonce representation

$$(3.65) \left[\begin{array}{ll} \text{INDEX} & \boxed{0} \\ \text{RELATION} & \text{AND} \\ \text{ARG1} & \boxed{1} \\ \text{ARG2} & \boxed{2} \end{array} \right]$$

is translated as (3.66).

$$(3.66) \tau(\boxed{1}) \wedge \tau(\boxed{2})$$

If the **RELATION** is a first order quantifier then the semantic representation is of the following schematic form.

$$(3.67) \left[\begin{array}{ll} \text{INDEX} & \boxed{0} \\ \text{RELATION} & \langle \textit{connective} \rangle \\ \text{QIND} & \boxed{1} \\ \text{PROPOSITION} & \langle \textit{semantic representation} \rangle \end{array} \right]$$

Then, the nonce representation

$$(3.68) \left[\begin{array}{ll} \text{INDEX} & \boxed{0} \\ \text{RELATION} & \text{FORALL} \\ \text{QIND} & \boxed{1} \\ \text{PROPOSITION} & \boxed{2} \end{array} \right]$$

is translated as (3.69).

$$(3.69) \forall x_1. (\tau(\boxed{2}))$$

Likewise, the nonce representation

$$(3.70) \left[\begin{array}{ll} \text{INDEX} & \boxed{0} \\ \text{RELATION} & \text{EXISTS} \\ \text{QIND} & \boxed{1} \\ \text{PROPOSITION} & \boxed{2} \end{array} \right]$$

is translated as (3.71).

$$(3.71) \exists x_1. (\tau(\boxed{2}))$$

Finally, $\tau(\boxed{2})$ is x_2 .

3–3.4 The Syntax of InL

The syntax of InL is defined as follows.⁸

⁸This section and the next are taken almost verbatim from §4.1 and §4.2 of [47, pp106–109].

The basic syntax of the representation language InL is given by definitions (3.72) and (3.73). Let L be a set of predicates, sorts and individual constants.

- (3.72) If n is a natural number and s is a sort of L, $\langle n, s \rangle$ is a variable of L.
If s is a sort of L and c is an individual constant of L, $\langle c, s \rangle$ is a constant of L.
- (3.73) 1. If P is an n -place predicate of L, t is a variable and t_1, \dots, t_n are terms then $[t]P(t_1, \dots, t_n)$ is a formula.
2. If t is a variable and t_1 is a term, then $[t] = t_1$ is a formula.
3. If t is a variable and A and B are formulae, then $[t][A \Rightarrow B]$ and $[t][A, B]$ are formulae.
4. \perp is a formula.

“Though already quite powerful, the syntax needs extensions to deal with phenomena like propositions, properties and other quantifiers in order to deal with a more substantial fragment of natural language. Some possible extensions are given in (3.74). We will later add (3.74) for quantifiers.”

- (3.74) if c is a variable and $[a]A$ and $[b]B$ are formulas then $[c]Q([a]A, [b]B)$ is a formula if Q is one of *most*, *few*, *many*, *three*, etc., i.e., if Q is a generalised quantifier.

Some examples of representations for expressions are given in (3.75).

- (3.75) John owns a donkey.
 $[s][[x]donkey, [s]owns(john, x)]$
- Every donkey cries.
 $[E][[x]donkey \rightarrow [e]cry(x)]$
- If John beats a donkey, it cries.
 $[E][[e][[x]donkey, [e]beat(j, x)] \rightarrow [e']cry(x)]$
- In the park, three men sleep.
 $[s'][[y]park, [s']in(y), [s']three([x]man, [s]sleep(x))]$
- Some donkeys cry in the park.
 $[E]some([x]donkey, [e][[y]park, [e]in(y), [e]cry(x)])$

“Every individual term is sorted. Sorts are taken from the following tree (Fig. 3-1). The place in the tree defines their behaviour under

unification: incomparable sorts do not unify, comparable sorts unify to the lower of the two. The tree is also the basis for the interpretation of the sorts.”

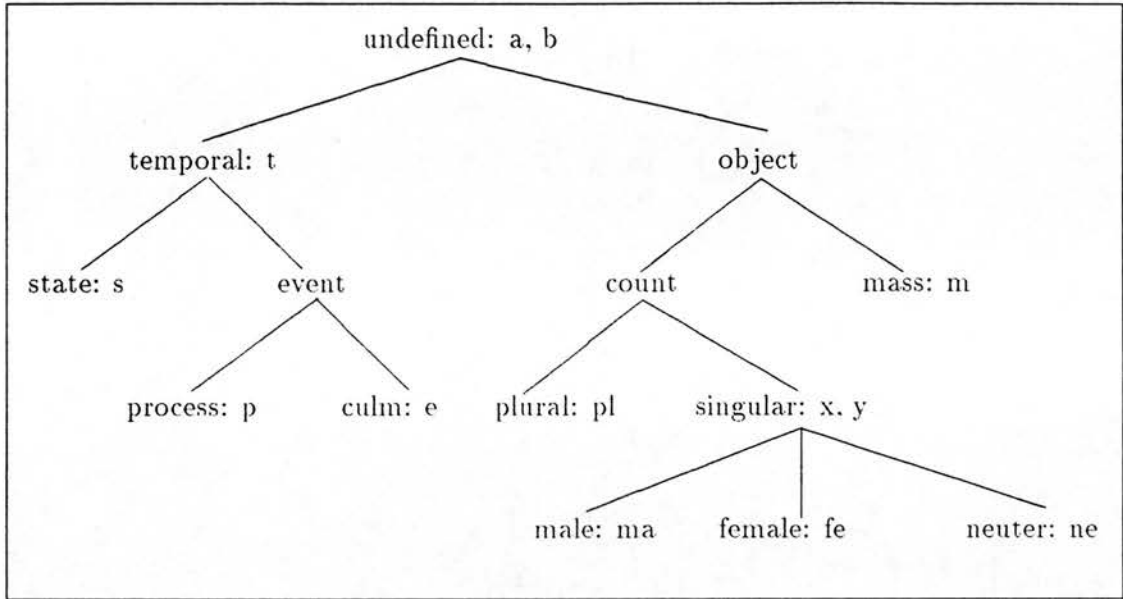


Figure 3-1: A tree of sorts

“Of course, (3-1) is not intended as a final statement of the sorts required. The basis of the selection above are the requirements for dealing with agreement, aspect and certain ambiguities in UCG.”

3-3.5 The Semantics of InL — InL and DRT

“The relation between InL and DRT is not complex given that InL just adds some structure to the DRT syntax. The notion of a discourse referent is recaptured by a recursion on the formula.”

$$(3.76) \quad \begin{aligned} DR(A) &= \{x\} \text{ iff } x \text{ is } A\text{'s index and } A \text{ is atomic or an implication} \\ DR(A) &= \{x\} \cup DR(B) \cup DR(C) \text{ iff } A = [x][A, B] \end{aligned}$$

This definition allows a DRT-style interpretation for the basic part of InL.

$$(3.77) \quad M = \langle U, S, F \rangle$$

where U is a nonempty set of objects, S is a function from sorts into U that respects the tree (i.e., $s_1 \leq s_2$ in the tree entails $S(s_1) \subseteq S(s_2)$ and if $s_1 \not\leq s_2$ and

if $s_2 \not\leq s_1$ then $S(s_1) \cap S(s_2) = \emptyset$ and F a function that interprets the constants in the usual way, depicting individual constants in the extension of their sort. g is a correct assignment for the variables if whenever x has sort s , $g(x) \in S(s)$.

We can now define t^g as usual

$$(3.78) \quad \begin{aligned} t^g &= g(t) \text{ iff } t \text{ is a variable} \\ t^g &= F(t) \text{ iff } t \text{ is a constant} \end{aligned}$$

and, using $f =_{DR(A)} g$ for the relation between assignments that holds iff (3.79), we can define $M \models \varphi[g]$ in (3.80).

$$(3.79) \quad \forall v (fv = gv \vee v \in DR(A))$$

$$(3.80) \quad \begin{aligned} 1. \quad & M \models [t]Pt_1 \dots t_n[g] \text{ iff } \langle t^g, t_1^g, \dots, t_n^g \rangle \in F(P) \\ 2. \quad & M \models [t_1] = [t_2][g] \text{ iff } t_1^g = t_2^g \\ 3. \quad & M \models [x][A, B][g] \text{ iff } M \models A[g] \text{ and } M \models B[g] \\ 4. \quad & M \models [x][A \Rightarrow B][g] \text{ iff } \forall h =_{DR(A)} g (M \models A[h] \Rightarrow \exists f =_{DR(A)} \\ & \quad gM \models B[f]) \\ 5. \quad & M \not\models \perp[g] \end{aligned}$$

Finally, truth is defined in (3.81).

$$(3.81) \quad M \models A \text{ iff there is an } h \text{ with } \text{dom}(h) = DR(A) \text{ such that } M \models A[h].$$

For a discussion of generalised quantifiers, see [47].

3–3.6 InL Translation of Semantic Representations

Rather than give an explicit semantics for the semantic representations proposed here, we implicitly give a semantics via translation of our semantic representations into formulas of InL by cases. Let ι be a translation function from semantic representations to formulas of InL.

If the representation is a universally quantified proposition of the type in (3.82),

it is translated as in (3.83).

$$(3.82) \left[\begin{array}{ll} \text{INDEX} & \boxed{1} \\ \text{RELATION} & \text{FORALL} \\ \text{QIND} & \boxed{2} \\ \text{PROPOSITION} & \left[\begin{array}{ll} \text{INDEX} & \boxed{1} \\ \text{RELATION} & \text{IMPLIES} \\ \text{ARG1} & \boxed{3} \\ \text{ARG2} & \boxed{4} \end{array} \right] \end{array} \right]$$

$$(3.83) [\iota(\boxed{1})][[\iota(\boxed{2})]\iota(\boxed{3}) \Rightarrow [\iota(\boxed{1})]\iota(\boxed{4})]$$

Similarly, if the representation is an existentially quantified proposition of the type in (3.84), it is translated as in (3.85).

$$(3.84) \left[\begin{array}{ll} \text{INDEX} & \boxed{1} \\ \text{RELATION} & \text{EXISTS} \\ \text{QIND} & \boxed{2} \\ \text{PROPOSITION} & \left[\begin{array}{ll} \text{INDEX} & \boxed{1} \\ \text{RELATION} & \text{AND} \\ \text{ARG1} & \boxed{3} \\ \text{ARG2} & \boxed{4} \end{array} \right] \end{array} \right]$$

$$(3.85) [\iota(\boxed{1})][[\iota(\boxed{2})]\iota(\boxed{3}), [\iota(\boxed{1})]\iota(\boxed{4})]$$

If the representation is a conjunction of the type in (3.86), it is translated as in (3.87).

$$(3.86) \left[\begin{array}{ll} \text{INDEX} & \boxed{0} \\ \text{RELATION} & \text{AND} \\ \text{ARG1} & \boxed{1} \\ \text{ARG2} & \boxed{2} \end{array} \right]$$

$$(3.87) [\iota(\boxed{0})][\iota(\boxed{1}), \iota(\boxed{2})]$$

If the representation is a predication of the type in (3.88), it is translated as in (3.89).

$$(3.88) \left[\begin{array}{ll} \text{INDEX} & \boxed{0} \\ \text{RELATION} & \text{FOO} \\ \text{ARG1} & \boxed{1} \\ \text{ARG2} & \boxed{2} \\ \text{ARG3} & \boxed{3} \end{array} \right]$$

$$(3.89) [\iota(\boxed{0})][foo(\iota(\boxed{1}), \iota(\boxed{2}), \iota(\boxed{3}))]$$

Notice that in the absence of any theory of argument or thematic roles, they are simply dropped in the translation as this information is recoverable from the position in the argument list of the translation.

If the representation is a property of the type in (3.90), it is translated as in (3.91).

$$(3.90) \left[\begin{array}{ll} \text{INDEX} & \boxed{0} \\ \text{RELATION} & \text{FOO} \end{array} \right]$$

$$(3.91) [\iota(\boxed{0})][foo]$$

Finally, $\iota(s^{\boxed{n}})$ is translated as the variable $\langle n, s \rangle$.

The translation function ι defines a complete map from semantic representations written as feature structures into formulas of InL. In particular, universal quantifications are translated into sorted implications whose semantics is defined in terms of universal quantifications over assignments to indexes or variables and existential quantifications are translated into simple conjunctions with an existential semantic interpretation. Thus, given ι , there is no need to supply the semantic representations discussed in this section and the following sections with a model theoretic semantics. Zeevat ([47, Ch. 4]) discusses a full range of extensions to InL to deal with generalised quantifiers, propositions, modification and the like. These extensions can also be encoded straightforwardly in the indexed semantic representations given here. However, since it is not my purpose to present a fully developed theory of natural language semantics here, I refer the reader to [47, Ch. 4].

3-4 The Syntax and Semantics of Adjuncts

3-4.1 The hpsg Treatment of Adjuncts

P&S1 presents an account of adjuncts in which heads select for the adjuncts

which modify them. The result is a “flat” *head-adjunct-complement-structure* of the form (3.92).

$$(3.92) \quad \left[\begin{array}{c} \text{DTRS} \\ \text{head-adjunct-complement-structure} \end{array} \left[\begin{array}{l} \text{HEAD-DTR} \quad [] \\ \text{COMP-DTRS} \quad \langle \dots \rangle \\ \text{ADJ-DTRS} \quad \langle \dots \rangle \end{array} \right] \right]$$

This structure allows for adjuncts to appear as sisters to the complements of the head that selects them. Although this treatment was already formally complicated enough, it resisted an adequately elegant semantic treatment. Therefore, P&S2 adopts a treatment where adjuncts select the single head which they modify.⁹

In P&S2, the sign for the attributive (non-predicative [PRD -]) restrictive adjective *red* is (PS2:63):

$$(PS2:63) \quad \left[\begin{array}{c} \text{CAT} \\ \text{CONTENT} \end{array} \left[\begin{array}{l} \text{HEAD} \quad \left[\begin{array}{l} \text{MOD } \bar{N}: \left[\begin{array}{l} \text{PARA} \quad [1] \\ \text{RESTR} \quad [2] \end{array} \right] \\ \text{PRD} \quad - \end{array} \right] \\ \text{SUBCAT} \quad \langle \rangle \end{array} \right] \quad \text{adj} \\ \text{ref-obj} \left[\begin{array}{l} \text{PARA} \quad [1] \\ \text{RESTR} \quad [2] \cup \left\{ \left[\begin{array}{l} \text{RELN} \quad \text{RED} \\ \text{ARG} \quad [1] \end{array} \right] \right\} \end{array} \right] \end{array} \right] \right]$$

where \bar{N} is (PS2:54).

$$(PS2:54) \quad \left[\text{LOC} \left[\text{CAT} \left[\begin{array}{l} \text{HEAD} \quad \text{NOUN} \\ \text{SUBCAT} \quad \langle \text{DET} \rangle \end{array} \right] \right] \right]$$

The feature **MODIFIER** (**MOD**) plays the role of the **SUBCAT** feature for adjuncts and selects a single argument for the adjunct.¹⁰ For a restrictive adjective like

⁹It should be noted that Bob Kasper has presented an account which allows multiple adjuncts to be sisters to the head and complements in [24] with a fully developed theory of semantic compositionality. It requires some relatively complex definitions of two recursive functional dependencies. For further information, see [24].

¹⁰It is unclear to me why **MOD** is a **HEAD** feature. Just as subcategorisation requirements are “discharged” by deletion of elements of the head’s **SUBCAT** list from the mother’s **SUBCAT** list, it appears that the feature **MOD** should be treated accordingly.

the attributive adjunct *red*, the **RESTRICTION** of the head \bar{n} is unioned with the restriction (PS2:64).

$$(PS2:64) \quad \left\{ \left[\begin{array}{ll} \mathbf{RELN} & \mathbf{RED} \\ \mathbf{ARG} & \boxed{1} \end{array} \right] \right\}$$

The ID schema which forms phrases from an adjunct and its head is Schema 5, 65.

(65) (Schema 5)

a phrase with **DTRS** value of sort *head-adjunct-structure*, such that the **MOD** value of the adjunct daughter is unified with the **SYNSEM** value of the head daughter.

In other words, the schematic structure is that of 3.93.

$$(3.93) \quad \left[\mathbf{DTRS} \left[\begin{array}{l} \mathbf{HEAD-DTR} | \mathbf{SYNSEM} \boxed{1} \\ \mathbf{ADJUNCT-DTR} | \mathbf{CAT} | \mathbf{HEAD} | \mathbf{MOD} \boxed{1} \end{array} \right] \right]$$

The adjunct in a sign formed by Schema 5 contains the **CONTENT** of the phrase due to the use of the \cup symbol in the definition of the **SYNSEM** value of the adjunct. Thus the **CONTENT** of the **ADJUNCT-DTR** should be unified with the **CONTENT** of the mother. The *Semantics Principle (second version)* ensures that this is the case.

(67) Semantics Principle (second version)

In a headed phrase, the **CONTENT** value is unified with that of the adjunct daughter if the **DTRS** value is of sort *head-adjunct-structure*, and with that of the head daughter otherwise.

Thus, the **SYNSEM** value of the sign of the phrase *red book* is (PS2:66).

$$(PS2:66) \quad \left[\begin{array}{l} \mathbf{CAT} \left[\begin{array}{ll} \mathbf{HEAD} & \mathbf{NOUN} \\ \mathbf{SUBCAT} & \langle \mathbf{DET} \rangle \end{array} \right] \\ \mathbf{CONT} \left[\begin{array}{l} \mathbf{PARA} \boxed{1} \\ \mathbf{RESTR} \left\{ \left[\begin{array}{ll} \mathbf{RELN} & \mathbf{BOOK} \\ \mathbf{INST} & \boxed{1} \end{array} \right] \cdot \left[\begin{array}{ll} \mathbf{RELN} & \mathbf{RED} \\ \mathbf{ARG} & \boxed{1} \end{array} \right] \right\} \end{array} \right] \end{array} \right]$$

by an NP which is coindexed with the NP subject of the VP argument. Finally, the head features of the rule sign and the argument VP will be coindexed.

The same strategy will work for all the other adjuncts. For example, consider *that*-less relative clauses. Then we will need a rule of the form (3.95).

$$(3.95) \left[\begin{array}{l} \text{SYN|LOC} \left[\begin{array}{l} \text{HEAD } \boxed{1} \\ \text{ARGS } \langle \text{N} \left[\begin{array}{l} \text{SYN|LOC} \left[\begin{array}{l} \text{HEAD } \boxed{1} \\ \text{ARGS } \langle \rangle \end{array} \right] \right] \rangle \end{array} \right] \\ \text{DTRS|TYPE-CHANGE-DTR } \text{S[COMP } -, \text{ SLASH } \{\text{NP}\}] \end{array} \right] \end{array} \right]$$

Clearly we are missing a generalisation though. All adjunct rules will be of the form $X/X \rightarrow Y$ as in (3.96) where $\boxed{3} = Y$ and X/X is the entire sign.

$$(3.96) \left[\begin{array}{l} \text{SYN|LOC} \left[\begin{array}{l} \text{HEAD } \boxed{1} \\ \text{ARGS } \langle \left[\begin{array}{l} \text{SYN|LOC} \left[\begin{array}{l} \text{HEAD } \boxed{1} \\ \text{ARGS } \boxed{2} \end{array} \right] \right] \rangle \circ \boxed{2} \end{array} \right] \\ \text{DTRS|TYPE-CHANGE-DTR } \boxed{3} \end{array} \right] \end{array} \right]$$

Since this information is invariant across type-changed modifiers, it should be stated independently as in (3.96). Then to specify a particular type change (which amounts to the same thing as letting an adjunct specify what it can modify) we can make a simple statement about what category Y a modifier X modifies. We could just state a relational dependency of the type *modifies*(X,Y) and then have the implication (3.97).

$$(3.97) \quad \text{modifies}(\boxed{1},\boxed{2}) \Rightarrow \left[\begin{array}{l} \text{SYN|LOC} \left[\begin{array}{l} \text{HEAD } \boxed{3} \\ \text{ARGS } \langle \boxed{2} \left[\begin{array}{l} \text{SYN|LOC} \left[\begin{array}{l} \text{HEAD } \boxed{3} \\ \text{ARGS } \boxed{4} \end{array} \right] \right] \rangle \circ \boxed{4} \end{array} \right] \\ \text{DTRS|TYPE-CHANGE-DTR } \boxed{1} \end{array} \right] \end{array} \right]$$

None of the rules above indicate how semantics is treated. First, we follow the convention of indicating the value of the path `SEMANTICS|CONTENT|INDEX` by a

In general, the SYNTAX and SEMANTICS of such a sign is (3.100).

$$(3.100) \left[\begin{array}{l} \text{SYN|LOC|ARGS} \left\langle \left[\begin{array}{l} \text{SYN|LOC|ARGS} \boxed{2} \\ \text{SEM|CONT} \boxed{3} [\text{INDEX} \boxed{4}] \end{array} \right] \right\rangle \circ \boxed{2} \\ \text{SEM|CONT} \left[\begin{array}{l} \text{INDEX} \boxed{4} \\ \text{RELN AND} \\ \text{CONJ1} [\text{INDEX} \boxed{4}] \\ \text{CONJ2} \boxed{3} \end{array} \right] \end{array} \right]$$

Notice that the SEM|CONT of the adjunct contains a conjunction relation AND which is also given an INDEX. This treatment of assigning every relation and connective an index is borrowed directly from the treatment of semantics in UCG ([48]). It has the advantage that the semantics of a modified VP has the same structure as the semantics of a nonmodified VP. Furthermore, the explicit specification of the path INDEX means that adjuncts do not have to look through the structure of modified phrases to find the relevant index. It is always percolated up by the semantics of the conjunctive relation's CONTENT value.

Now the question is what do we have to do to the VP/VP \rightarrow PP rule to handle the semantics correctly. If we consider the sign for the preposition *at* (3.101), we see that it (and all other predicative (PRD +) prepositions) subcategorise for an NP and an NP[ACC] prepositional object. This gives rise to predicative copula constructions such as *The race is at 6:00* and small clauses such as *the start of the race at 10:00* in *With the start of the race at 10:00, it will be too late to go out afterwards*. Furthermore, the variables of the two arguments are coindexed with the values of the subpaths INDEX and TIME respectively of the path SEM|CONT of the preposition. Therefore the first NP argument of the PP[PRD +] will be unsaturated. So, a PP[PRD +] is in fact a nonlexical head and so the type-changing rule must apply to it.

$$(3.101) \left[\begin{array}{l} \text{PHON } \langle \text{at} \rangle \\ \text{SYN|LOC} \left[\begin{array}{l} \text{HEAD} \left[\begin{array}{l} \text{MAJ P} \\ \text{PRD +} \end{array} \right] \\ \text{LEX +} \\ \text{ARGS} \langle \text{NP} \boxed{1} \cdot \text{NP} [\text{ACC}] \boxed{2} \rangle \end{array} \right] \\ \text{SEM|CONT} \left[\begin{array}{l} \text{INDEX} \boxed{1} \\ \text{RELN AT} \\ \text{TIME} \boxed{2} \end{array} \right] \end{array} \right]$$

What the general adjunct rule has to do with semantics is coindex the INDEX of the modified head with the INDEX of the adjunct and create an AND structure with

To get this to work out, we have to coindex the INDEX of the NP[ACC] and the INDEX of the noun *man* (3.104). (Notice that the entry for *man* only contains an INDEX and not an INSTANCE feature. This is because the INDEX plays the dual role of INDEX and INSTANCE for common nouns. Again this is consistent with UCG practice.)

$$(3.104) \left[\begin{array}{l} \text{PHON } \langle \text{man} \rangle \\ \text{SYN } N \\ \text{SEM|CONT } \left[\begin{array}{l} \text{INDEX } \boxed{1} \\ \text{RELN } \text{MAN} \end{array} \right] \end{array} \right]$$

Then the sign for *John saw* as an N/N is (3.105).

$$(3.105) \left[\begin{array}{l} \text{SYN|LOC } \left[\begin{array}{l} \text{HEAD } \boxed{1} \\ \text{ARGS } \langle N \left[\begin{array}{l} \text{SYN|LOC } \left[\begin{array}{l} \text{HEAD } \boxed{1} \\ \text{ARGS } \langle \rangle \\ \text{SEM|CONT } \boxed{2} [\text{INDEX } \boxed{3}] \end{array} \right] \rangle \end{array} \right] \right. \end{array} \right] \\ \text{SEM|CONT } \left[\begin{array}{l} \text{INDEX } \boxed{3} \\ \text{RELN } \text{AND} \\ \text{CONJ1 } \left[\begin{array}{l} \text{INDEX } \boxed{4} \\ \text{RELN } \text{SEE} \\ \text{EXPERIENCER } \text{John} \\ \text{THEME } \boxed{3} \end{array} \right] \\ \text{CONJ2 } \boxed{2} \end{array} \right] \end{array} \right]$$

Then the sign of *man John saw* is (3.106).

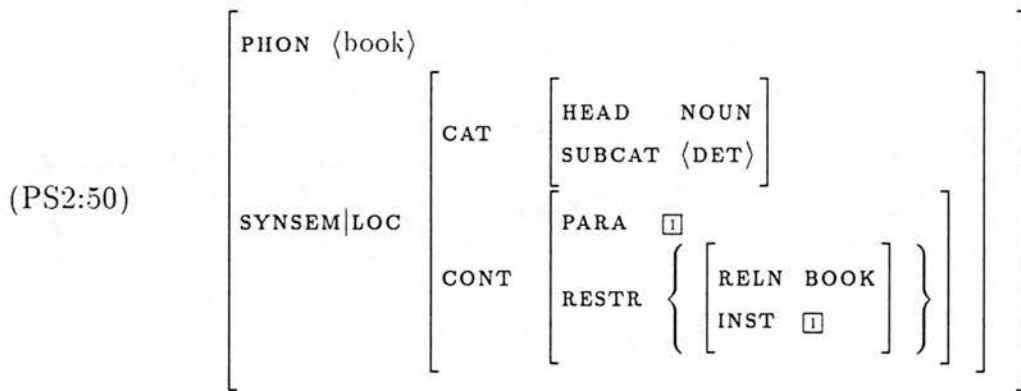
$$(3.106) \left[\begin{array}{l} \text{SYN } N \\ \text{SEM|CONT } \left[\begin{array}{l} \text{INDEX } \boxed{3} \\ \text{RELN } \text{AND} \\ \text{CONJ1 } \left[\begin{array}{l} \text{INDEX } \boxed{4} \\ \text{RELN } \text{SEE} \\ \text{EXPERIENCER } \text{John} \\ \text{THEME } \boxed{3} \end{array} \right] \\ \text{CONJ2 } \left[\begin{array}{l} \text{INDEX } \boxed{3} \\ \text{RELN } \text{MAN} \end{array} \right] \end{array} \right] \end{array} \right]$$

In this case, the index of the modified phrase is not the index of the adjunct *that*-less relative clause *John saw* but rather the index of the NP[ACC] which is in the SLASH set. This is at odds with the treatment of VPs. Therefore a different rule has to be used to accomplish the type-change. Notice that a gap elimination rule which reintroduces the SLASHed element as an argument might solve the problem since in the case of PP adjuncts and relative clauses, the index is that of the unsaturated argument. However, with VPs which are controlled (either functionally or anaphorically) this doesn't seem available. Compare *Running down the road, I came across a very old man*. In this case the variable which is relevant is that of the unsaturated subject of the VP which should be bound with the index of *I*. Again, I will leave such speculations for the future.

3-5 The Syntax and Semantics of Specifiers

3-5.1 The HPSG Treatment of Specifiers

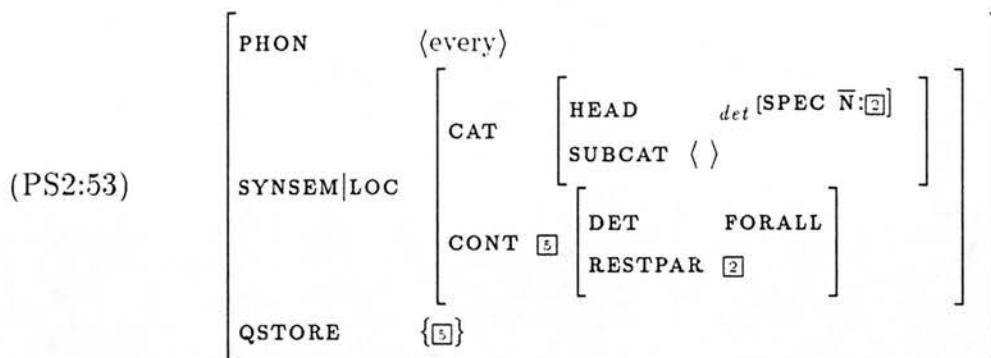
Among the specifier types which are discussed in detail in P&S1 or P&S2 are determiners and possessives. With regard to both syntax and semantics they are very similar, although possessives are more complicated semantically. In both cases, the head noun subcategorises for the determiner or possessive. It is the "last" or least oblique argument. Thus the sign for *book* (PS2:50) subcategorises for a determiner DET.



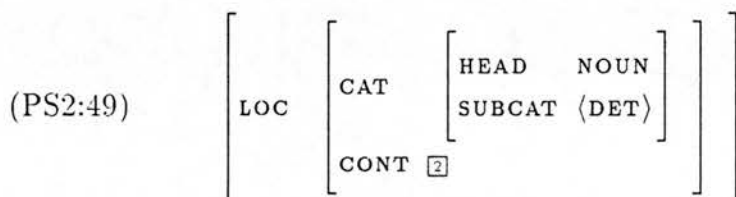
The schematic sign for DET is (PS2:51).



The sign for the determiner *every* is (PS2:53).



There are several things to note about this sign. First, the sign contains a **SPEC HEAD** feature. The **SPEC** feature takes as its value an \bar{N} with semantic content $\boxed{2}$. In more detail, \bar{N} stands for the sign P&S2, Ch. 1, (49). That is, the noun subcategorises for the determiner and the determiner simultaneously selects for the \bar{N} . (\bar{N} 's and N 's are distinguished by the fact that an \bar{N} has exactly one subcategorised element, the determiner.)



The **CONTENTS** of the determiner consists of a **DET** feature which takes the quantifier type as its value (in this case **FORALL**) and a **RESTRICTED-PARAMETER** which is the semantic content of the head \bar{N} . The entire semantic content ($\boxed{5}$) is the singleton set value of the feature **QSTORE** which percolates quantifiers up the tree until they are retrieved. We will have more to say about this in later chapters.

In order for the **SPEC** specification to simultaneously select for the head \bar{N} , something else needs to be done. The *Specifier Principle* P&S2, Ch. 1, (50) requires that the **SPEC** value of a specifier be token-identical to the **SYNSEM** value of the head thus effecting the simultaneous subcategorisation of the head by the determiner. The *Specifier Principle* guarantees that the content of the determiner is unified with the content of the \bar{N} . This unification cannot be lexically specified by the subcategorisation of the noun for then any semantic contribution made by adjuncts to the head noun would be lost in the content of the determiner and likewise in the restricted parameter of the stored quantifier.

(PS2:50) **Spec Principle**

If a non-head daughter in a headed structure bears a **SPEC** value, it is unified with the **SYNSEM** value of the head daughter.

3-5.2 An Alternative Treatment of Specifiers

The categorial treatment of specifiers usually treats them as being of the schematic category XP/X . If we instantiate X to N , then the category type is NP/N . This is the category of determiners. A determiner is a functor with a single $N[\text{MAX } -]$ saturated argument which the functor embeds in its own semantics. Specifiers are of the type $x[\text{MAX } +]/x[\text{MAX } -, \text{SUBCAT } ()]$ where MAX is a feature which indicates that a category is maximal or not. $[\text{MAX } +]$ indicates that a constituent is phrasal and has combined with its specifier. Singular count nouns are specified as $N[\text{MAX } -]$. Plural nouns and mass nouns are unspecified for MAX as they may appear as NPs with or without a specifier.

Contrary to most versions of \bar{X} theory, the subject is not considered to be the specifier of either the VP or IP. This is because only a clause is considered to be saturated and to receive a specifier. This position is more or less forced on us if we allow all of the verb's arguments to be sisters of the verb, i.e., to have a *VP-internal* or *nonconfigurational subject* as we argue for in German. The specifier position in this case is what GB calls [Spec, CP] or the topic position in V2 clauses. S is treated as a [BAR 0] category. $s[\text{COMP } +]$ is treated as a [BAR 1] category and SP (=CP) is treated as a [BAR 2] or phrasal category. That is, [Spec, CP] is the specifier of the clause. So CP is really SP since we are treating the clause as the head instead of the complementiser. SP is then specified as $[\text{MAX } +]$ and S is specified as $[\text{MAX } -]$.

To summarise, instead of the GB \bar{X} treatment we use the following scheme:

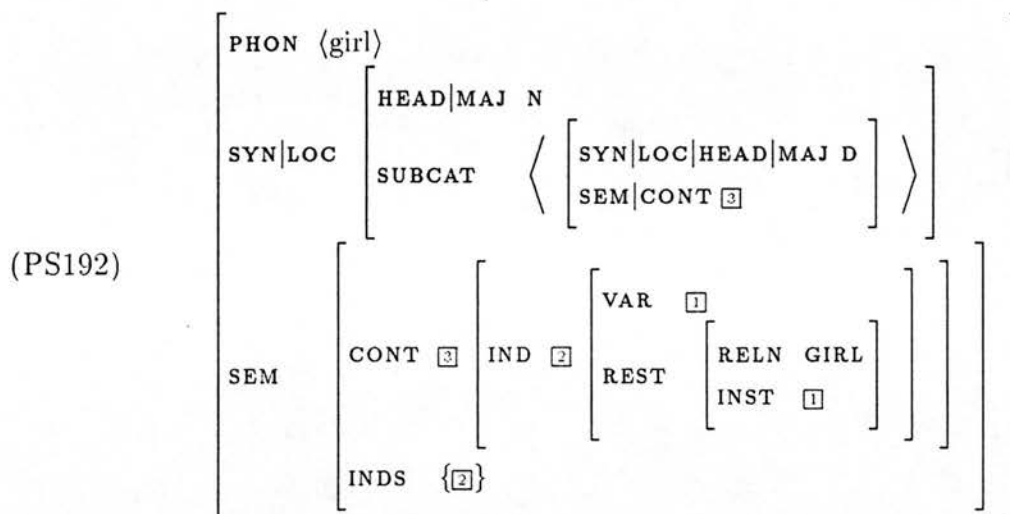
- $s[\text{COMP } -]$ a noncomplementised clause
- $s[\text{COMP } +]$ a complementised clause
- $SP[\text{COMP } -]$ a noncomplementised clause with a filled [Spec, CP] position
- $SP[\text{COMP } +]$ a complementised clause with a filled [Spec, CP] position

Complementised clause have a complementiser which takes the clause as an argument, e.g., *He said that this is a fine paper* or *I asked for him to arrive early*. This has been called *S'* in the past. A complementised clause with a filled [Spec, CP] position appears in Bavarian German and other languages. Typically the Spec position is filled with a *wh*-element, a topic constituent or an expletive.

3-6 The hpsg Treatment of Quantifier Scope

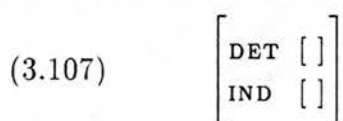
3-6.1 Pollard and Sag (1987)

In P&S1, head nouns subcategorise for determiners. The sign of *girl* is given in (PS192).

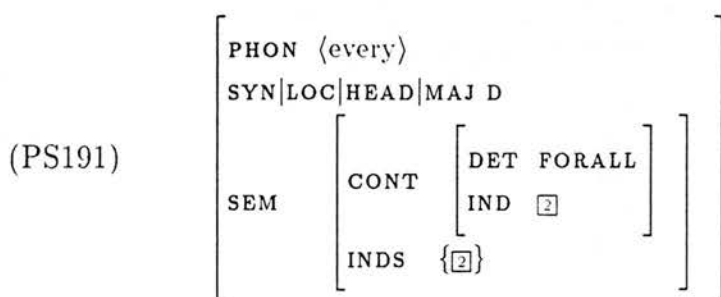


Notice that the **CONTENTS** of the subcategorised determiner and the **CONTENTS** of the noun are coindexed. That is, the **CONTENTS** of the noun and the determiner are unified.

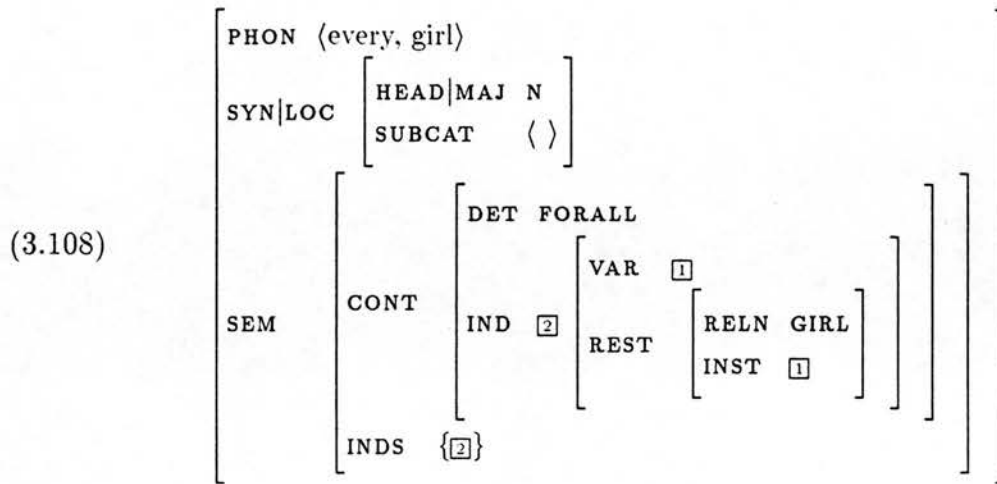
The form of the **CONTENTS** of a determiner (and a noun which subcategorises for a determiner) is of type *quantifier* and has attributes **DETERMINER** (**DET**) and **INDEX** (**IND**) as in (3.107). **DET** indicates what type of determiner (or quantifier) the determiner is and **IND** is the “restricted index” quantified over by the determiner.



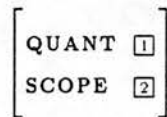
The sign of *every* is given in (PS191). The **CONTENTS** value of *every* is of type *quantifier* and has attributes **DET** and **IND** with values **FORALL** and $\boxed{2}$ respectively.



Then the sign for *every girl* is (3.108). Notice that the **CONTENTS** of the NP is just the unification of the **CONTENTS** of the sign for *every* and the sign for *girl*.



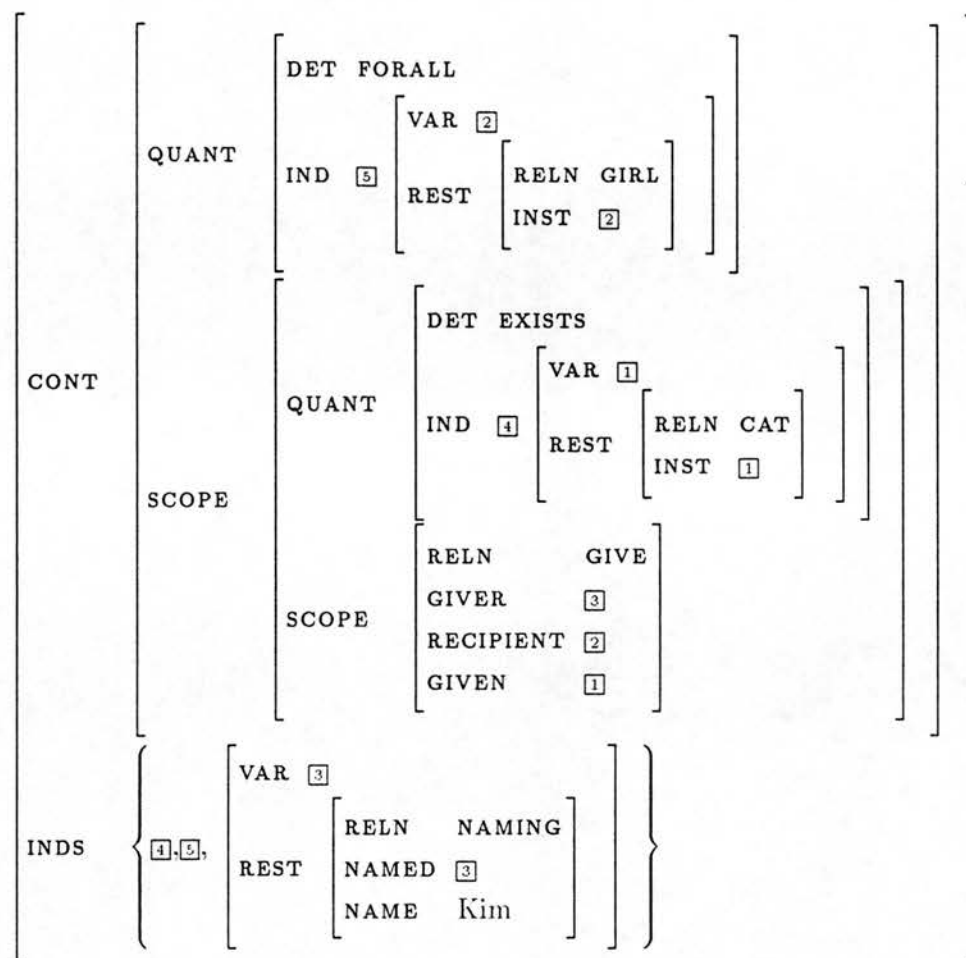
Since *every girl* is one of the **DTRS** in a *head-complement-structure* its semantics will be combined with the semantics of the head that subcategorised for it by the *Subcategorisation Principle*. All NPs whose semantics are of type *quantifier* will produce a structure of the form



where $\boxed{1}$ is the **CONTENTS** value of the quantified NP and $\boxed{2}$ is the **CONTENTS** value of the sign that subcategorises for the quantified NP by the definition of *combine-semantics*.

Therefore, the sign for *Kim gives every girl a cat* is (3.109).

(3.109)



If the order of argument daughters is determined by grammatical function hierarchy from most oblique to least oblique (as in P&S1) then the *Semantics Principle* produces only the “natural scope” reading where less oblique quantified phrases take wide scope over more oblique quantified noun phrases. However, any other scope readings are unattainable with the definition of the *Semantics Principle* defined as it is. P&S reconstruct the treatment of quantifiers in P&S2 in terms of a form of “Cooper Storage”. We will turn to this reconstruction next.

3–6.2 Pollard and Sag (1992)

In P&S2, the treatment of quantifier scope is modified completely. First, the structure of signs is redefined. Instead of every sign being defined for **PHONOLOGY**, **SYNTAX** and **SEMANTICS**, they are defined for the attributes **PHONOLOGY**, **SYNSEM** (which contains all syntactic and semantic information) and **QUANTIFIER-STORE** (**QSTORE**) which is used for the treatment of quantifier scope. **QSTORE** takes (a possibly empty) set of *quantifiers* as its value. I will briefly summarize the treatment of quantified NPs and quantifier scope here. The key change of the new

account is that all quantifiers start out in storage and that the final scope that a quantifier receives depends on which node it is *retrieved* at and on the order of its retrieval relative to the retrieval of other quantifiers in storage. This is defined informally in terms of the *Quantifier Inheritance Principle*.

(3.110) *Quantifier Inheritance Principle (QIP), informal version*

The **QUANTIFIER-STORE (QSTORE)** value of a phrasal node is the union of the **QSTORE** value of the daughters less those quantifiers that are retrieved at that node.

Instead of appearing as the **CONTENT** value, the quantifiers appear in the **QSTORE** value. Only the restricted parameter of a quantifier will appear in the **CONTENT** value. The sign for the phrase *every book* is shown in (3.111).

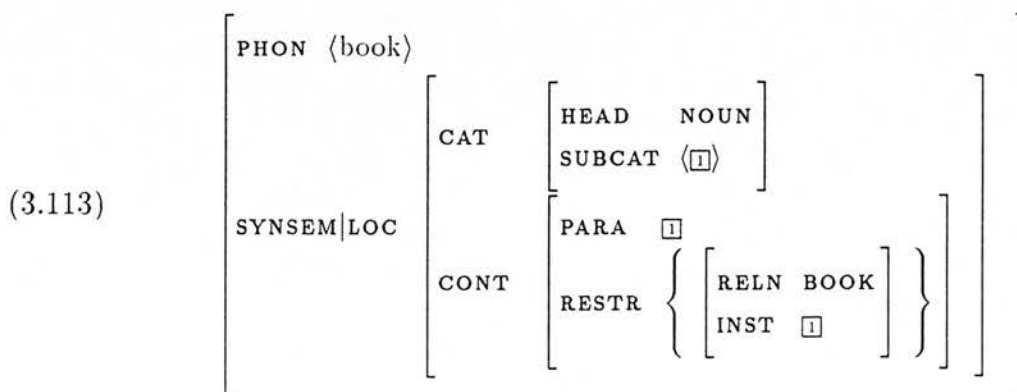


Then a quantified NP's **CONTENT** value has the same sort as a nonquantified, i.e., *referential-object*. In fact, the **CONTENT** value of *every book* is the same as the head noun's *book*. This makes it possible to modify the *Semantics Principle* as follows.

(3.112) *Semantics Principle (preliminary version)*

The **CONTENT** value of a phrase is unified with that of the head daughter.

Then the sign for *book* is (3.113)



where \boxed{n} is

$$(3.114) \quad \left[\text{LOC|CAT} \left[\begin{array}{cc} \text{HEAD} & \text{DET} \\ \text{SUBCAT} & \langle \rangle \end{array} \right] \right]$$

3-7 The Syntax and Semantics of Quantifiers

The treatment of the semantics of quantifiers here builds on the treatment of semantics outlined in the previous sections. The sign for the determiner a is the following.

$$(3.115)_{\text{NP}} \quad \left[\begin{array}{l} \text{PHON} \langle a \rangle \\ \text{SYN|LOC|ARGS} \left\langle \text{N} \left[\begin{array}{c} \text{SYN|LOC|ARGS} \langle \rangle \\ \text{SEM|CONT} \boxed{2} \end{array} \right] \right\rangle \\ \text{SEM|CONT} \left[\begin{array}{l} \text{INDEX} \boxed{4} \\ \text{RELN} \text{ EXISTS} \\ \text{QIND} \boxed{1} \\ \text{PROP} \left[\begin{array}{l} \text{INDEX} \boxed{1} \\ \text{RELN} \text{ AND} \\ \text{ARG1} \boxed{2}[\text{INDEX} \boxed{1}] \\ \text{ARG2} [\text{INDEX} \boxed{3}] \end{array} \right] \end{array} \right] \end{array} \right]$$

The outermost RELATION is EXISTS and the innermost RELATION is AND. This corresponds to the first order translation $\exists x.(P(x) \wedge Q(x))$.

There are several things to notice about this sign. First of all, it is a functor from nouns into NPs. This is consistent with categorial grammar treatments and with the traditional \bar{X} view of determiners as specifiers. That is, nouns (even nouns with all of their complements saturated and with possible adjuncts) are of type \bar{N} , i.e., not phrasal (not NPs) and not maximal. Actually the sign should indicate that the argument noun is nonmaximal ([MAX -]) since saturated nouns are taken to be NPs in HPSG. However, this route is not available to us, since heads do not subcategorise for their specifiers in this treatment. (Presumably, MAX would be a local nonhead feature. Singular common nouns would be specified as [MAX -] while plural common nouns and mass nouns would be unspecified for MAX. NPs would be specified as N[MAX +, SUBCAT ⟨ ⟩]. However, we will ignore such issues in the sequel.)

The value of the path SEM|CONT needs some explanation. It is a feature structure containing the attributes INDEX, RELATION (RELN), QUANTIFIED-INDEX (QIND)

and **PROPOSITION (PROP)**. The **INDEX** is the index of the entire semantics of the quantified NP. In general, it will be coinstantiated with the head daughter. This is so that the right ontological index can be assigned at the sentential level for example when the NP is the complement of the head verb. The **RELATION** is the actual quantifier relation, for example, **EXISTS**, **FORALL**, **MANY**, **MOST**, etc. This is so because we take a roughly generalised quantifier approach to quantification. The **QUANTIFIED-INDEX** is actually the index which is quantified over by the quantifier. It will always be the index of the argument noun. The **PROPOSITION** is the “formula” which is quantified over. This attribute is perhaps misnamed but allows some uniformity in signs.

The **PROPOSITION** is itself a complex feature structure with four features: **INDEX**, **RELATION**, **ARGUMENT1 (ARG1)** and **ARGUMENT2 (ARG2)**. The **INDEX** is the index of the quantified noun (which is coindexed with **QIND**). The **RELATION** is the actual relation introduced by the natural language quantifier (in this case **AND**). **ARGUMENT1** is the semantics of the argument noun (whose index is coindexed with the **INDEX** and with **QIND**). **ARGUMENT2** is the semantics of the phrase with which the quantified NP will combine as a complement. Notice that it’s index is coindexed with the index of the entire sign.

The sign for the determiner *every* is similar:

$$(3.11(x)P) \left[\begin{array}{l} \text{PHON } \langle \text{every} \rangle \\ \text{SYN|LOC|ARGS } \left\langle \begin{array}{l} \text{N } \left[\begin{array}{l} \text{SYN|LOC|ARGS } \langle \rangle \\ \text{SEM|CONT } \boxed{2} \end{array} \right] \right\rangle \\ \text{SEM|CONT } \left[\begin{array}{l} \text{INDEX } \boxed{4} \\ \text{RELN } \text{FORALL} \\ \text{QIND } \boxed{1} \\ \text{PROP } \left[\begin{array}{l} \text{INDEX } \boxed{1} \\ \text{RELN } \text{IMPLIES} \\ \text{ARG1 } \boxed{2}[\text{INDEX } \boxed{1}] \\ \text{ARG2 } [\text{INDEX } \boxed{4}] \end{array} \right] \end{array} \right. \end{array} \right. \end{array} \right]$$

The outermost **RELATION** is **FORALL** and the innermost **RELATION** is **IMPLIES**. This corresponds to the first order translation $\forall x.(P(x) \rightarrow Q(x))$.

The sign for the determiner *the* is similar:

$$(3.117)_{\text{NP}} \left[\begin{array}{l} \text{PHON } \langle \text{the} \rangle \\ \text{SYN|LOC|ARGS } \left\langle \begin{array}{l} \text{N } \left[\begin{array}{l} \text{SYN|LOC|ARGS } \langle \rangle \\ \text{SEM|CONT } \boxed{2} \end{array} \right] \end{array} \right\rangle \\ \text{SEM|CONT } \left[\begin{array}{l} \text{INDEX } \boxed{4} \\ \text{RELN } \text{ DEFINITE} \\ \text{QIND } \boxed{1} \\ \text{PROP } \left[\begin{array}{l} \text{INDEX } \boxed{1} \\ \text{RELN } \text{ AND} \\ \text{ARG1 } \boxed{2}[\text{INDEX } \boxed{1}] \\ \text{ARG2 } [\text{INDEX } \boxed{4}] \end{array} \right] \end{array} \right] \end{array} \right]$$

The outermost **RELATION** is **DEFINITE** and the innermost **RELATION** is **AND**. This corresponds to the first order translation $\exists x.(P(x) \wedge Q(x))$.

We also treat proper names as quantified NPs. The sign for *Kim* is similar to the sign for *the*.

$$(3.118) \left[\begin{array}{l} \text{PHON } \langle \text{Kim} \rangle \\ \text{SYN } \text{ NP} \\ \text{SEM|CONT } \left[\begin{array}{l} \text{INDEX } \boxed{4} \\ \text{RELN } \text{ DEFINITE} \\ \text{QIND } \boxed{1} \\ \text{PROP } \left[\begin{array}{l} \text{INDEX } \boxed{4} \\ \text{RELN } \text{ AND} \\ \text{ARG1 } \left[\begin{array}{l} \text{INDEX } \boxed{1} \\ \text{RELN } \text{ NAME} \\ \text{NAME } \text{ Kim} \end{array} \right] \\ \text{ARG2 } [\text{INDEX } \boxed{4}] \end{array} \right] \end{array} \right] \end{array} \right]$$

There are a couple of points to note about proper name signs. First, the outermost **RELATION** is **DEFINITE**. This indicates that the semantics is unscoped. I.e., it acts as a logical constant. This means that the semantics of a proper name can be combined with the semantics of other quantified NPs in any order without affecting the reading of the sentence. Second, the **RELN** of **ARG1** is **NAME** which indicates the naming relation. Its **NAME** role is the actual proper name constant. Otherwise, the semantics is like the semantics of the definite article which it must be in order for the semantics of the sign to be able to combine with the semantics of other signs like any other NP.

3-8 Scope of Quantification

Although the value of the `SEMANTICS` feature is a complex feature structure, the approach taken to quantifiers here will be much more like that of first order logic except that quantifier structures can be *unscoped*. The infix binary relation \otimes is used to combine two quantifier structures without specifying their relative scope. \otimes is associative and commutative. For example, given the feature structures $\boxed{1}$ and $\boxed{2}$, $\boxed{1} \otimes \boxed{2}$ is equivalent to (3.119).

$$(3.119) \quad \left[\begin{array}{l} \text{INDEX } \boxed{4} \\ \text{RELN } [] \\ \text{QIND } \boxed{3} \\ \text{PROP } \left[\begin{array}{l} \text{INDEX } \boxed{3} \\ \text{RELN } [] \\ \text{ARG1 } [\text{INDEX } \boxed{3}] \\ \text{ARG2 } \boxed{2}[\text{INDEX } \boxed{4}] \end{array} \right] \end{array} \right] \vee \boxed{2} \left[\begin{array}{l} \text{INDEX } \boxed{4} \\ \text{RELN } [] \\ \text{QIND } \boxed{3} \\ \text{PROP } \left[\begin{array}{l} \text{INDEX } \boxed{3} \\ \text{RELN } [] \\ \text{ARG1 } [\text{INDEX } \boxed{3}] \\ \text{ARG2 } \boxed{1}[\text{INDEX } \boxed{4}] \end{array} \right] \end{array} \right]$$

\otimes implicitly encodes possible quantifier scopes without the use of Cooper storage or similar mechanisms (as discussed in [32, Ch. 7]). \otimes is defined in terms of the following relational dependencies.

$$(3.120) \quad \boxed{1} \otimes \boxed{2} \Leftrightarrow \begin{array}{l} \text{sequence}(\boxed{1} \otimes \boxed{2}, \boxed{3}) \wedge \\ \text{permutations}(\boxed{3}, \boxed{4}) \wedge \\ \text{disjunction-of-permutations}(\boxed{4}) \end{array}$$

$$(3.121) \quad \begin{array}{l} \text{disjunction-of-permutations}(\langle \boxed{1} \rangle) \Leftrightarrow \\ \text{combine-quantifiers}(\boxed{1}) \\ \text{disjunction-of-permutations}(\langle \boxed{1} \rangle \circ \boxed{2}) \Leftrightarrow \\ \text{combine-quantifiers}(\boxed{1}) \vee \text{disjunction-of-permutations}(\boxed{2}) \end{array}$$

$$(3.122) \quad \begin{array}{l} \text{combine-quantifiers}(\langle \boxed{1} \rangle) \Leftrightarrow \boxed{1} \\ \text{combine-quantifiers}(\boxed{1} \circ \langle \boxed{2}, \boxed{3} \rangle) \Leftrightarrow \\ \text{combine-quantifiers}(\boxed{1} \circ \langle \text{combine-two-quantifiers}(\boxed{2}, \boxed{3}) \rangle) \end{array}$$

$$(3.123) \quad \text{combine-two-quantifiers}(\boxed{1}, \boxed{2}) \Leftrightarrow \boxed{1} \left[\begin{array}{l} \text{INDEX } \boxed{4} \\ \text{RELN } [] \\ \text{QIND } \boxed{3} \\ \text{PROP } \left[\begin{array}{l} \text{INDEX } \boxed{3} \\ \text{RELN } [] \\ \text{ARG1 } [\text{INDEX } \boxed{3}] \\ \text{ARG2 } \boxed{2}[\text{INDEX } \boxed{4}] \end{array} \right] \end{array} \right]$$

sequence takes a \otimes term of arbitrary associativity as first argument and produces a sequence of the \otimes operands as its second argument. **permutations** takes an input sequence as its first argument and returns a sequence containing all permutations of the sequence as its second argument. **disjunction-of-permutations** is a function which “returns” a disjunction of all of the possible quantifier scopings implicit in $\boxed{1} \otimes \boxed{2}$. Thus, $\boxed{1} \otimes \boxed{2}$ will be equivalent to a disjunction of all the possible quantifier scopings implicit in $\boxed{1} \otimes \boxed{2}$. **combine-quantifiers** combines all the quantifiers in a sequence from last to first. Therefore, the last quantifier will have narrowest scope and the first will have widest scope. **combine-two-quantifiers** combines its two arguments such that the first argument has immediate wide scope over the second quantifier.

The *Semantics Principle* is stated as in (3.124). It is the same as the original P&S1 *Semantics Principle* except it is simplified by eliminating the collection of indexes. This is because we make no use of the feature **INDEXES** in this thesis.

$$(3.124) \left[\begin{array}{l} \text{SEM|CONT successively-combine-semantic}(\boxed{1}, \boxed{2}) \\ \text{DTRS} \left[\begin{array}{l} \text{FUN|SEM|CONT } \boxed{1} \\ \text{ARG-DTRS } \boxed{2} \end{array} \right] \end{array} \right]$$

The definition of the functional dependency **successively-combine-semantic** is similar to that in P&S. It is stated in (3.125)-(3.127).

$$(3.125) \text{successively-combine-semantic}(\boxed{1}, \langle \rangle) \Leftrightarrow \boxed{1}$$

$$(3.126) \text{successively-combine-semantic}(\boxed{1}, \langle [\text{SEM|CONT } \textit{quantifier} \boxed{2}] \rangle \circ \boxed{3}) \Leftrightarrow \text{successively-combine-semantic}(\boxed{1} \otimes \boxed{2}, \boxed{3})$$

$$(3.127) \text{successively-combine-semantic}(\boxed{1}, \langle [\text{SEM|CONT } \textit{circumstance} \boxed{1}] \rangle \circ \boxed{2}) \Leftrightarrow \text{successively-combine-semantic}(\boxed{1}, \boxed{2})$$

(3.125) states that the semantics of a functor with an empty argument list is simply the head semantics.

(3.126) states that the semantics of a functor with a nonempty argument list whose first element is of sort *quantifier* is the value of **successively-combine-semantic** applied to the functor’s semantics \otimes the quantifier semantics and the rest of the list. In other words, the functor’s semantics and the quantifier’s semantics are combined (with scope unspecified) using \otimes to become the new functor semantics which is then combined with the semantics of the rest of the list.

(3.127) states that the semantics of a functor with a nonempty argument list whose first element is of sort *circumstance* is simply the value of **successively-combine-semantic** applied to the functor’s semantics and the rest of the list.

That is, in this case, the functor takes care of semantic compositionality through coindexation in the argument list.

$quantifier []$ and $circumstance []$ are defined as follows.

$$(3.128)_{quantifier []} \Leftrightarrow \boxed{1} \left[\begin{array}{l} \text{INDEX } \boxed{3} \\ \text{RELN } [] \\ \text{QIND } \boxed{2} \\ \text{PROP } \left[\begin{array}{l} \text{INDEX } \boxed{2} \\ \text{RELN } [] \\ \text{ARG1 } [\text{INDEX } \boxed{2}] \\ \text{ARG2 } [\text{INDEX } \boxed{3}] \end{array} \right] \end{array} \right]$$

$$(3.129)_{circumstance []} \Leftrightarrow \left[\begin{array}{l} \text{INDEX } [] \\ \text{RELN } [] \end{array} \right] \wedge \neg[\text{QIND } []]$$

Then the semantics for *Kim gives every girl a cat* is the quantifier unscoped semantic representation (3.130)

$$(3.130) \boxed{7} \otimes \boxed{6} \otimes \boxed{7} \otimes \boxed{8}$$

where (3.131), (3.132) and (3.133) are the signs for *a cat*, *every girl* and *Kim* respectively.

$$(3.131) \boxed{7} \left[\begin{array}{l} \text{INDEX } \boxed{4} \\ \text{RELN } \text{ EXISTS} \\ \text{QIND } \boxed{1} \\ \text{PROP } \left[\begin{array}{l} \text{INDEX } \boxed{1} \\ \text{RELN } \text{ AND} \\ \text{ARG1 } \left[\begin{array}{l} \text{INDEX } \boxed{1} \\ \text{RELN } \text{ CAT} \end{array} \right] \\ \text{ARG2 } [] \end{array} \right] \end{array} \right]$$

$$(3.132) \left[\begin{array}{l} \text{INDEX } \boxed{4} \\ \text{RELN } \text{FORALL} \\ \text{QIND } \boxed{2} \\ \text{PROP } \left[\begin{array}{l} \text{INDEX } \boxed{2} \\ \text{RELN } \text{IMPLIES} \\ \text{ARG1 } \left[\begin{array}{l} \text{INDEX } \boxed{2} \\ \text{RELN } \text{GIRL} \end{array} \right] \\ \text{ARG2 } [] \end{array} \right] \end{array} \right]$$

$$(3.133) \left[\begin{array}{l} \text{INDEX } \boxed{4} \\ \text{RELN } \text{DEFINITE} \\ \text{QIND } \boxed{3} \\ \text{PROP } \left[\begin{array}{l} \text{INDEX } \boxed{4} \\ \text{RELN } \text{AND} \\ \text{ARG1 } \left[\begin{array}{l} \text{INDEX } \boxed{3} \\ \text{RELN } \text{NAME} \\ \text{NAME } \text{Kim} \end{array} \right] \\ \text{ARG2 } [\text{INDEX } \boxed{4}] \end{array} \right] \end{array} \right]$$

$$(3.134) \left[\begin{array}{l} \text{INDEX } \boxed{4} \\ \text{RELN } \text{GIVE} \\ \text{GIVER } \boxed{3} \\ \text{RECIPIENT } \boxed{2} \\ \text{GIVEN } \boxed{1} \end{array} \right]$$

The sign for *Kim gives every girl a cookie* is then

$$(3.135) \left[\begin{array}{l} \text{PHON } \langle \text{Kim, gives, every, girl, a, cookie} \rangle \\ \text{SYN|LOC|HEAD } \text{S} \\ \text{SEM|CONT } \boxed{5} \circ \boxed{6} \circ \boxed{7} \circ \boxed{8} \end{array} \right]$$

So far, we have said nothing about how a particular scope gets forced. We introduce the relation \rightarrow to indicate that one SEM|CONT value takes *immediate*

wide scope over another. \leftarrow is defined in (3.136).

$$(3.136) \quad \boxed{2} \leftarrow \boxed{1} \Leftrightarrow \text{true} \quad \left(\begin{array}{c} \boxed{1} \left[\begin{array}{l} \text{INDEX } \boxed{4} \\ \text{RELN } [] \\ \text{QIND } \boxed{3} \\ \text{PROP } \left[\begin{array}{l} \text{INDEX } \boxed{3} \\ \text{RELN } [] \\ \text{ARG1 } [\text{INDEX } \boxed{3}] \\ \text{ARG2 } \boxed{2}[\text{INDEX } \boxed{4}] \end{array} \right] \end{array} \right] \end{array} \right)$$

The nonce relation **true** is used to ensure that its argument formula is true *somewhere* in the model.¹² **true** is axiomatised by the single axiom $\text{true}(\top)$.

We can also define a less restrictive scope relation \leftarrow in terms of \leftarrow which indicates that one **SEM|CONT** takes wide scope over another but not necessarily immediate scope as in (3.138).

$$(3.137) \quad \boxed{2} \leftarrow \boxed{1} \Rightarrow \boxed{2} \leftarrow \boxed{1}$$

$$(3.138) \quad \boxed{3} \leftarrow \boxed{2} \leftarrow \boxed{1} \Rightarrow \boxed{3} \leftarrow \boxed{1}$$

Then for example if we believe that the scope of quantified NPs in German is determined by “natural scope”, i.e., by left to right order we could state a *Quantifier Scope Principle* (3.139).

$$(3.139) \quad \text{DOM NP}[\text{SEM|CONT}\boxed{1}] \prec \text{NP}[\text{SEM|CONT}\boxed{2}] \Rightarrow \boxed{2} \leftarrow \boxed{1}$$

Furthermore, we could state the more restrictive principle (3.140).

$$(3.140) \quad \text{DOM NP}[\text{SEM|CONT}\boxed{1}] \prec \text{NP}[\text{SEM|CONT}\boxed{2}] \Rightarrow \boxed{2} \leftarrow \boxed{1}$$

Similar principles can be stated for adverbials whose scope is determined by linear order in the *Mittelfeld* and is not inherent in the semantics of the adverbials. Cf. [24] and [2].

This concludes the technical treatment of *head-complement*, *head-adjunct* and *head-specifier* structures. We have not completed the goal of eliminating the

¹²It's effect is similar to the **L** operator of modal logic. $\mathbf{L}\phi$ is true if there exists some world w in the model \mathcal{M} such that $\mathcal{M} \models_w \phi$. However **L** is not deriveable from the rest of the logic although it can always be simulated with nonce relations like **true**.

Semantics Principle entirely however. Therefore, I will make one or two speculative remarks at this point.

A technical exposition of the solution will show that the question of whether or not to eliminate the *Semantics Principle* is essentially a philosophical one. Therefore, I will address the technical issues first. The technical solution depends on making the semantics of every functor which takes NP arguments a function. To illustrate the idea I repeat the sign for *kick* presented earlier in this chapter but in modified form.

$$(3.141) \left[\begin{array}{l} \text{PHON} \quad \text{kicks} \\ \text{SYN|LOC|ARGS} \quad \langle \text{NP[NOM]: [1], NP[ACC]: [2]} \rangle \\ \text{SEM|CONT} \quad \left[\text{RELN} \quad \text{kick}([0].[1].[2]) \right] \end{array} \right]$$

That is, we now have a standard relational translation of the verb *kick*. However, it is nonstandard in that it takes quantified NP structures as arguments. This is taken care of by making the predicate ‘kick’ a function defined as follows.

$$(3.142) \quad \text{kick}([0].[1][\text{PROP|ARG1|INDEX } [5]], [2][\text{PROP|ARG1|INDEX } [6]]) \Leftrightarrow$$

$$[1] \otimes [2] \otimes \left[\begin{array}{l} \text{INDEX} \quad [4] \\ \text{RELN} \quad \text{KICK} \\ \text{AGENT} \quad [5] \\ \text{PATIENT} \quad [6] \end{array} \right]$$

Thus a saturated verb projection will have a semantics of (3.142) just as if it were created by the *Semantics Principle*. (This assumes that all argument requirements get saturated, either syntactically, functionally or anaphorically, an assumption which I presume is uncontroversial.) Modulo possible objections about different scope of adjuncts, the results for saturated functor projections is the same. Any functor which takes quantified NPs as arguments can have its semantics translated like this. As things stand, a functional dependency would be required for each lexical entry since the thematic roles differ from predicate to predicate. One step towards easing this inventory of functional definitions is to eliminate specific roles like **AGENT** and **PATIENT** and replace them with a single list of argument roles as in (3.143).

$$(3.143) \quad \text{kick}([1][\text{PROP|ARG1|INDEX } [5]], [2][\text{PROP|ARG1|INDEX } [6]]) \Leftrightarrow$$

$$[1] \otimes [2] \otimes \left[\begin{array}{l} \text{INDEX} \quad [4] \\ \text{RELN} \quad \text{KICK} \\ \text{ARGS} \quad \langle [5].[6] \rangle \end{array} \right]$$

This makes all the definitions similar but still doesn't solve the problem of assigning one general schema to define all of them. This problem is solvable with

suitable recursive definitions but we can make a simpler observation. The following sign is all we need for *kicks*

$$(3.144) \left[\begin{array}{l} \text{PHON } \textit{kicks} \\ \text{SYN|LOC|ARGS} \left\langle \left[\begin{array}{l} \text{SYN NP[NOM]} \\ \text{SEM|CONT } \boxed{1} \end{array} \right], \left[\begin{array}{l} \text{SYN NP[ACC]} \\ \text{SEM|CONT } \boxed{2} \end{array} \right] \right\rangle \\ \text{SEM|CONT } \boxed{1} \otimes \boxed{2} \otimes \left[\begin{array}{l} \text{INDEX } \boxed{4} \\ \text{RELN } \textit{KICK} \\ \text{AGENT } \boxed{5} \\ \text{PATIENT } \boxed{6} \end{array} \right] \end{array} \right]$$

where $\boxed{1}$ and $\boxed{2}$ are as follows.

$$(3.145) \boxed{1} \left[\text{PROP|ARG1|INDEX } \boxed{5} \right]$$

$$(3.146) \boxed{2} \left[\text{PROP|ARG1|INDEX } \boxed{6} \right]$$

The philosophical problem which arises is whether the lexical semantics of a verb contains the (very underspecified) semantics of each of its NP arguments. If there are no philosophical objections then the *Semantics Principle* can be eliminated, otherwise it must be retained in its current form.

3–9 Scope of Modification

Kasper ([24]) points out what he sees as potential problems for my approach with respect to multiple adverbial VP modifiers. He lists three types of multiple adverbial modifier configurations which bear on the issue of the scope of multiple adverbials. They are:

1. scope of modification unchanged by permutation;
2. scope of modification changed by permutation and
3. ambiguous scope of modification.

I'll briefly review his data and propose solutions to the potential problems Kasper raises.

3-9.1 Scope of modification: unchanged by permutation

The scope of modification of adverbials can be unchanged by permutation of the adverbials either because there is either only one coherent order of modification as in (Kas6) (where the order of composition is shown in (Kas7a)) or else because the adverbials are not inherently semantic operators as in (Kas8).

(Kas6) a Peter hat täglich 3 Stunden lang trainiert
 Peter has daily 3 hours long trained
 'Peter trained 3 hours a day'

b Peter hat 3 Stunden täglich lang trainiert

(Kas7) a daily-frequency(3-hour-duration(Peter-trained))

b *3-hour-duration(daily-frequency(Peter-trained))

In (Kas6), *täglich* and *3 Stunden* can be permuted without affecting the quantifier scope. This is because the only coherent scope is where *täglich* takes wide scope over *3 Stunden*. The opposite case where *3 Stunden* takes wide scope over *täglich* is semantically incoherent. The null hypothesis running through the treatment of word order is that both orders are possible in this case precisely because there is no need to indicate scope of modification by word order.

(Kas8) a Peter hat gestern im Park Tennis gespielt
 Peter has yesterday in the park tennis played
 'Peter played tennis in the park yesterday'

b Peter hat im Park gestern Tennis gespielt

In (Kas8), the modifier *gestern* introduces a temporal location predicate which modifies the index of *gespielt* while *im Park* introduces a spatial location predicate which also modifies the index of *gespielt*. In our terms they are not semantic operators and so are conjunctive in the InL translations. This means that they are commutative in the corresponding InL translations. Thus neither expression can really be said to be in the scope of the other. Again, the fact that scope of modification does not need to be indicated by word order leads us to expect that either order is possible (although they may be subject to weak preferential word order constraints).

This means that there is spurious ambiguity in terms of the derivation of *im Park gestern Tennis gespielt*. That is, both derivations in (3.147) are wellformed and the semantic translation is the same (modulo commutativity).

(3.147) a [VP [VP/VP im Park] [VP [VP/VP gestern] [VP Tennis gespielt]]]

b [VP [VP/VP gestern] [VP [VP/VP im Park] [VP Tennis gespielt]]]

Thus, (Kas8) presents no problem for our account aside from the spurious ambiguity. I do not see the spurious ambiguity as a theoretical problem but rather as a parsing problem (but not a *human* parsing problem). I prefer to treat the spurious ambiguity using techniques developed within the categorial grammar community for dealing with it. It must be said however that this is arguably a strong argument for a configurational structure where all adverbials are sisters to their head, as proposed by Kasper ([24]).

The two adverbials may then be subject to (weak) ordering constraints such as the wellknown general preference for temporal adverbials to precede locative adverbials (in contrast to English).

- (Kas9) a Peter liest gut wegen der Nachhilfestunden
 Peter reads well because of the tutoring
 'Peter reads well because of the tutoring'
- b Peter liest wegen der Nachhilfestunden gut

- (Kas10) a because-of-tutoring(good-manner(Peter-reads))
 b *good-manner(because-of-tutoring(Peter-reads))

In this example, the reading (Kas10b) is semantically incoherent. In this case there is only one wellformed derivation but the adverbials may be freely ordered (modulo preferences). (Kas9) is similar to (Kas6).

3-9.2 Scope of modification: affected by permutation

The scope of modification of adverbials can also be determined by linear order. In these cases, the adverbials are semantic operators.

- (Kas12) a Peter kommt oft vergeblich
 Peter comes often in vain
 'It is often that Peter comes in vain'
- b Peter kommt vergeblich oft
 Peter comes in vain often
 'It is in vain that Peter comes often'

The two examples in (Kas12) follow the general preference that adverbial modifier scope is wide scope to the left and narrow scope to the right. Thus *oft* takes wide scope over *vergeblich* in (Kas12a) and *vergeblich* takes wide scope over *oft* in (Kas12b). In our terminology, both *oft* and *vergeblich* are semantic operators since they embed the semantics of their argument in their own semantics. I.e., in

(Kas12a), an informal indication of the semantics is (Kas14) while the semantics of (Kas12b) is (3.148).

(Kas14) `often(in-vain(Peter-comes))`

(3.148) `in-vain(often(Peter-comes))`

The problem for our account is to link the left-to-right scope reading of the linear order with the semantic operator compositionality. An implication of the form (3.149) is all that is required. The ordering implication is of exactly the same form as the ordering implication for scope of quantified NPs.

(3.149) $\text{dom VP/VP : } \boxed{1} \prec \text{VP/VP : } \boxed{2} \Rightarrow \boxed{2} \prec \boxed{1}$

- (Kas13) a Oft kommt Peter vergeblich
 often comes Peter in vain
 ‘It is often that Peter comes in vain’
- b vergeblich kommt Peter oft
 in vain comes Peter often
 ‘It is in vain that Peter comes often’

(Kas13) is an example which Kasper presents as problematic for our approach but for which he makes no comment. In both examples, one of the adverbials appears in topicalised position. At this point, it is worth recalling that quantified NPs in topic position invariably take wide scope over other NPs in the clause. Similarly, an adverbial modifier in topic position takes wide scope over other adverbial modifiers in the clause. This can be treated by requiring that the topic always takes wide scope (whether quantified NP or adverbial) over the rest of the clause.

3–9.3 Scope of modification: ambiguous

Finally, Kasper notes that there are a class of examples where the scope of modification is ambiguous and appears not to be influenced by linear order.

- (Kas16) Peter kommt wegen der Verabredung während der
 Peter comes because of the appointment during the
 Vorstellung
 performance
 ‘Peter comes because of the appointment during the
 performance’

- (Kas17) a 'Because of the appointment he is coming during the performance.'
- b 'During the performance he is coming because of the appointment.'

According to [2], there are two interpretations, shown in (Kas17), available where *wegen der Verabredung* and *während der Vorstellung* are VP modifiers. Linear order appears to play no part in determining scope. If these two readings are equally available, then they are not subject to the normal left-to-right scope preference constraint. In the absence of the applicability of this constraint, then the derivations corresponding to (Kas17a) and (Kas17b) are free to linearise the two adverbials in either order (under the assumption that no other LP constraints apply). Thus, this class of examples poses no problems for our approach either.

3-10 Raising

To implement raising in HPSG, we will use an approach which is similar to the treatment of unbounded dependencies in HPSG or GPSG. That is, there will be a "top", a "bottom" and a "middle". The bottom will move elements to be raised from argument lists to a sequence of the SYN|NONLOC feature RAISE which takes a sequence as its value. The middle will percolate RAISE values up the functor-argument structure. Finally, the top will take all of the raised elements in a clause and add them to the clause's domain.

Raising introduction is accomplished by the *raising trace* which has a null PHON and which has a SLASH value which is a singleton set containing the PHON, SEM and SYN|LOC features of the sign corresponding to the gap. The following trace sign is modelled on that in P&S2.

(3.150) Raising Trace

$$\left[\begin{array}{l} \text{PHON } \langle \rangle \\ \text{SYN } \left[\begin{array}{l} \text{LOC } \boxed{1} \\ \text{NONLOC|SLASH } \left[\begin{array}{l} \text{SYN|LOC } \boxed{1} \\ \text{SEM } \boxed{2} \end{array} \right] \end{array} \right] \\ \text{SEM } \boxed{2} \end{array} \right]$$

To percolate RAISE values up the functor-argument structure, we need a *Raising Inheritance Principle* which is defined in terms of the functional dependency **concatenate-raise** which is defined below. It concatenates the RAISE sequences

of a sequence of signs.

(3.151) concatenate-raise

$$\begin{aligned} \text{concatenate-raise}(\langle \rangle) &= \langle \rangle \\ \text{concatenate-raise}(\langle \langle \text{SYN|NONLOC|RAISE } \boxed{1} \rangle \circ \boxed{2} \rangle) &= \\ &\boxed{1} \circ \text{concatenate-raise}(\boxed{2}) \end{aligned}$$

The *Raising Inheritance Principle* only applies to non-clause phrasal signs. It concatenates the **RAISE** values of the **ARG-DTRS** and the **FUN-DTR** and makes the resulting sequence the value of the phrasal mother's **RAISE** value.

(3.152) *Raising Inheritance Principle*

$$\begin{aligned} &\text{phrasal} \left[\text{LOC|HEAD } \neg\text{S} \right] \Rightarrow \\ &\left[\begin{array}{l} \text{SYN|NONLOC|RAISE concatenate-raise}(\langle \boxed{1} \rangle \circ \boxed{2}) \\ \text{DTRS} \left[\begin{array}{l} \text{FUN-DTR } \boxed{1} \\ \text{ARG-DTRS } \boxed{2} \end{array} \right] \end{array} \right] \end{aligned}$$

According to the empirical facts, raised elements must all be discharged into their smallest governing clause. Since, the *Raising Inheritance Principle* does not inherit **RAISE** values onto clauses and since raised elements must all be part of a clause's domain, a special version of the *Domain Principle* is needed for clauses. It is defined in terms of the two relational dependencies, **permute** and **concatenate-raise**. **permute** is itself defined in terms of the relational dependency **delete**. These are both defined below. They should be immediately familiar to any Prolog programmer. $\boxed{1} \Leftrightarrow \text{permute}(\boxed{2})$ is true iff $\boxed{1}$ is a permutation of the sequence $\boxed{2}$. $\boxed{1} \Leftrightarrow \text{delete}(\boxed{2}, \boxed{3})$ is true iff $\boxed{1}$ is the list $\boxed{3}$ with the element $\boxed{2}$ removed.

(3.153) permute

$$\begin{aligned} \text{permute}(\langle \rangle) &\Leftrightarrow \langle \rangle \\ \text{permute}(\boxed{1}) &\Leftarrow \\ &\langle \boxed{2} \rangle \circ \text{permute}(\text{delete}(\boxed{2}, \boxed{1})) \end{aligned}$$

(3.154) delete

$$\begin{aligned} \text{delete}(\boxed{1}, \langle \boxed{1} \rangle \circ \boxed{2}) &\Leftrightarrow \boxed{2} \\ \text{delete}(\boxed{1}, \langle \boxed{2} \rangle \circ \boxed{3}) &\Leftarrow \\ &\langle \boxed{2} \rangle \circ \text{delete}(\boxed{1}, \boxed{3}) \end{aligned}$$

The *Domain Principle for Clauses* is the same as the *Domain Principle* except that some permutation of the concatenation of the **RAISE** values of the **FUN-DTR**

and **ARG-DTRS** is domain unioned into the clauses domain. Thus raised elements are raised to a clausal domain and not to the clause's head's argument list or to some filler gap position. This is crucial for getting the empirical facts right.

(3.155) *Domain Principle for Clauses*

$$\left[\begin{array}{l} \text{SYN|LOC|HEAD S} \\ \text{SYN|NONLOC|RAISE } \langle \rangle \\ \text{DTRS } \left[\begin{array}{l} \text{FUN-DTR } \boxed{1} \\ \text{ARG-DTRS } \boxed{2} \end{array} \right] \\ \text{DOM } \left(\langle \boxed{1} \rangle \circ \langle \boxed{3} \rangle \circ \dots \circ \langle \boxed{1} \rangle \circ \langle \boxed{i+1} \rangle \circ \dots \circ \langle \boxed{n} \rangle \right) \\ \text{permute}(\text{concatenate-raise}(\langle \boxed{1} \rangle \circ \langle \boxed{2} \rangle)) \end{array} \right] \Rightarrow$$

where $\boxed{2}$ is

$$(3.156) \quad \left(\langle \boxed{3}[\text{UNIONED } -], \dots, \boxed{2}[\text{UNIONED } -] \rangle \circ \langle [\text{UNIONED } +, \text{DOM } \boxed{i+1}], \dots, [\text{UNIONED } +, \text{DOM } \boxed{n}] \rangle \right)$$

The fact that **permute** and **concatenate-raise** are used in the definition is a direct consequence of treating **RAISE** values as sequences. This was done mainly as a matter of technical convenience. Similar definitions and principles could be used to implement a treatment where **RAISE** takes sets as its values at the expense of making the *Raising Introduction Rule* and the *Domain Principle for Clauses* slightly more complicated. It makes no difference which approach is taken.

Chapter 4

An Account of German Word Order

4-1 Introduction

We will now give an informal and somewhat idealised presentation of how this approach accounts for word order in German V2 (verb-second) and subordinate clauses. We will assume the 'TVX' analysis for V2 clauses, i.e., a topic followed by the finite verb followed by 'everything else'. The order of constituents in the 'X' domain is the same as the order of the post-complementiser field in subordinate clauses so we will primarily consider subordinate clauses. We assume, following [45], that the structure of a V2 clause is [CP XP [S/XP V[FIN] ...]], i.e., that the topic is filled by unbounded movement of a phrasal constituent from the post-verbal field. Furthermore, we assume that the finite verb is initial in the domain of the S in contrast to GB assumptions about the verb moving to COMP. In this structure there is no COMP position. The category of the V2 clause is labelled CP even though there is no COMP position present. CP is used because the topic does seem to be in a specifier-like position.

As in the discussion of (1.15), the characteristic Mittelfeld order of a sequence of NPs followed by a sequence of verbs is produced by unioning the domains of VPs and Ss together. We maintain the two previous LP constraints that NPs precede verbs and for canonical verb order, a governing verb is preceded by all the verbs it governs.

- (4.1)
- a. daß der Mann versuchte, das Buch zu lesen
that the man tried the book to read
'that the man tried to read the book'
 - b. daß der Mann versucht hat, das Buch zu lesen
that the man tried has, the book to read
'that the man has tried to read the book'

- (4.2) daß der Mann versucht hat, zu behaupten, das Buch gelesen zu
 that the man tried has, to claim, the book read to
 haben
 have
 ‘that the man has tried to claim to have read the book’

(4.1) and (4.2) contain *extraposed* VPs. An extraposed VP is one which occurs to the right of the verb cluster which contains the verb that governs it. In (4.1a), the VP *das Buch zu lesen* is not unioned but is extraposed and appears in clause-final position after the finite verb. To analyse this, we prohibit extraposed clauses from being unioned and add the LP constraint [EXTRA -] < [EXTRA +]. **EXTRA** (**EXTRAPOSED**) is a binary-valued feature which indicates whether a constituent is extraposed or not. This LP constraint forces all nonextraposed elements to proceed all extraposed elements within a domain. (In this fragment, only VPs are allowed to be extraposed.) Since the VP is extraposed it won't be unioned and will be marked [EXTRA +]. (Cf Figures (4-1) and (4-2).)

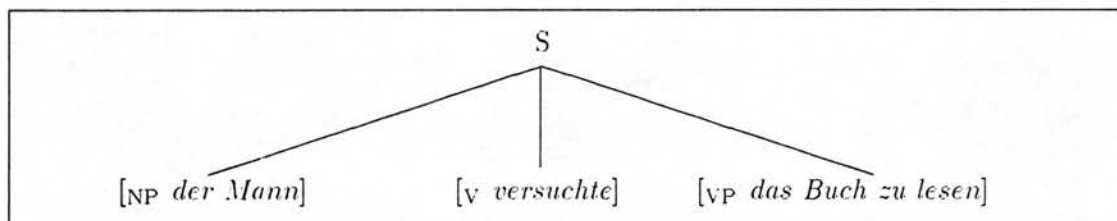


Figure 4-1: Syntax tree for (4.1a)

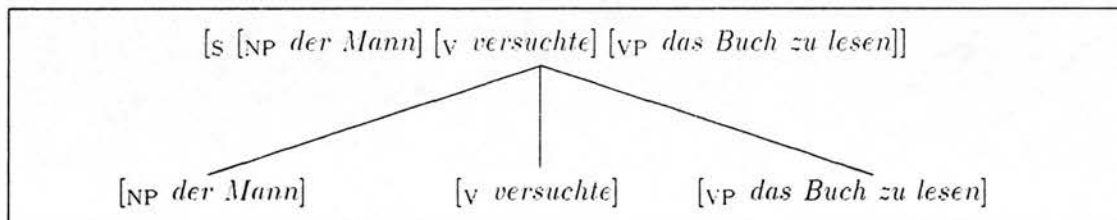


Figure 4-2: Domain tree for (4.1a)

(4.1b) is slightly more interesting. Here, *das Buch zu lesen* is subcategorised by the participle *versucht* from which it is separated and is not subcategorised by the finite auxiliary *hat*. The domain of the VP *das Buch zu lesen* is not unioned and so the VP is in domain-final position in the domain of the VP *versucht, das Buch zu lesen* due to the LP constraint. Its domain is (4.3).

- (4.3) [VP [v versucht] [VP das Buch zu lesen]]

(4.3) is then unioned into the finite clause domain resulting in domain (4.4).

(4.4) [VP [NP *der Mann*] [v *versucht*] [v *hat*] [VP *das Buch zu lesen*]]

Since [VP *das Buch zu lesen*] is [EXTRA +] and an element of the finite clause's domain, it must be domain-final, which it is. *versucht* appears to the left of *hat* since *hat* governs it. Finally, the NP *der Mann* appears to the left of all the verbs as required by the LP constraint. (Cf. Figures (4-3) and (4-4).)

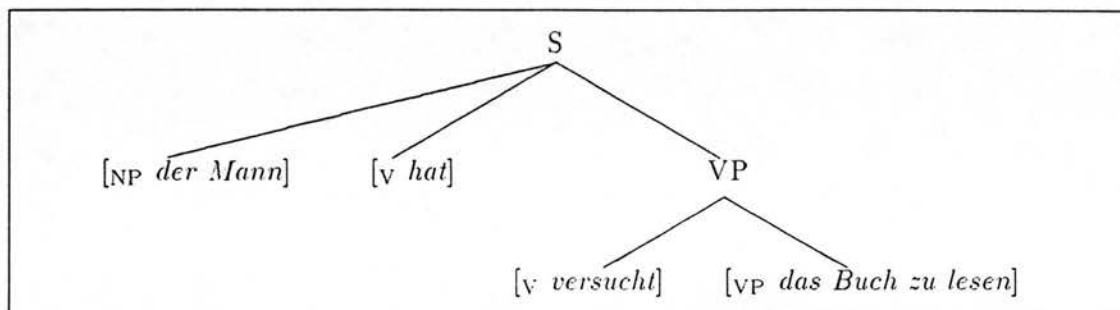


Figure 4-3: Syntax tree for (4.1b)

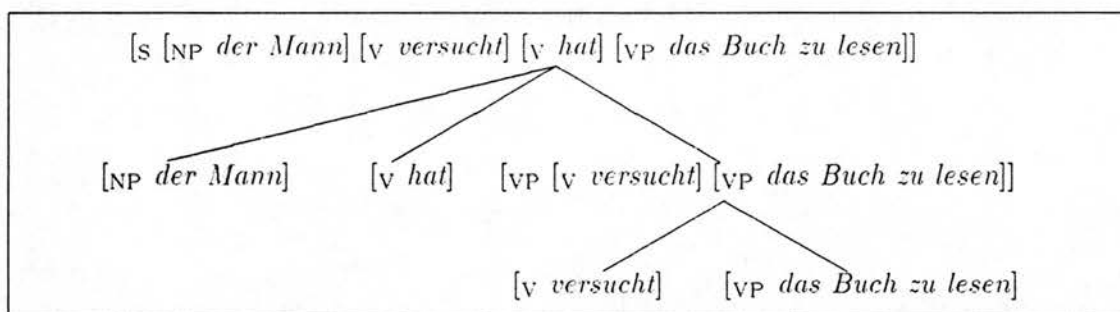


Figure 4-4: Domain tree for (4.1b)

(4.2) is a bit more complicated. The VP *das Buch gelesen* and the verb *zu haben* form the VP *das Buch gelesen zu haben*. It may not look like it but this actually involves unioning the domain of *das Buch gelesen* into the VP domain of which *zu haben* is the head. In general, we require that nonextraposed VPs are unioned. The extraposed VP *das Buch gelesen zu haben* forms a VP with the verb *zu behaupten*. Its domain is (4.5).

(4.5) [VP [v *zu behaupten*] [VP *das Buch gelesen zu haben*]]

Then this VP is extraposed in the VP *versucht, zu behaupten, das Buch gelesen zu haben*. Its domain is (4.6).

(4.6) [VP [v *versucht*] [VP [v *zu behaupten*] [VP *das Buch gelesen zu haben*]]]

Finally, this VP domain is unioned into the domain of the finite clause as in (4.1b). Cf. Figures (4-5) and (4-6).

Notice that VP extraposition is not taken to be clause-bounded rightward movement as in English relative clause extraposition. Rather VP extraposition is analyzed as a VP occurring in domain final position in a VP or S domain. In cases of "recursive" extraposition like (4.2), the VP *das Buch gelesen zu haben* is "trapped" inside of the domain of the VP *zu behaupten, das Buch gelesen zu haben*. Therefore when this VP is extraposed within the finite clause, we will get the characteristic recursive extraposition order of [... VC VP₁ VP₂ VP₃ ...] where VC is the verb cluster.

This treatment of VP extraposition has interesting implications for topicalisation of VPs in V2 clauses. Assume that constituents are not unioned in topic position. In other words, the topic must be in clause-initial position. Then it should be possible to topicalise entire VPs and also recursively extraposed VPs. This is precisely what we find.

(4.7) dem Jungen das Buch schenken wollte Peter
 the boy the book give wanted Peter
 'Peter wanted to give the boy the book'

(4.8) a. Hans hat sich geweigert, dem Richter zu gestehen, die Tat
 Hans has himself refused the judge to confess, the act
 begangen zu haben
 committed to have
 'Hans has refused to confess to having committed the act to the
 judge'

b. Dem Richter zu gestehen, die Tat begangen zu haben, hat Hans
 sich geweigert

In (4.7), the VP complement *dem Jungen das Buch schenken* of the finite verb *wollte* has been topicalised. (4.8a) is a V2 clause where the VP *die Tat begangen zu haben* is recursively extraposed in the VP *dem Richter zu gestehen, die Tat begangen zu haben* which is extraposed in the finite clause itself. (4.8b) shows that the VP *dem Richter zu gestehen, die Tat begangen zu haben* can be topicalised instead of extraposed and that the VP *die Tat begangen zu haben* is in fact a daughter of the VP headed by *zu gestehen* and not the finite clause. (Cf. Figure 4-7.)

Examples like (4.9) would appear to cause problems however.

(4.9) ?versucht zu behaupten hat er, das Buch gelesen zu haben

At first sight, (4.9) looks very problematic.¹ The VP *das Buch gelesen zu haben* is dependent on *zu behaupten* (which in turn is extraposed in the topicalised VP

¹This example is not accepted by many speakers.

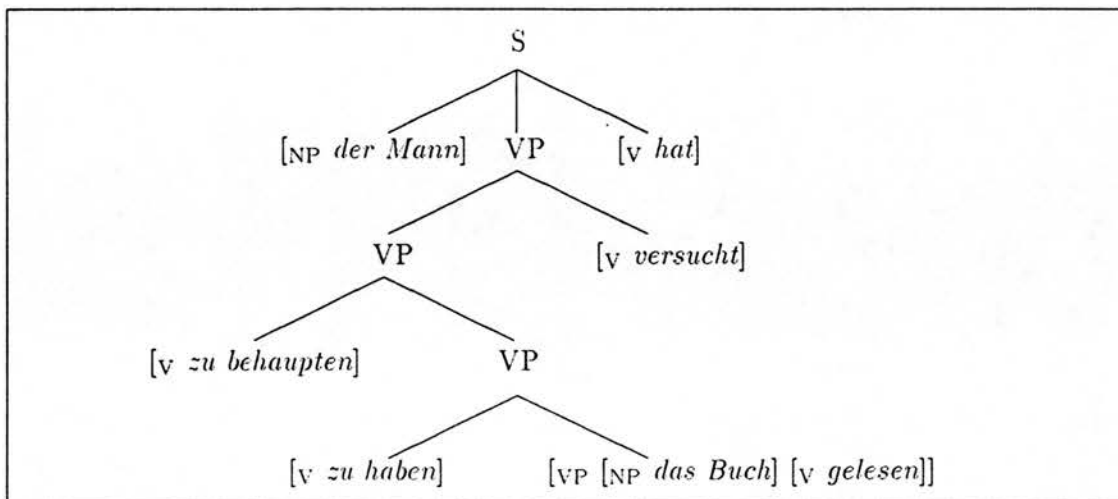


Figure 4-5: Syntax tree for (4.2)

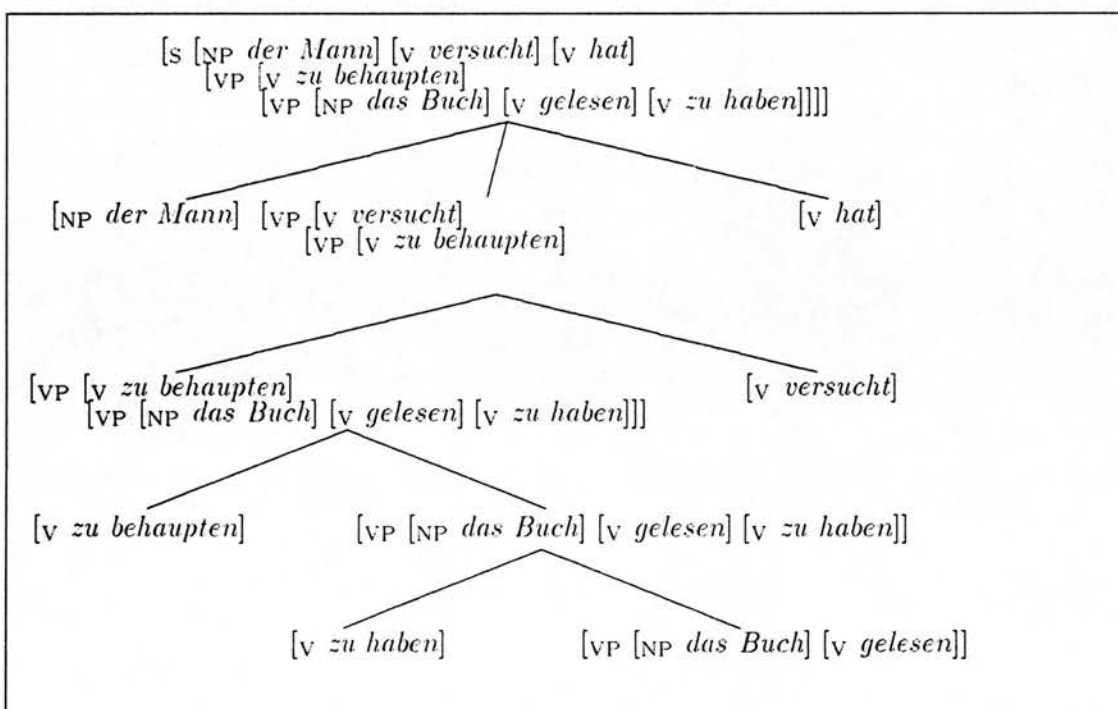


Figure 4-6: Domain tree for (4.2)

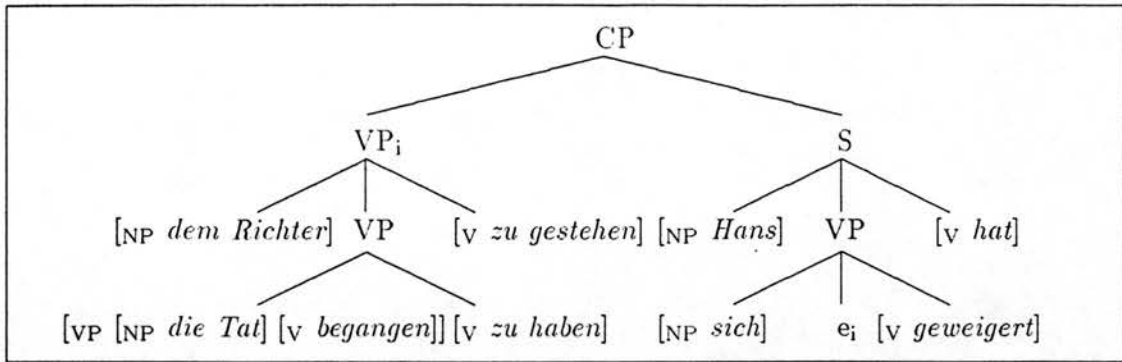


Figure 4-7: Syntax tree for (4.8b)

versucht zu behaupten) but it is in clause-final position rather than in topic-final position as we would expect from the account of VP extraposition above. This is rather disturbing since it would appear to require a special rule or device to allow clause-final extraposition in just this one case when we have derived the other cases from general properties of the analysis. However, there is other data which is relevant to this problem. As is well known, it is possible to front a verb with some or none of its complements.

- (4.10) a. dem Jungen das Buch schenken wollte Peter
 b. dem Jungen schenken wollte Peter das Buch
 c. das Buch schenken wollte Peter dem Jungen
 d. schenken wollte Peter dem Jungen das Buch

One explanation that has been proposed by den Besten and Webelhuth ([den Besten and Webelhuth 88]) is that the complements which are “left behind” have been “moved out” of the VP to the finite clause prior to its topicalisation. I.e., *dem Jungen schenken*, *das Buch schenken* and *schicken* in (4.10b)-(4.10d) respectively are really VPs and not nonmaximal verb projections. Given this analysis, the null hypothesis predicts that constituents can be “moved up” from extraposed VPs as well. Uszkoreit ([45]) provides examples of just this type. (Cf. (4.11).)

- (4.11) Letztes Jahr hatte Peter [das große Haus]; der Stadt
 last year had Peter the big house the city
 versprochen e_i zu reparieren
 promised to repair
 ‘Peter promised the city to repair the big house last year’

(4.11) shows that this movement can cooccur with topicalisation. Uszkoreit also provides examples that show that it can cooccur with *wh*-movement and that

more than one constituent can be moved. All of these examples, however, involve movement to the post-finite verb field. The null hypothesis should be that these constituents can move up to VPs which intervene between the extraction site and the finite clause. However, the only way to see this theory-neutrally is to see if constituents move from an extraposed VP governed by a VP which is itself extraposed (i.e., a case of recursive extraposition) to the extraposed governing VP itself. Hans den Besten (p.c.) has shown that there are examples in both German and Dutch where this is in fact possible.

How then can these facts be integrated into the account outlined above? At a purely descriptive level, we want to say that such "raised" constituents are moved to the domain of the VP or S that they appear in. Then the LP constraints and other grammatical factors come into play to treat them just like any other element of that domain. All the evidence seems to support that this "equal treatment" does in fact hold. We're now in a position to explain (4.9). Assume that raised constituents can also be marked [EXTRA +], i.e., can be extraposed. Then (4.9) can be explained by assuming that *das Buch gelesen zu haben* is marked [EXTRA +] and raised out of the VP *versucht zu behaupten, das Buch gelesen zu haben* into the domain of the finite clause. Since *das Buch gelesen zu haben* is marked [EXTRA +] it will appear in clause-final position.

A few words are in order concerning the style of analysis. First, the word order domains are very flat. There isn't even an identifiable verb cluster domain in contrast with most assumptions about the constituent structure of German subordinate clauses. Coordination data suggests that verb clusters are constituents, and therefore, have continuous domains. However, there is a great deal of nonconstituent coordination evidence in the Mittelfeld which suggests that coordination is not a good metric for determining whether verb clusters are constituents or not. Furthermore, topicalisation of verb clusters or partial verb clusters is not necessarily an indication of constituency either as this can be explained in terms of raising and remnant topicalisation as discussed above. Neither is the fact that verb clusters cannot be interrupted by any other material. Dowty ([12]) presents an analysis of English which contains *attachment operators* in addition to operations equivalent to sequence union. These attachment operators are just like sequence union except that they force the heads of the two sequences being unioned to be "attached", i.e., immediately adjacent and not interrupted by any other material in any domain. These operators could explain why verb clusters must not be interrupted and are prosodic phrases. That is, domain union may not only involve sequence union but also adjacency of the head verbs. The fact that verb clusters are prosodic phrases does not imply that a verb cluster must be a constituent in a VP or S domain though. So far, it has not been necessary to use an immediate precedence relation although its role in the treatment of clitics and clitic-like elements is of obvious utility.

A second reason that verb clusters contain no internal domain structure is based on empirical evidence. Typical verb raising analyses of verb clusters assume a nested \bar{V} structure where each level is of the form $[\bar{V} V \bar{V}]$ or $[\bar{V} \bar{V} V]$. This covers the possible 1-2, 2-1 and 1-2-3 government orders of Dutch and

all of the possible Standard German verb cluster orders (and many more). These can all be dealt with very easily within the current account since a verb can be lexically specified to occur to the right or left of the verbs it governs. This is equivalent to “direction of status government” in von Stechow’s terms ([46]). However, a government order of 4 – 1 – 3 – 2 could not be accounted for in terms of a verb raising analysis unless adjunction to the highest \bar{V} was allowed and then movement of the V_4 to the adjoined position. Such orders can be found in nonstandard dialects of German. In Zurich German, there seems to be no restriction on the relative order of auxiliaries, modals, verbs of perception and the causative within a domain. In fact, these verbs need not even form a verb cluster. (Cf. [7]). Furthermore, it can be shown that such instances are not examples of extraposition or verb projection raising. This dialect evidence suggests strongly that the verb cluster is not a constituent in word order domains.

4–2 Netter (1991)

In [28], Klaus Netter presents a two dimensional classification of verb projections in German as a way of investigating certain difficulties for configurational analyses of German. The two dimensions are:

- *clause union*, which can be roughly equated with the merging of subcategorisation frames, and
- *verb raising*, which can be roughly equated with the formation of a verb cluster.

Furthermore, Netter assumes a more or less traditional analysis of control verbs and causative verbs as taking clausal complements. He then presents seven different categories of data that one might investigate to try to determine whether a given construction involves clause union and/or verb raising. He then shows that each of these categories presents difficulties for configurational analyses with respect to a wide range of data.

In what follows, I’ll present an alternative to the “clause union-verb raising” account which captures some of the generalizations that Netter’s data seem to need to be given but which is unorthodox in many respects. My purpose is not to criticise or review Netter’s analysis or proposals but to review Netter’s data and the empirical generalisations which he presents and to provide an alternative account for the empirical facts which does not suffer from some of the difficulties that configurational accounts suffer.² Before I can present the linguistic assumptions that I make about German, I must first say something about the basic approach to word order and constituency that I assume.

²For a discussion of the difficulties which configurational analyses of German suffer, the reader is referred to Netter’s paper, [28].

4-3 Some assumptions about German verbs

In what follows, I will make some tentative hypotheses about German verbs and investigate their consequences in terms of Netter's data given the model of word order and constituency that I have outlined. These analyses are rather preliminary but they do indicate the power to capture empirical generalizations that the approach provides.

As Netter points out, extraposed VP complements must be *zu*-infinitival VPs and they must not be the complements of a raising verb.³ The first fact I simply stipulate since I do not have an explanation for it.⁴ The second fact I will try to provide a tentative explanation for.

I make the following assumptions.

1. Raising verbs take infinitival *clauses* (Ss) as their only complements. Agreement between the subject of the nonfinite clause and the raising verb is accomplished via some means other than subcategorisation as in HPSG or the type of "syntactic" agreement found in GB which I will not investigate here.⁵ *scheinen* takes a *zu* S and *müssen* takes a bare infinitival S. Raising verbs also obligatorily domain union their single S complement.
2. Only VP[ZU]s and \bar{S} s can be extraposed. We have no explanation for the fact that Ss cannot be extraposed.
3. Verbs which have both a raising and non-raising reading have two different subcategorisation frames corresponding to their semantics. The raising reading is both syntactically and semantically a one-place predicate. The non-raising reading is both syntactically and semantically a two-place predicate. Thus, a raising reading subcategorises for a single clause and a non-raising reading subcategorises for an NP and a VP.
4. The causative *lassen* and object control verbs with an overt controlled NP constituent subcategorise for an NP and a VP complement. They do not take clausal complements as in GB.
5. Occurrences of *lassen* which do not take a controlled object subcategorise for a subject and a single VP.

The following are the key ideas upon which the architecture of the analysis is based.

³There are at least two apparent exceptions to this second restriction which I will discuss later.

⁴However, Zurich German seems to provide evidence that it allows the extraposition of bare infinitival VPs.

⁵One way in which this could be accomplished is by making the "agreement features" HEAD features of the head verb of a clause. Then a raising verb could select a clause with the appropriate agreement features which agree with the inflectional morphology of the raising verb. Alternatively raising verbs could select for the index of the subject of a level of functional structure partly similar to the *functional structure* of LFG.

1. Nonextraposed VPs are domain unioned, which accounts for scrambling.
2. Apparent instances of scrambling out of topicalised or extraposed VPs are really instances of a different phenomenon which Uszkoreit ([45]) calls *focus-raising*.
3. In spite of appearances, scrambling (domain union) is possible from the VP complement even in the presence of an NP object of the same governing verb.
4. Raising verbs (*scheinen*, epistemic modals, and optionally auxiliaries) take S complements but *lassen* subcategorises for NP and VP.

We will now turn to the data.

4-4 Extraposition

As Netter points out, the crucial empirical generalizations about extraposition of VPs is that an extraposed VP must be a *zu*-infinitival VP and it cannot be the complement of a raising verb. Thus the VP *zu arbeiten* can be in coherent position (KN3) or extraposed (KN4) when it is the complement of a non-raising verb like *versuchen*. However, although the coherent variant (KN1) is okay with the raising verb *scheint*, the extraposed variant (KNex2) is ungrammatical.

(KN3) weil Fritz zu arbeiten versucht
 because Fritz to work tries
 ‘because Fritz tries to work’

(KN4) weil Fritz versucht, zu arbeiten
 because Fritz tries to work

(KN1) weil Fritz zu arbeiten scheint
 because Fritz to work seems
 ‘because Fritz seems to work’

(KN2) *weil Fritz scheint, zu arbeiten
 because Fritz seems to work

These facts follow immediately from our assumptions. We predict the grammaticality of (KN1) since *scheint* will take the S[ZU] *Fritz zu arbeiten* as its only complement and union it into the mother’s domain. That is, the syntactic structure of (KN1) is as in (4.12) and the domain structure is (4.13).

(4.12) [s [s [NP Fritz] [v zu arbeiten]] [v scheint]]

(4.13) [s [NP Fritz] [v zu arbeiten] [v scheint]]

We predict the ungrammaticality of (KN2) simply because 'zu arbeiten' is not a VP in the constituent structure of (KN2). As we stated before, VPs do not form clauses with NPs. Rather, the head verb is sister to all of its complements in a clause. Since *scheint* subcategorises for a clause, *zu arbeiten* is not a VP in the constituent structure of (KN2).

On the other hand, nonraising verbs which take a VP[ZU] complement allow the VP to be domain unioned (KN3) or extraposed (KN4).

(KN7) indicates that *versprechen* can have both a raising and non-raising reading. In this case, we would say that the two readings result from different syntactic and semantic structures according to our assumptions.

(KN7) weil das Baby sich zu einem Genie zu entwickeln
 because the baby itself into a genius to develop
 versprach
 promised
 (a) 'because the baby promised to become a genius'
 (b) 'because it was very likely that the baby would turn into a
 genius'

(KN7a) has the syntactic structure (4.14) and domain structure (4.15) and (KN7b) has the syntactic structure (4.16) and domain structure (4.17).

(4.14) [s [NP das Baby] [VP [NP sich] [NP zu einem Genie] [v zu entwickeln]]
 [v versprach]]

(4.15) [s [NP das Baby] [NP sich] [NP zu einem Genie] [v zu entwickeln]
 [v versprach]]

(4.16) [s [s [NP das Baby] [NP sich] [NP zu einem Genie] [v zu entwickeln]]
 [v versprach]]

(4.17) [s [NP das Baby] [NP sich] [NP zu einem Genie] [v zu entwickeln]
 [v versprach]]

Notice that the two domains (4.16) and (4.17) are identical.

In (KN8) we now predict that the raising reading is out since there is no VP to extrapose under that reading (corresponding to the syntactic structure (KN7b-syntax)). This is similar to the reasoning behind the ungrammaticality of (KN2).

Since (KNS) is ungrammatical with the raising reading, the raising reading is not available.

- (KNS) weil das Baby versprach, sich zu einem Genie zu
 because the baby promised itself into a genius to
 entwickeln
 develop
 (a) 'because the baby promised to become a genius'
 (b) 'because it was very likely, that the baby would turn into a
 genius'

There are (at least) two raising verbs which appear to allow VP[ZU] extraposition.

- (4.18) daß ihm begann schlecht zu werden
 that him begann sick to get
 'that he started to get sick'

- (4.19) als es anfing zu regnen
 when it started to rain
 'when it started to rain'

(4.18) makes it clear that this is no ordinary English-style raising verb. As we will see in the next section, *ihm schlecht zu werden* is a subjectless construction. This means that it is syntactically saturated and since a syntactically saturated verb projection is a clause *ihm schlecht zu werden* is an S[ZU]. *ihm* semantically and syntactically is part of this clause. So, it looks as if a (partial) clause has been extraposed in violation of our assumptions. For *ihm* to appear to the left of *begann*, it would have to be moved there from the extraposed S[ZU]. Perhaps this arises from factors which rule out the sequence C V[FIN] Another possibility is that *begann* domain unions with the clause. This would explain the position of *ihm*. *schlecht* could be where it is because of its "fixed" interpretation. Incorporation of elements like *schlecht* into the verb cluster are common enough. More evidence needs to be found to determine *beginnen's* properties. In any event, it most definitely does not extrapose a VP[ZU] in (4.18). Similarly, in (4.19), *anfing* looks as if it extraposes the VP *zu regnen*. However, this could be another case of clause extraposition plus movement or domain union with the clause. In either case, it will be difficult to tell what is going on because the results will look like ordinary VP extraposition.

4-5 Subjectless Constructions

German systematically exhibits so-called *subjectless constructions*, i.e., clauses or clause-like constituents which do not contain a subject. As Netter points out,

such subjectless constructions can only appear as the complements of raising verbs, as in (KN17)-(KN19).

(KN17) *weil Hans ihm schlecht zu werden behauptete
 because Hans him (DAT) sick to get claims
 'because Hans claims to get sick'

(KN18) weil gearbeitet zu werden scheint
 because worked to be seems
 'because it seems that someone works'

(KN19) weil ihnen schlecht zu werden droht
 because them (DAT) sick to get threatens
 'because there is some danger that they get sick'

The problem for the account here is to explain this fact. There is one main point to note. Verb projections with saturated subcategorisation requirements and which contain no subject, i.e., so-called *impersonal constructions* are perfectly wellformed structures and by the definition of *clause* in HPSG, they must be *clauses*, i.e., Ss. I.e., since they are of the form $v[\text{ARGS } \langle \rangle]$, they must be sentences. They need not have a subject structurally or functionally.

The ungrammaticality of (KN17) follows automatically from our assumption that raising verbs take Ss as complements. *ihm schlecht zu werden* is a S[ZU] which has no SUBJ function. However, the finite EQUI verb *behauptete* subcategorises for a SUBJ NP and a VP[ZU]. Thus, (KN17) is ungrammatical because the subcategorisation requirements are not satisfied.

The grammaticality of (KN18) and (KN19) is also predicted by our assumptions. *Gearbeitet zu werden* is also a S[ZU]. We can see this by comparing it with the finite *daß gearbeitet wurde* which is a perfectly acceptable subordinate clause. Since *Gearbeitet zu werden* is a S[ZU], it is an acceptable complement for *scheint* and so (KN18) is grammatical. Similarly (KN19) is fine since *droht* has a raising reading in this context. (Cf. (KN17).)

(KN20) is also predicted to be fine except that this time *müssen* subcategorises for an infinitival S instead of a S[ZU] like *scheinen*. On the other hand (KN21) is out because *wollen* is not a raising verb but an EQUI verb, subcategorising for a subject NP and a VP complement.

(KN20) weil gearbeitet werden muß
 because worked be must
 'because there has to be somebody working'

(KN21) *weil gearbeitet werden will
 because worked be wants
 'because it is desirable that somebody is working'

4-6 Pronominalization

As we stated at the outset, we assume that control verbs subcategorise for a (possible) NP complement and a VP complement and that raising verbs subcategorise for Ss. In what follows, we will investigate *es*-pronominalisation of VPs. The key empirical generalisation is that the VP complement of non-raising verbs can be pronominalised by the pronoun *es* but apparent VP complements of raising verbs cannot be. We will show that this follows straightforwardly from our assumption that raising verbs subcategorise for clauses and not VPs.

(KN24) Fritz versucht den Roman zu lesen und Maria versucht es auch
 Fritz tries the novel to read and Maria tries it also
 'Fritz tries to read the novel and Maria does too'

(KN25) *Fritz scheint den Roman zu lesen und Maria scheint es auch
 Fritz seems the novel to read and Maria seems it also
 'Fritz seems to read the novel and Maria does too'

(KN24) then follows straightforwardly from our assumptions. *den Roman zu lesen* is a VP[ZU] in the first conjunct and so *es* can be a pronominal for it in the second conjunct. (KN25) is ungrammatical since *Fritz den Roman zu lesen* is an S[ZU] and *den Roman zu lesen* is not a VP[ZU] in (KN25), there is no nonfinite VP for *es* to be the pronominal of. That is, the configuration [_S NP[NOM] VP] is ungrammatical.

scheinen can occur with the pronoun *es* however. Consider (4.20).

(4.20) Fritz scheint dir den Roman zu lesen und mir scheint es auch
 Fritz seems you the novel to read and me seems it also
 'Fritz seems to you to read the novel and so does it to me too'

Here we see that the pronominal *es* is coreferential with the clause *Fritz den Roman zu lesen*. *dir* and *mir* are the different indirect objects of *scheint*.

(KN26) Fritz versprach sich gut zu entwickeln und Peter versprach
 Fritz promised himself well to develop and Peter promised
 es auch
 it also
 (a) 'Fritz promised to develop well and Peter promised it too'
 (b)*'There were good chances that Fritz would develop well and
 that Peter would too'

(KN26) follows by the same reasoning. (KN26a) is fine since *sich gut zu entwickeln* is a VP[ZU] and therefore an acceptable complement for the *control* reading of *versprach* and can be pronominalized by *es*. On the other hand, in (KN26b), *sich*

gut zu entwickeln is not a constituent (*Fritz sich gut zu entwickeln* is) and so it cannot be pronominalized.

- (KN27) Fritz muß Klavier spielen
 Fritz must piano play
 (a) 'Fritz must play the piano' (= has the obligation to)
 (b) 'Fritz should be playing the piano' (= it is likely that)
- (KN28) Fritz muß Klavier spielen und Maria muß es auch
 Fritz must piano play and Maria must it also
 (a) 'Fritz has the obligation to play the piano and Maria has the obligation too'
 (b)*'It should be the case that Fritz is playing the piano and Maria too'
- (KN29) Fritz muß in der Bibliothek sein und Maria muß es auch
 Fritz must in the library be and Maria must it also
 (a) 'Fritz has the obligation to be in the library and Maria has the obligation too'
 (b)*'It should be the case that Fritz is in the library and Maria too'

(KN27) illustrates once again that *müssen* can have both a raising and a non-raising reading and (KN28) and (KN29) illustrate that pronominalization is possible with the non-raising reading but not the raising reading for the same reasons as outlined above.

- (KN31) Fritz läßt ihn den Roman lesen und Maria läßt es ihn auch
 Fritz lets him the novel read and Maria lets it him also
 'Fritz lets him read the novel and Maria lets him (do it) also'
- (KN32) *Fritz läßt ihn den Roman lesen und Maria läßt es auch
 Fritz lets him the novel read and Maria lets it also
 'Fritz lets him read the novel and Maria lets (do) it also'

(KN31) and (KN32) provide evidence for our assumption that *lassen* subcategorizes for an accusative NP and a VP[INF] complement. I.e., it is essentially an object control verb. In (KN31), the VP[INF] *den Roman lesen* can be pronominalized. This would not be predicted if a clausal complement was hypothesized for *lassen*. Conversely, in (KN32), *ihn den Roman lesen* cannot be pronominalized since it is not a constituent, i.e., not a VP or an S. This also partially supports the hypothesis that a clausal analysis of *ihn den Roman lesen* is incorrect.

- (KN33) Fritz läßt den Roman lesen und Maria läßt es auch
 Fritz lets the novel read and Maria lets it also
 'Fritz lets the novel be read and Maria lets it (be done) also'

- (KN34) *Fritz läßt den Roman lesen und Maria läßt es ihn
 Fritz lets the novel read and Maria lets it it (the novel)
 auch
 also
 'Fritz lets the novel be read and Maria lets it (be done) also'

(KN33) indicates that *den Roman lesen* is a VP in the context where it is the only nonsubject complement of *lassen*. I.e., it is not the case that *den Roman* is the "controlled" complement of *lassen* where *lesen* is essentially taken to be passive despite its morphology. This is the so-called "unannounced passive". (KN34) likewise indicates that it is impossible to pronominalize *lesen* thus indicating that it is not a VP but that *den Roman lesen* is. Therefore, we assume in these cases that the "controlled" complement of *lassen* has been eliminated through some process of lexical *argument reduction*.

Note that if these examples had contained instances of scrambling in which the pronominalized VP had not been continuous that pronominalization would still be predicted to be possible. This is in fact possible and is problematic for some other theories which would have to reconstruct a contiguous VP for the pronominal to be coreferential with.

4-7 Scrambling

Here we use "scrambling" as a descriptive term to mean that an NP complement appears to the left of its "canonical" position. The canonical position is given by its non-domain-unioned position. Netter claims that if a verb takes an object in addition to a VP complement, then scrambling is impossible out of the VP complement. He presents (KN42)-(KN44) in support of this claim. (KN42) contains an object control verb, (KN43) a subject control verb and (KN44) the causative *lassen*. In all three cases, the object of the controlled verb is scrambled to the left of the object of the control verb resulting in unacceptable sentences.

- (KN42) *weil ihn der Fritz dem Hans abzuholen empfahl
 because him Fritz Hans to pick up recommended
 'because Fritz recommended to Hans to pick him up'
- (KN43) *weil ihn der Fritz dem Hans abzuholen versprach
 because him Fritz Hans to pick up promised
 'because Fritz promised to Hans to pick him up'
- (KN44) *weil der Fritz ihm den Hans nicht helfen läßt
 because Fritz him (DAT) Hans (ACC) not help lets
 'because Fritz doesn't let Hans help him'

In our terms, the inability to scramble means that domain union is impossible. Instead, the entire VP complement simply appears to the left of the verb that subcategorises for it. In general, the structure would look like (4.21).

(4.21) [... [VP_{n+1} ... V_{n+1}] V_n ...]

(KN45) and (KN46) provide additional evidence that *lassen* takes an accusative NP object and a VP[INF] complement.

(KN45) weil ihn der Fritz dem Hans nicht helfen läßt
 because him (ACC) Fritz Hans (DAT) not help lets
 'because Fritz doesn't let him help Hans'

(KN46) *weil ihm der Fritz den Hans nicht helfen läßt
 because him Fritz Hans not help lets
 'because Fritz doesn't let Hans help him'

In (KN45), we see that the object *ihn* of *lassen* can be scrambled in front of the subject *Fritz*. We would expect this to be acceptable if the domain structure of (KN45) is (4.22) since *ihn* and *der Fritz* are elements of the same clausal domain and pronouns may in general precede NPs.

(4.22) [s ihn der Fritz [VP dem Hans nicht helfen] läßt]

(Of course, the major difficulty for the domain union account is to explain the position of *nicht* if the VP *dem Hans helfen* has not been domain unioned. We will return to this below.)

On the other hand, (KN46) is predicted to be ungrammatical since the object *ihm* of the controlled verb *helfen* has been scrambled out of its domain to the left of the matrix subject *der Fritz* as in (4.23).

(4.23) [s ihm_i der Fritz den Hans [VP [e_i nicht helfen] läßt]

The generalization about scrambling being restricted to verbs without an object would predict the acceptability of (KN47) immediately. However, (KN48) fits no such pattern. It seems that the unaccusativity of *fallen* is the factor which allows the scrambling to occur. Although we have no explanation for the identical behaviour of *lassen* with unannounced passive complements and unaccusative complements, reflexivisation data indicates that they form a natural class with respect to reflexivisation (namely, that neither complement has an agentive subject).⁶

(KN47) weil ihm der Fritz helfen läßt
 because him Fritz help lets
 'because Fritz lets him be helped'

⁶Cf. ([16]) for some discussion of this issue.

- (KN48) weil ihm der Fritz den Stein auf den Kopf
 because him (DAT) Fritz the stone (ACC) on the head
 fallen läßt
 fall lets
 'because Fritz lets the stone fall on his head'

However, the apparent prohibition against domain union of control verbs with an object NP complement has some problems. For example, if we consider our very first example *es ihm jemand zu lesen versprochen hat* we see that *versprochen* takes a dative object *ihm* which is separated from it by the subject *jemand*.

- (4.24) daß es₃ ihm₂ jemand₁ zu lesen₃ versprochen₂ hat₁
 that it him someone to read promised has
 'that some promised him to read it'

(KN50) and (KN51) are two instances of *lassen* where the object of the governed verb is scrambled to the left of the subject *niemand* in (KN50) and where it is scrambled to the left of both the subject *niemand* and the object *ihn* of *läßt* in (KN51).

- (KN50) weil ihn₁ das₂ niemand₁ machen₂ läßt₁
 because him this nobody do lets
 'because nobody lets him do this'

- (KN51) weil es₂ ihn₁ niemand₁ lesen₂ läßt₁
 because it him nobody read lets
 'because nobody lets him read it'

(KN52) is an instance where the object pronoun *das* of the verb *zu tun* governed by the object control verb *bat* is scrambled to the left of the subject *niemand*.

- (KN52) weil ihn₁ das₂ niemand₁ zu tun bat₁
 because him this nobody to do asked
 'because nobody asked him to do this'

(KN55a) is an instance where the object pronoun *es* of the verb *zu lesen* governed by the control verb *bat* is scrambled to the left of the object *ihn* of *bat*. If the scrambling is taken as evidence of domain union (or clause union) then we would expect the negation to take both narrow and wide scope over the verbs. This is precisely what we find. Furthermore, if we reverse the scrambling, i.e., reverse the order of *es ihn* to *ihn es* the example becomes ungrammatical (KN55b).⁷

⁷(KN55b) is acceptable to some speakers. This causes no particular problem for our analysis. It simply suggests that the "canonical order" *ihn es* is always acceptable to some speakers.

However, the sequence *ihn es* is very strange. *es* is “lighter” than *ihn* and almost always precedes it. This suggests that *es* and *ihn* are in the same domain even when one of them is the object of a control verb and the other is the object of the governed verb. Therefore, unless pronouns triggers domain union in object control constructions while phrasal NPs do not, then we have to conclude that all VP complements are domain unioned when they are in coherent position (i.e., when they are not extraposed or topicalised).

- (KN55) a. weil der Fritz₁ es₂ ihn₁ nicht zu lesen₂ bat₁
 because Fritz it him not to read asked
 (a) ‘because Fritz asked him not to read it’
 (b) ‘because Fritz didn’t ask him to read it’
- b. *weil der Fritz₁ ihn₁ es₂ nicht zu lesen₂ bat₁

There is considerable other evidence that control verbs with object complements do undergo domain union with their VP complement. Consider (4.25).

- (4.25) weil dieses Machwerk kein Vater seinen Kindern zu lesen
 because this sorry effort no father his kids to
 erlauben würde
 read permit would
 ‘because no father would permit his kids to read such a sorry
 effort’

In fact, there is a considerable body of data which indicates that scrambling is sometimes permitted and sometimes not permitted depending on various sentential factors.

First, we might try to find examples of extraposed control verbs with an object complement which allow scrambling. As the following data indicates, this is rather difficult for simple examples.

- (4.26) a. ‘because Fritz promised to persuade Hans to help him’
- b. *weil der Fritz versprochen hat, ihm den Hans zu helfen zu überreden

Neither does this work if both NPs are pronouns.

- (4.27) a. weil der Fritz versprochen hat, den Hans ihm zu helfen zu
 because the Fritz promised has the Hans him to help to
 überreden weil der Fritz versprochen hat, ihn es zu lesen zu
 persuade because the Fritz promised has him it to read to
 überreden
 persuade
 'because Fritz promised to persuade him to read it'
- b. *weil der Fritz versprochen hat, es ihn zu lesen zu überreden

However, the following examples do seem to be better.

- (4.28) a. weil der Fritz versprochen hat, ihn es zu lesen zu bitten
 because the Fritz promised has him it to read to ask
 'because Fritz promised to ask him to read it'
- b. ?weil der Fritz versprochen hat, es ihn zu lesen zu bitten
- (4.29) a. weil der Fritz versprochen hat, niemanden es zu lesen zu
 because the Fritz promised has noone it to read to
 überreden
 persuade
 'because Fritz promised to persuaded noone to read it'
- b?⁺ weil der Fritz versprochen hat, es niemanden zu lesen zu
 überreden
- (4.30) a. weil der Fritz versucht hat, ihm es zu lesen zu versprechen
 because the Fritz tried has him it to read to promise
 'because Fritz tried to promise him to read it'
- b?⁺ weil der Fritz versucht hat, es ihm zu lesen zu versprechen
- (4.31) a. weil jemand versucht hat, ihn es nicht zu lesen zu
 because someone tried has him it not to read to
 überrededen
 persuade
 'because someone tried to persuade him not to read it'
- b. ?weil jemand versucht hat, es ihn nicht zu lesen zu versprechen

- (4.32) a. weil der Fritz versuchte niemanden es zu lesen zu
 because the Fritz tried noone it to read to
 versprechen
 promise
 'because Fritz tried to promise noone to read it'
- b. weil der Fritz versuchte es niemanden zu lesen zu versprechen

In (4.33), the fact that both objects are pronominals seems to license the scrambling that we find. We can make several observations. First, the "canonical" order corresponding to absence of domain union is acceptable in (4.33d). It appears that this "canonical" order is always possible. The order in (4.33f) is expected also if there is no domain union. However, the pronominal sequences *es ihn* in (4.33b) and *ihn es* in (4.33e) appear scrambled to the left of the subject in clear violation of the empirical generalisation. Furthermore, the swapping of *es* and *ihn* to the right of the subject in (4.33c) and the fronting of the pronoun *es* in (4.33b) is marginally acceptable or acceptable for some speakers.

- (4.33) a. ?
 because it the Fritz him to read persuaded weil es der Fritz ihn
 zu lesen überredete
 'because Fritz persuaded him to read it'
- b. weil es ihn der Fritz zu lesen überredete
- c. ?weil der Fritz es ihn zu lesen überredete
- d. weil der Fritz ihn es zu lesen überredete
- e. weil ihn es der Fritz zu lesen überredete
- f. weil ihn der Fritz es zu lesen überredete

Another example which shows even more scrambling is (4.34).

- (4.34) a. ?
 because it some him to read persuaded weil es jemand ihn zu
 lesen überredete
 'because someone persuaded him to read it'
- b. weil es ihn jemand zu lesen überredete
- c. ?weil jemand es ihn zu lesen überredete
- d. weil jemand ihn es zu lesen überredete
- e. weil ihn es jemand zu lesen überredete
- f. weil ihn jemand es zu lesen überredete

In this case the indefinite pronominal subject NP seems to license the scrambling. Of the six possible orders of the three NPs, only two are judged marginal. The other four are judged acceptable. The two marginal variants correspond to the marginal variants in (4.33).

If we now substitute the negative pronoun *niemand* for the indefinite pronoun *jemand* then all six possible orders become acceptable as in (4.35).

- (4.35)
- a. weil es niemand ihn zu lesen überredete
because it noone him to read persuaded
'because noone persuaded him to read it'
 - b. weil es ihn niemand zu lesen überredete
 - c. weil niemand es ihn zu lesen überredete
 - d. weil niemand ihn es zu lesen überredete
 - e. weil ihn es niemand zu lesen überredete
 - f. weil ihn niemand es zu lesen überredete

On the other hand, if the subject is a nonlexical NP but the sentence is negated, then all six possible orders are acceptable as in (4.36).

- (4.36)
- a. weil es ihn der Fritz nicht zu lesen bat
because it him the Fritz not to read asked
'because Fritz didn't ask him to read it'
 - b. weil es der Fritz ihn nicht zu lesen bat
 - c. ihn es der Fritz nicht zu lesen bat
 - d. ihn der Fritz es nicht zu lesen bat
 - e. weil der Fritz es ihn nicht zu lesen bat
 - f. weil der Fritz ihn es nicht zu lesen bat

If the object of the control verb is the negative pronoun *niemanden* then scrambling is also licensed as in (4.37). Here we see that only preposing of the embedded object *es* to the front of the subject by itself is judged marginal and the sequence

niemanden es before the subject is judged unacceptable.

- (4.37)
- a. ?weil es der Fritz niemanden zu lesen bat
because it the Fritz noone to read asked
'because Fritz asked noone to read it'
 - b. weil es niemanden der Fritz zu lesen bat
 - c. weil der Fritz es niemanden zu lesen bat
 - d. weil der Fritz niemanden es zu lesen bat
 - e. *weil niemanden es der Fritz zu lesen bat
 - f. weil niemanden der Fritz es zu lesen bat

Another class of examples involves demonstrative objects. (4.38) is similar to (4.25). Surprisingly, fronting of the embedded demonstrative object in front of the subject by itself is fully acceptable as in (4.38c). Object-swap as in (4.38)a and the canonical order as in (4.38)b are also both fully acceptable. Surprisingly, the other three possible orders are judged marginal as well.

- (4.38)
- a. weil der Fritz dieses Haus niemanden betreten lässt
because the Fritz this house noone enter lets
because Fritz lets noone enter this house
 - b. weil der Fritz niemanden dieses Haus betreten lässt
 - c. weil dieses Haus der Fritz niemanden betreten lässt
 - d. ?weil dieses Haus niemanden der Fritz betreten lässt
 - e. ?weil niemanden der Fritz dieses Haus betreten lässt
 - f. ?weil niemanden dieses Haus der Fritz betreten lässt

(4.39) indicates that scrambling is also possible with *lassen* when it takes an NP complement. Surprisingly, the object *es* of *lesen* can be scrambled to the left of the subject as in (4.39a). In addition to the "canonical" order (4.39d), object swap is possible if the object of *lassen* is the indefinite pronoun *jemanden* as in (4.39c). The pronominal sequence *jemanden es* is acceptable for some speakers to

the left of the subject as in (4.39e). Only (4.39b) and (4.39f) are fully marginal.

- (4.39)
- a. als es der Hans jemanden lesen ließ, ...
if it the Hans someone read lets ...
if Hans lets someone read it
 - b. ?als es jemanden der Hans lesen ließ, ...
 - c. als der Hans es jemanden lesen ließ, ...
 - d. als der Hans jemanden es lesen ließ, ...
 - e. ?als jemanden es der Hans lesen ließ, ...
 - f. ?als jemanden der Hans es lesen ließ, ...

There are two approaches to take to this data which violates Netter's empirical generalisation that control verbs which take an additional object complement do not clause union. First, we can claim that particular factors, like the presence of the negation element *nicht*, the indefinite or negative pronouns or the fact that both objects are pronouns, triggers clause union. However, in this case, we would still have to explain why certain permutations of the NPs are grammatical and others are not. The alternative is to state that such control verbs always domain union their VP complements (which allows scrambling to occur) but that there are strong ordering preferences which "block" scrambling in the absence of negation, indefinite and negative pronouns, demonstratives, multiple pronouns and the like. Since the former option would still have to develop a theory of linear precedence constraints, the latter option is the simpler.

We already know that a complex set of linear precedence constraints come into play in German, partly influenced by semantic, pragmatic, prosodic and many other factors. That we should find partial evidence of scrambling for control verbs with an additional object complement and that that evidence should involve complex linear precedence constraints should come as no surprise.

For German, we can have domain union or not have it. There is no problem restricting the class of verbs which allow domain union if this is necessary.⁸ However, this contraindicating data leads me to suspect that control verbs with an object complement domain union coherent VP complements but that there are strong ordering preferences on NPs in object control constructions. Furthermore, if we look beyond German to Dutch we find that NP complements are almost always nonadjacent to their head verbs. I.e., the fact that German tends to exhibit 3-2-1 government order while Dutch tends to exhibit 1-2-3 government order may mislead us into thinking that there is less domain union or "scrambling" in German than there actually is, or at least, that there is less in German than in

⁸Although, we then have to work hard to explain the position and scope of *nicht* in *lassen* constructions. Cf. the discussion in §4-8.

Dutch. Only cross-linguistic and cross-dialectal studies of dialects which “straddle” German and Dutch will reveal whether German avails itself of domain union to an extent comparable to which Dutch does.

A reviewer has pointed out the existence of examples like the following.

- (4.40) a. *weil mit ihm wahrscheinlich zu reden langweilt
 because with him to speak bores
 ‘because to speak with him is boring’
 b. *weil es wahrscheinlich zu lesen faszinierend ist
 because it to read fascinating is
 ‘because to read it is fascinating’

which appear to involve verbs with no object complement which do not domain union their VP complement. I have not investigated such constructions. However, (4.40a) appears to involve a non-domain-unioned VP subject complement and (4.40b) appears to involve a non-domain-unioned VP subject complement of the copula. If this is correct, neither example presents a problem for the analysis presented here.

4–8 Scope of Adjuncts and Negation

In this section, we will assume that the negation element *nicht* and adverbials take a single VP or S argument which they obligatorily domain union with. Furthermore, we’ll assume that they are required to appear to the left of any non-inverted verbs in any domain they occur in.

Then we can explain the asymmetries in (KN68) and (KN69) straightforwardly. In (KN68), *nicht* occurs to the left of *gewagt hat* so it can modify either *gewagt* or *hat*. Semantically, though this gives us the one reading (KN68a). (Cf. Figure 4–8.) However, if it had modified *das Buch zu lesen*, then by domain union, the domain of the modified, extraposed VP would have been *das Buch nicht zu lesen*. But this explains the asymmetry in (KN69) as well. (Cf. Figure 4–9.) To summarize, a negation in pre-verb cluster position cannot take narrow scope over an extraposed VP and conversely, a negation in an extraposed VP cannot take wide scope over any of the verbs in the verb cluster.

- (KN68) weil Fritz nicht gewagt hat, das Buch zu lesen
 because Fritz not dared has the book to read
 (a) ‘because Fritz didn’t dare to read the book’
 (b)*‘because Fritz dared not to read the book’

- (KN69) weil Fritz gewagt hat, das Buch nicht zu lesen
 because Fritz dared has the book not to read
 (a)*‘because Fritz didn’t dare to read the book’
 (b) ‘because Fritz dared not to read the book’

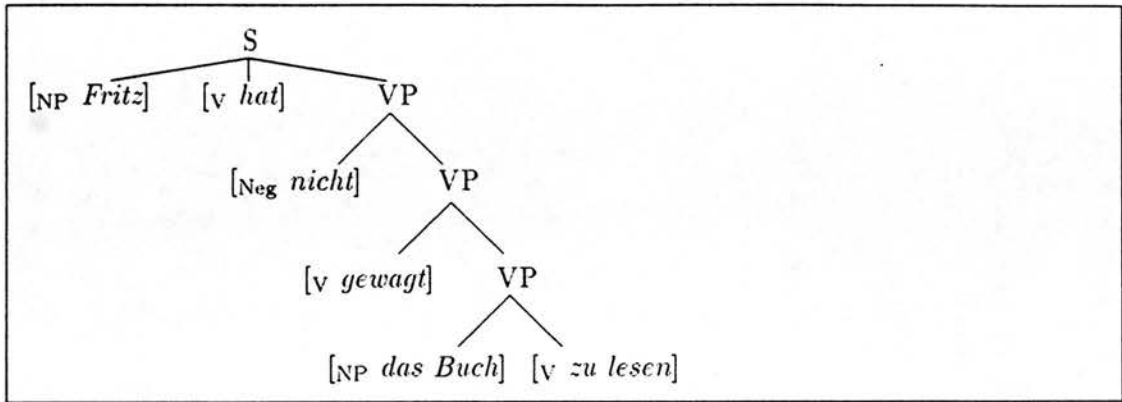


Figure 4-8: Syntax tree of (KN68)

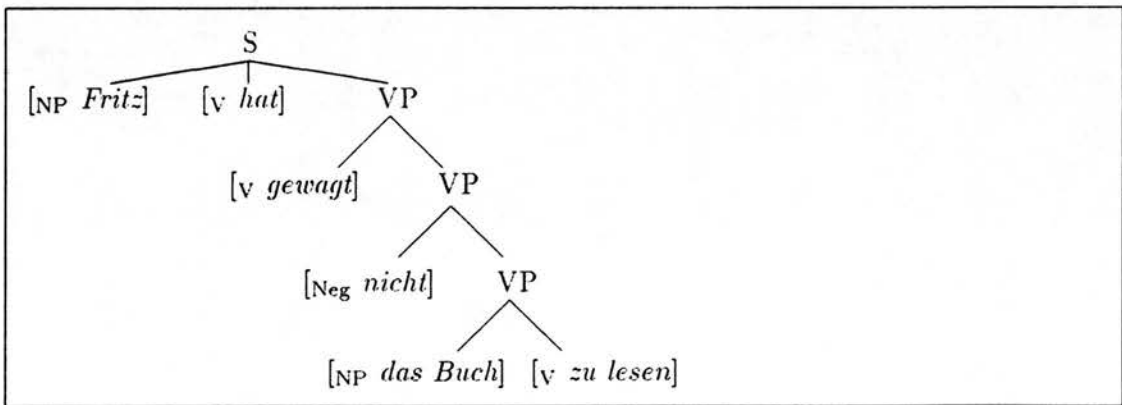


Figure 4-9: Syntax tree of (KN69)

(KN70), (KN71) and (KN72) indicate that the class of verbs which putatively do not allow scrambling also do not allow the negation element *nicht* to take wide scope over the finite verb but do take narrow scope over the embedded VP. This follows automatically, if we continue with the hypothesis that control verbs which take an object complement do not domain union. Therefore, in (KN70), *das Buch mehrmals zu lesen* is a single non-domain-unioned VP domain. But then we can reconstruct via domain union that *mehrmals* modifies *das Buch zu lesen*. This gives reading (KN70a). Furthermore, no other configurational structure could lead to this domain. (Cf. Figure 4-10). (KN71) and (KN72) are exactly analogous.⁹

- (KN70) weil der Fritz ihn [v_{VP} das Buch mehrmals zu lesen]
 because Fritz him the book several times to read
 überredet hat
 persuaded has
 (a) 'because Fritz has persuaded him to read the book several times'
 (b)*'because Fritz has several times persuaded him to read the book'

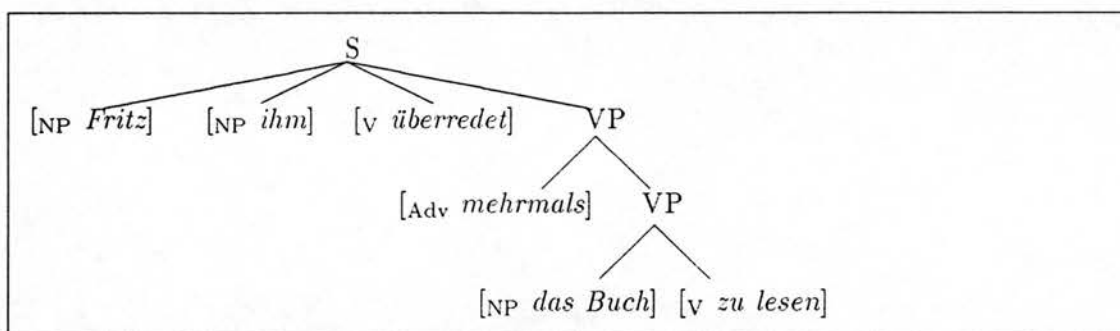


Figure 4-10: Syntax tree of (KN70)

- (KN71) weil der Fritz ihn [v_{VP} das Buch nicht zu lesen] überredet
 because Fritz him the book not to read persuaded
 hat
 has
 (a) 'because Fritz has persuaded him not to read the book'
 (b)*'because Fritz has not persuaded him to read the book'

⁹One reviewer points out that (s)he does not agree with the grammaticality judgements in (KN71b) and (KN72b). This lends further support to the domain union hypothesis for control verbs with object complements.

- (KN72) weil der Fritz ihm [_{VP} das Buch mehrmals zu lesen]
 because Fritz him the book several times to read
 versprochen hat
 promised has
 (a) 'because Fritz has promised him to read the book several times'
 (b)*'because Fritz has several times promised him to read the book'

(KN73) is an example where two different syntactic configurations produce the same word order domain. In (KN73a), the *S* *der Fritz die Maria zu lieben* is modified by *seit langem*. Via obligatory domain union, this produces *der Fritz die Maria seit langem zu lieben*. This becomes the *S* complement of *scheint* and one more step of domain union yields the example. (Cf. Figure 4-11.) In (KN37b), the *S* *der Fritz die Maria zu lieben* is the *S* complement of *scheinen*. Via obligatory domain union, this produces *der Fritz die Maria zu lieben scheint*. This is modified by *seit langem* to yield the example via obligatory domain union. This pattern will occur repeatedly below. (Cf. Figure 4-12).

- (KN73) weil der Fritz die Maria seit langem zu lieben scheint
 because Fritz Maria for a long time to love seems
 (a) 'because Fritz seems to have love Maria for a long time'
 (b) 'because it has seemed for a long time that Fritz loves Maria'

(KN74) is another example of "linearisation" ambiguity. Because of domain union two structures produce the same word order domain. In (KN74a), *nicht* modifies *gewagt* whereas in (KN74b) it modifies *zu lesen*.

- (KN74) weil der Fritz das Buch nicht zu lesen gewagt hat
 because Fritz the book not to read dared has
 (a) 'because Fritz has not dared to read the book'
 (b) 'because Fritz has dared not to read the book'

So, we see that two classes of verbs which allow scrambling, namely, raising verbs and subject control verbs which take no extra complement, also allow scope ambiguities. (KN75) and (KN76) also indicate *lassen* constructions with inherently unaccusative complements or argument-reduced "unannounced passive" complements also allow scope ambiguities. Again, this follows automatically from the assumption that *lassen* domain unions for these complement types.

- (KN75) weil Fritz den Ball mehrmals am Boden auftreffen ließ
 because Fritz the ball several times on the floor bounce let
 (a) 'because Fritz let the ball bounce on the floor several times'
 (b) 'because Fritz several times let the ball bounce on the floor'

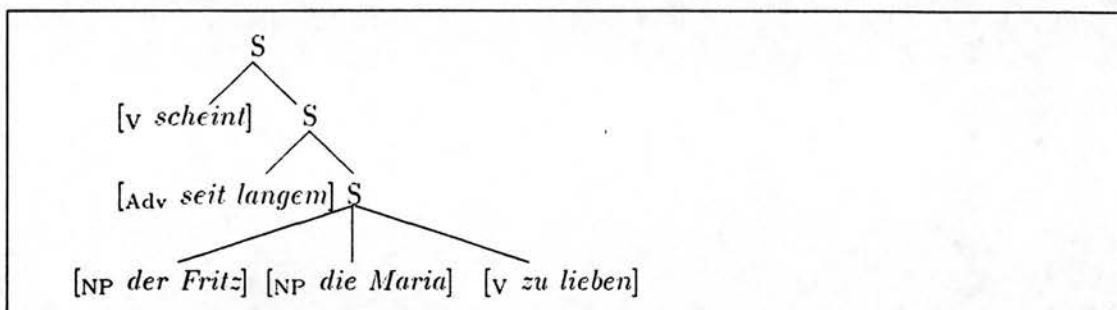


Figure 4-11: Syntax tree of (KN73a)

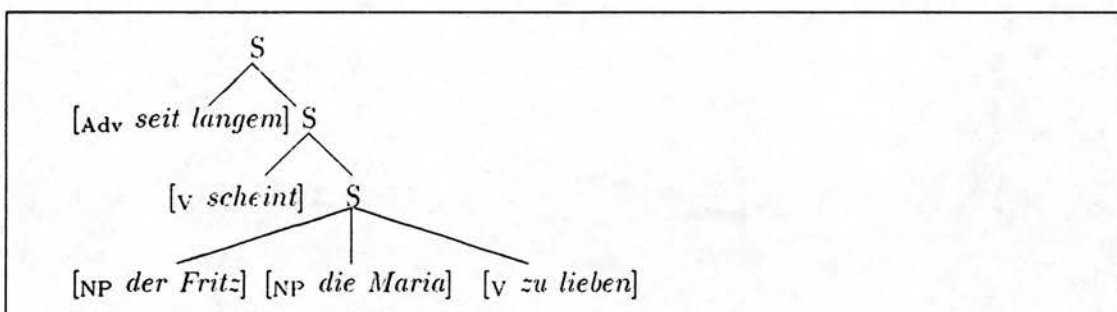


Figure 4-12: Syntax tree of (KN73b)

- (KN76) weil Fritz die Neuigkeit nicht verbreiten ließ
 because Fritz the story not spread let
 (a) 'because Fritz didn't make the news be spread'
 (b) 'because Fritz caused the news not to be spread'

(KN77)-(KN79) show that *lassen* with an "agentive object" also allows both wide and narrow scope of adverbials. This is at odds with the fact that this type of construction appears not to allow scrambling which suggests that *lassen* does not domain union with its VP complement.

- (KN77) weil Fritz ihn den Brief mehrmals vom Original
 because Fritz him the letter several times from the original
 abschreiben ließ
 copy let
 (a) 'because Fritz made him copy the letter several times from
 the original'
 (b) 'because Fritz several times made him copy the letter from
 the original'

- (KN78) weil Fritz ihn den Brief nicht an alle Mitarbeiter schicken
 because Fritz him the letter not to all employees send
 ließ
 let
 (a) 'because Fritz made him send the letter not to all employees'
 (b) 'because Fritz didn't make him send the letter to all
 employees'

- (KN79) weil Fritz ihn seinem Freund nicht helfen ließ
 because Fritz him his friend not help let
 (a) 'because Fritz didn't make him help his friend'
 (b) 'because Fritz made him not help his friend'

This is especially problematic given that (KN77) and (KN78) lend support to the domain union hypothesis, since *mehrmals* in (KN77) and *nicht* in (KN78) are separated from the verb clusters by other constituents. If there was no domain union, then for (KN78) we would expect a domain structure like

[_S weil [_S Fritz ihn [_{VP} den Brief nicht an alle Mitarbeiter schicken]
 ließ]]

Thus, there could be no possibility of explaining the fact that *nicht* can take scope over *ließ* if there is no domain union. Netter notes that this is the crucial question because these constructions pattern with object control verbs with respect to scrambling and reflexivisation but with raising verbs with respect to

extraposition. The second of these concerns is easily dismissed. If *lassen* takes an NP and a VP complement in these constructions (as we have assumed all along), then the VP complement cannot be extraposed because it is not a VP[ZU] but a VP[INF]. On the other hand, if *lassen* takes a clausal complement (contrary to our assumptions) then it cannot be extraposed simply because Ss cannot be extraposed. So this distinction is not of much importance.

Now, there are two possible ways to explain the remaining problem that adverbials can take either wide or narrow scope (in contrast with object control verbs). First, we might assume that *lassen* always obligatorily domain unions its VP complement. We would then look for evidence of scrambling involving *lassen* with an agentive object. We have already seen that *lassen* does indeed allow scrambling sometimes. We might then look for evidence which suggests that *nicht* sometimes takes wide scope over control verbs which take an NP complement.

In fact, there is strong evidence which shows precisely this when prosodic factors are taken into account. Consider the following example (due to Ede Zimmerman) where the verbs in bold face are given heavy stress. The example is perfectly acceptable to most speakers.

- (4.41) weil der Fritz ihm das Buch nicht zu lesen **gebeten** hat
 because the Fritz him the book not to read asked has
 sondern **gezwungen**
 but forced
 ‘because Fritz didn’t ask him to read the book, but forced him
 to’

In this case the heavy stress forces *nicht* to take wide scope over *gebeten* and *gezwungen* rather than narrow scope over *zu lesen*.

An implicit methodological assumption running through this paper is that if there is *one* instance of a grammatical pattern in a language, then the associated syntactic and domain structures are grammatical according to the grammar of the language. There may be strong preferences against such a pattern in general, but if unacceptable examples can be made acceptable by adding contrastive stress or substituting pronouns for phrasal NPs or whatever, then the pattern itself is grammatical. It is then left to the linguist to try to explain why other instances of the pattern are slightly odd, awkward, marginal or completely unacceptable. That is, it is unnecessary to completely conflate grammaticality with acceptability. This point of view is similar to recent work by Chomsky and others to seek explanation for “semi-grammaticality”. It is clear that both processes are at work in natural language.

We hinted at the acceptability issue in the section on scrambling. Clearly in German, there are many so-called “weak word order constraints” over the sequence of NPs and PPs in the Mittelfeld. These constraints are really preferences and are themselves ordered in the sense that some are stronger than others. In general, for any acceptable sequence of NPs and PPs in the Mittelfeld we can probably

find another example (by substitution) with the same pattern which is unacceptable. Under the view that every instance of a pattern must be acceptable for that pattern to be grammatical, this would mean that every Mittelfeld sequence is unacceptable. It is also clear that intonation and stress plays a clear role in the Mittelfeld in altering the preferences of the weak word order effects. We should not be at all surprised then that it plays a role in determining the scope of *nicht* and adverbials in German.

This means that we should not view prosody as a derivational process which takes ungrammatical output syntactic structures and maps them into acceptable structures; rather, prosody is just one of many factors (like case, thematic role, definiteness, etc. with respect to scrambling and governing verb type with respect to scope) which interacts with syntactic and semantic structure to limit the range of interpretations available. Furthermore, the role of prosody in acceptability is wrapped up with issues of topic and focus. It is beyond the scope of this paper to pursue these matters here. Suffice it to say that there is clear evidence that the hypothesis that control verbs which take an object NP complement do indeed domain union in coherent (i.e., nonextraposed and nontopicalised) constructions.

A second *functional* explanation, independent of the first, can also be offered to explain the scope facts of the causative *lassen*. As noted above, *lassen* takes a VP[INF] complement. This means that it cannot be extraposed. Now assume that *lassen* does not domain union its VP complement. Then there is no way to express unambiguously *because Fritz didn't make him help his friend*. In particular, we cannot get wide scope by extraposing the embedded VP (since extraposition of a VP[INF] is impossible). The only type of string that could result would be *weil Fritz ihn nicht [seinem Freund helfen] ließ* with the stated interpretation. I.e., *nicht* has to appear further to the left since the string *helfen nicht ließ* is ungrammatical. However, for any such grammatical variants, we get both scope readings.

What I would like to propose is that if *lassen* is indeed one of the verbs which doesn't domain union, then given the assumptions we have made so far, there is no way to express wide scope. However, this has to be possible in the language. That is, there has to be a way to say the equivalent of *because Fritz didn't make him help his friend*. Therefore, a plausible explanation is to say that semantic interpretation is able to assign either scope regardless of what the actual syntax is for *lassen* constructions with an object NP complement. Basically, what I am suggesting is a possible mismatch between syntactic configuration and semantic compositionality. This possibility is corroborated by the presence of the negative determiner *kein* whose semantics contains a negation and an existential. *kein* is a single morpheme which is the agglutination of the negation element *nicht* plus the indefinite article *ein*.

- (KN82) weil Fritz ihn keinem Menschen helfen ließ
 because Fritz him no man help let
 (a) 'because Fritz didn't let him help anybody'
 (b) 'because Fritz made him help nobody'

Here we see that the implicit negation can take wide scope over the matrix verb. In fact, Netter notes that the wide scope reading is preferred. Clearly, unless a separate level of morphological analysis is provided which decomposes *kein* into *nicht* and *ein*, a lexicalist theory of grammar will have the same problem explaining (KN82) as the account proposed for *lassen* does. That is, one is forced to recognize at some point that scope of negation and adverbials is not entirely determined by syntactic structure. We will return to this below.

To return briefly to a point mentioned above, (AE156) ([13]) is an example from Evers which nicely illustrates the point that the verb cluster cannot be interrupted by *nicht* and also indicates the power of the domain union account to explain “double *nicht*” facts correctly.

- (AE156) a. weil er nicht versuchte das Lied nicht zu singen
 because he not tried the song not to sing
 ‘because he didn’t try not to sing the song’
- b. *weil er das Lied nicht zu singen nicht versuchte

The domain for (AE156a) is

[C_i weil [S er nicht versuchte [VP das Lied nicht zu singen]]]

whereas the domain for the ungrammatical (AE156b) would be

[C_i weil [S er [VP das Lied nicht zu singen] nicht versuchte]]]

and the syntax for both would be

[C_i weil [S er [VP nicht [VP [VP nicht [VP das Lied zu singen]]] versuchte
]]]]

Under a clause union account, (AE156b) would be expected. However, it is clearly ungrammatical as there is no extraposed VP, just the VP *das Lied nicht zu singen* which is almost in “canonical” position, being separated from its governing verb only by *nicht* which takes wide scope over

[VP [VP das Lied nicht zu singen] versuchte]

Descriptively of course, we can see that (AE156b) is out because the verb sequence *zu singen nicht versuchte* is “interrupted” by *nicht*. But this just confirms that

domain union is obligatory here or else structures like (AE156b) would be predicted to be grammatical.¹⁰

As Netter notes, one might try to take the opposite position and claim that scope is determined entirely by semantics. (KN83) and (KN84) might be taken to be examples which support the claim. However, I would analyse both of these examples in terms of syntactic ambiguity with respect to a single word order domain. In (KN83), *nicht* can modify *kommen* giving reading (KN83a) and also modify *darf* giving reading (KN83b). Domain union provides the same word order in either case.

- (KN83) weil er nicht kommen darf
 because he not come may
 (a) 'because he was not allowed to come'
 (b) 'because he was allowed not to come'

(KN84) is a very nice example due to Hubert Haider. Again, three structures give the same word order domain. In (a), *nicht* modifies *muß*, in (b) *können* and in (c) *anstarren*. (Cf. Figures 4-13, 4-14 and 4-15.) The point to realize here is that we are not *interpreting* syntax to get different readings; rather, different syntactic and semantic functor-argument structures produce the same word order domains leading to semantic ambiguity.

- (KN84) weil ein Wachsoldat die Königin nicht anstarren können
 because a guard the queen not stare at can
 muß
 must
 (a) 'because a guard does not have to be able to stare at the
 queen'
 (b) 'because a guard has to be unable to stare at the queen'
 (c) 'because a guard has to be able not to stare at the queen'

There is a potential difficulty with the analysis of reading (a). Because the account here assumes that German clauses do not contain a VP daughter, the syntax of reading (a) would appear to force the article *ein* to be within the scope of the negation *nicht*. This is very likely to be wrong. However, a treatment of quantifier scope is beyond the scope of this thesis.

However, (KN87) and (KN88) provide counterevidence to a purely semantic account of scope. In these two examples, we find the so-called *Ersatzinfinitiv* where

¹⁰(AE156b) is acceptable to some speakers. In this case we would analyze *das Lied nicht zu singen* as a non-domain-unioned VP occurring to the left of the verb cluster *nicht versuchte*. This is made more apparent by the increased acceptability of the example *weil er das Lied nicht zu singen nicht einmal versuchte* where the presence of the adverbial *einmal* clearly indicates that the sequence *nicht zu singen nicht einmal versuchte* is not a single verb cluster. For further discussion, cf. the next section.

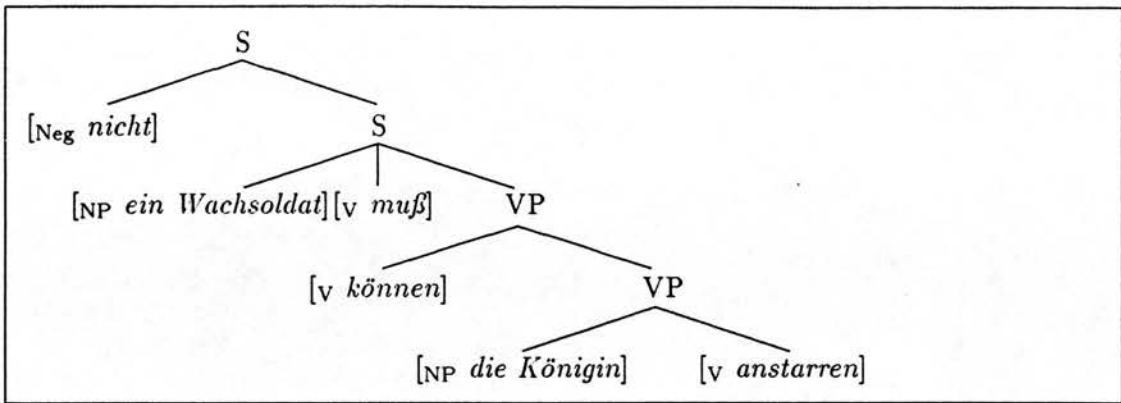


Figure 4-13: Syntax tree of (KN84a)

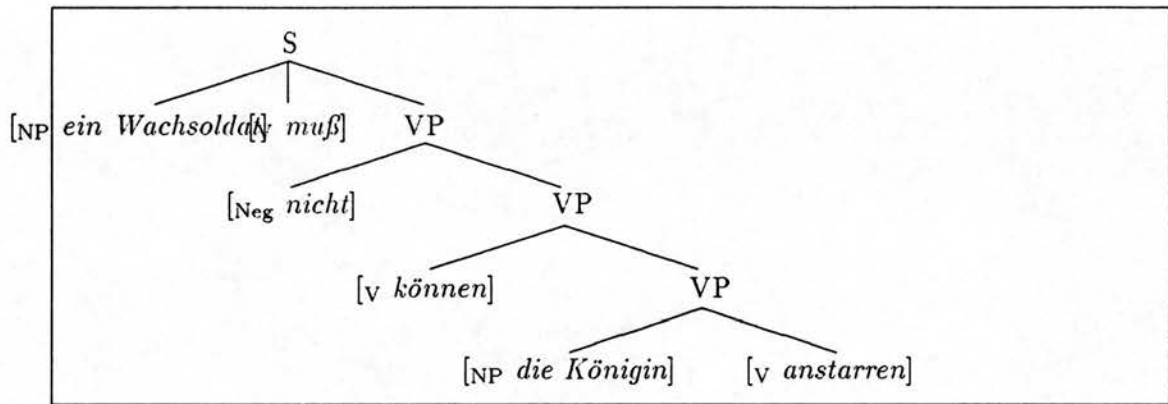


Figure 4-14: Syntax tree of (KN84b)

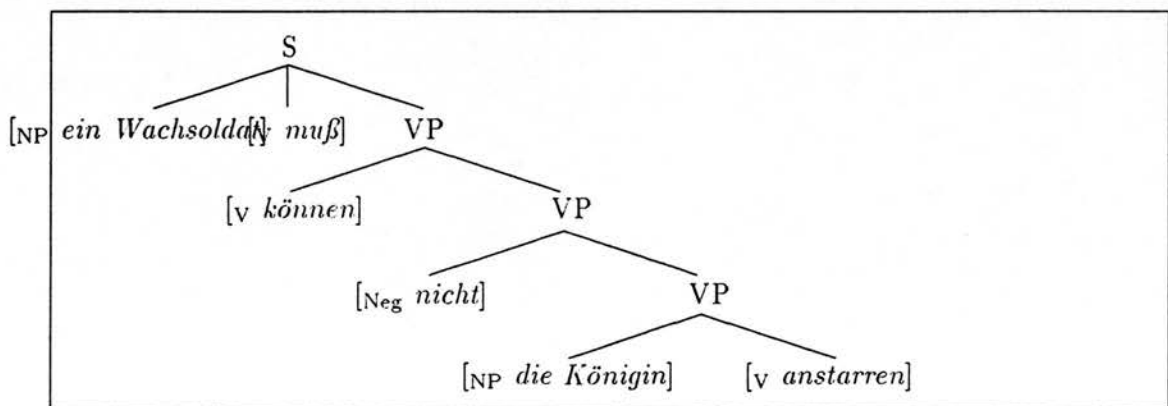


Figure 4-15: Syntax tree of (KN84c)

the finite verbs *hat* and *muß* are the first verb in the verb cluster and the past participles expected to be governed by *hat* and *muß* are replaced by the infinitives *dürfen* and *können*. Therefore, the government order is 1-3-2 instead of 3-2-1. But only the wide scope reading is available so it seems that a purely semantic account is impossible.

(KN87) weil er nicht hat kommen dürfen
 because he not has come may
 (a) 'because he was not allowed to come'
 (b)*'because he was allowed not to come'

(KN88) weil ein Wachsoldat die Königin nicht muß anstarren
 because a guard the queen not must stare at
 können
 can
 (a) 'because a guard does not have to be able to stare at the
 queen'
 (b)*'because a guard has to be unable to stare at the queen'
 (c)*'because a guard has to be able not to stare at the queen'

Unfortunately, I do not see how to capture (KN87) and (KN88) naturally in the account given. The answer which suggest itself is to invoke an LP constraint of the form $DTRS|HEAD-DTR|V \preceq NEG$ where *NEG* is an abbreviation for the category of the negation element *nicht* (whatever it may happen to be). That is, the fronted auxiliary or modal in the *Ersatzinfinitiv* construction would specify that it has to precede *nicht*. The idea is that a fronted auxiliary or modal must precede not only any verbs which it governs but also any adverbials and negation elements which it governs.

If we then tried to derive the narrower scope readings, e.g., (KN87b), we would have to hypothesize that *nicht* modified *kommen*. But then our assumption about *nicht* would require that it appear before any of the verbs in the verb cluster, i.e., to the left of the auxiliary *hat*. But *hat* requires that *nicht* appear to the right of it and so the structure is not welldefined. Similar examples can be found with four and five verbs but wide scope is always the only interpretation available. The crucial question is whether the same generalisation holds for adverbials in general or whether this is limited to *nicht*.

I readily agree that this "solution" is rather ad hoc. There is evidence from Dutch however which lends support to this idea. If a verb cluster is of the form $V[FIN]_1 V_2 V_3$, as in example (4.42), then *niet* may take wide scope.

(4.42) omdat een lijfwacht de koningin niet hoeft te kunnen
 because a guard the queen not has to can
 aanstaren
 stare at
 'because a guard does not have to be able to stare at the queen'

It may also take narrow scope with respect to the second verb as in (4.43).

- (4.43) omdat een lijfwacht de koningin niet mag kunnen aanstaren
 because a guard the queen not must can stare at
 'because a guard must be unable to stare at the queen'

However, surprisingly it may not take narrow scope with respect to the third verb as in (4.44). (The Dutch word corresponding to *must* in English must change to maintain the same meaning.)

- (4.44) omdat een lijfwacht de koningin niet moet kunnen aanstaren
 because a guard the queen not must can stare at
 *'because a guard has to be able to not stare at the queen'

What is even more surprising is that (4.44) can be partially repaired if *niet* is placed immediately before the third verb as in (4.45).

- (4.45) omdat een lijfwacht de koningin moet kunnen niet aanstaren
 because a guard the queen not must can stare at
 ?'because a guard has to be able to not stare at the queen'

Clearly, *aanstaren* is part of the verb cluster and is not extraposed since the clitic *te* does not immediately precede it. It is also unlikely that we could justify it on the grounds that *niet aanstaren* is an instance of verb projection raising since the direct object *de koningin* occurs to the left of the verb cluster. At least some speakers find (4.45) interpretable with the reading given. Although it is admittedly rather awkward, it is surprising that it is acceptable at all in light of the normal assumptions made about verb clusters. We will return to this point momentarily.

Examples (TH14a-c) (due to Tilman Höhle) indicate that although Dutch has a "normal" government order of verbs in Dutch of 1-2-3... , the negative determiner *geen* can take either wide or narrow scope over the verbs in the verb cluster.

- (TH14) a. dat hij geen fouten lijkt te maken
 that he no faults seems to make
 (a) 'that he doesn't seem to make any mistakes'
 (b) 'that he seems not to make any mistakes'
- b. dat Jan geen vlees wil eten
 that Jan no meat wants eat
 (a) 'that Jan doesn't want to eat any meat'
 (b) 'that Jan wants to eat no meat (wants to not eat any meat)'
- c. dat Jan geen fouten heeft proberen te maken
 that Jan no faults has try to make
 (a) 'that Jan hasn't tried to make any mistakes'
 (b) 'that Jan tried to make no mistakes'

(TH14c) is especially convincing in that the two interpretations are very different. Notice furthermore, that in (TH14c) we have an instance of the Dutch variant of the *Ersatzinfinitiv* where the bare infinitive *proberen* replaces the past participle *geprobeerd* and it is still the case that both the wide and narrow scope readings are available. This suggests that it is the “movement” of the finite verb which eliminates the narrower scope readings in German and not the form of the verb governed by the finite verb.

- (4.46)
- a. dat hij fouten niet lijkt te maken
that he faults not seems to make
(a) ‘that he doesn’t seem to make mistakes’
(b) ‘that he seems to not make mistakes’
 - b. dat Jan vlees niet wil eten
that Jan meat not wants eat
(a) ‘that Jan doesn’t want to eat meat’
(b) ‘that Jan wants to not eat meat’
 - c. dat Jan fouten niet heeft proberen te maken
that Jan faults not has try to make
(a) ‘that Jan hasn’t tried to make mistakes’
(b) ‘that Jan tried not to make mistakes’

These variants of (TH14a-c) are given to indicate that it is not just the determiner *geen* which allows for both narrow and wide scope readings. We see in (4.46a-c) that the negation *niet* takes both wide and narrow scope over the verb cluster, even in the *Ersatzinfinitiv* construction.

The data above suggest that Dutch is not as strict as German in determining the scope of negation. Once more, a functional argument can be made here. Since the “normal” government order of verbs in Dutch is 1-2-3, if the scope of negation was required to be wide with respect to verb 1, then there would be no way to express the narrow scope readings with respect to verbs 2 and 3 (assuming domain union and canonical Dutch verb order). Therefore, all the scope readings with respect to the verbs in the verb cluster must be allowed. In the one case where scope over the third verb is impossible (4.44), we find that the clause giving the appropriate interpretation can be partially repaired by placing the negation directly in front of the third verb. Otherwise, when the negative element is *niet*, then it must simply appear to the left of any verbs in its domain even when there is an *Ersatzinfinitiv* construction (in contrast to German where only the wide scope is available). Thus the major difference between the grammars of the two languages is that the LP constraint $\text{DTRS|HEAD-DTR|V} \preceq \text{NEG}$ is required for German *Ersatzinfinitiv* constructions but not for Dutch. This is a very minor difference between the two grammars but does seem to account for most of the phenomena discussed here.

There is one other point to be made about this seemingly ad hoc ordering constraint of German. The ordering constraint $\text{DTRS|HEAD-DTR|V} \preceq \text{v}$ is certainly

motivated in cases of finite verb inversion. This established the inverted pattern. Now from a categorial unification grammar point of view, the MAJ feature of both *nicht* and adverbials is in fact *v*. Therefore, the ad hoc constraint $DTRS|HEAD-DTR|v \preceq NEG$ may simply be an instance of the more general constraint $DTRS|HEAD-DTR|v \preceq v$.

I only intend this to be a preliminary sketch of how the analysis of scope might be rescued when it is clear that a constituent has been domain unioned but only the wide scope is available. In any event, it is well known that these types of “inversions” have unusual behaviour in both German and Dutch. Further investigation is beyond the scope of this paper.

The crucial question for this section is whether there are control verbs which take an object complement and exhibit scope variations but exhibit absolutely no scrambling whatsoever or whether there are control verbs which take an object complement and exhibit scrambling but exhibits no scope variation whatsoever. If either of these two types of verbs exist, then the analysis proposed here faces serious difficulties. Whether such verbs exist is an open question. However, given the amount of counterevidence to the generalisations that only narrow scope is available in control verb constructions with an object complement and that scrambling is unavailable in control verb constructions with an object complement, it seems unlikely that this is the case.

4-9 Topicalization and Extraposition

Topicalization in German is rather notorious for apparently allowing nonconstituents to appear in topic position. Examples (KN92)-(KN96) illustrate some of the simpler possibilities. (The brackets delimit the fronted topic.)

(KN92) [einen Freund vorgestellt] hat er ihr noch nie
 a friend introduced has he her (DAT) yet never
 ‘He hasn’t introduced a friend to her yet’

(KN93) [ihr vorgestellt] hat er seine Freunde noch nie
 her (DAT) introduced has he his friends yet never
 ‘He hasn’t introduced his friends to her yet’

(KN94) [vorgestellt] hat er ihr seine Freunde noch nie
 introduced has he her (DAT) his friends yet never
 ‘He hasn’t introduced his friends to her yet’

(KN95) [zu lesen versucht] hat er das Buch nicht
 to read tried has he the book not
 ‘He didn’t try to read the book’

- (KN96) [einen Freund vorzustellen versucht] hat er ihr noch
 a friend to introduce tried has he her (DAT) yet
 nie
 never
 'He hasn't tried to introduce a friend to her yet'

Following den Besten and others, I take it that such cases of "partial VPs" in topic position are the result of a clause bounded movement process of one or more complements (of one or more nested VPs) into a governing verb projection. (Strictly speaking the bounding category is $\bar{S}(= \bar{C})$). This type of movement is distinct from the type of unbounded movement found in *wh*-movement and the effects of domain union (or scrambling) as we will see. What is left after movement can be called the "remnant VP". It can then be topicalized.¹¹ We will also see that the same process is involved in extraposed VPs.

This means that the following syntactic analyses are assigned to (KN92)-(KN94)

- (KN92') [einen Freund vorgestellt] hat er ihr noch nie
 [CP [VP e_i einen Freund vorgestellt]_j [S noch nie [S hat er ihr_i e_j]]]
- (KN93') [ihr vorgestellt] hat er seine Freunde noch nie
 [CP [VP ihr e_i vorgestellt]_j [S noch nie [S hat er seine Freunde_i e_j]]]
- (KN94') [vorgestellt] hat er ihr seine Freunde noch nie
 [CP [VP e_i e_j vorgestellt]_k [S noch nie [S hat er ihr_i seine Freunde_j e_k]]]

In these three examples, we see that either or both of the complements of *vorgestellt* can be "left behind" in the Mittelfeld. We can already see some evidence that such "raised"¹² constituents are ordered in the domain they occur in as though they were introduced to the domain via one of the normal domain construction mechanisms.¹³ For example, **NOM** < **DAT** < **ACC** and **+PRONOMINAL** < **-PRONOMINAL**.

- (KN95') [zu lesen versucht] hat er das Buch nicht
 [CP [VP [VP e_i zu lesen] versucht]_j [S nicht [S hat er [NP das Buch]_i e_j]]]

¹¹Of course, I mean "movement" only in a figurative sense. Apparently "dislocated" material is base generated in situ.

¹²Raising here is not to be confused with raising verbs or argument inheritance.

¹³This suggests that the proper analysis is to accomplish such movement directly between domains. This would allow the simple functor-argument structure of both syntax and semantics to be maintained without additional complications. However, I will proceed in this chapter as if raising is a syntactic phenomena. Once again, empirical evidence is needed to decide on this issue.

(KN96') [einen Freund vorzustellen versucht] hat er ihr noch nie

$$\begin{array}{l} \text{[CP [VP [VP } e_i \text{ einen Freund vorzustellen] versucht]}_k \\ \quad \text{[S noch nie} \\ \quad \quad \text{[S hat er ihr}_i \text{ e}_k\text{]]}] \end{array}$$

It is also possible to front a verbal head with the nonfinite VP complement with none (KN95') or some (KN96') of the nonfinite verb's complements. Furthermore, we also find examples where there is an extraposed VP inside the topicalized VP in topic position with none or some of the extraposed VP's head verb's complements. (KN97) and (KN98) are exactly analogous to (KN95) and (KN96) except that the embedded "remnant VP" in topic position is extraposed.

(KN97) [versucht, zu lesen], hat er das Buch nicht
 tried, to read, has he the book not
 'He hasn't tried to read the book'

(KN98) [versucht, einen Freund vorzustellen], hat er ihr noch
 tried, a friend to introduce, has he her (DAT) yet
 nie
 never
 'He hasn't tried to introduce a friend to her yet'

According to the account offered here, (KN95) and (KN97) have the same syntactic structure and (KN96) and (KN98) have the same syntactic structure. The difference between the two pairs is purely to do with the way that the word order domains are constructed. The syntactic structure of the topic of (KN95) and (KN97) will be

(4.47) [VP₁ [VP₂ e_i zu lesen₂] versucht₁]

and the syntactic structure of the topic of (KN96) and (KN98) will be

(4.48) [VP₁ [VP₂ e_i einen Freund vorzustellen₂] versucht₁]

That is the syntactic structure of both topics is of the form

(4.49) [VP₁ [VP₂ ... V[ZU]₂] versucht₁]

where "..." indicates a possibly empty string of NP complements and adverbials. So, according to the account so far, there are two possibilities. Either VP₂ can be domain unioned with VP₁ giving (KN95) and (KN96) or VP₂ can be extraposed

within VP_1 's domain giving (KN97) and (KN98). This is a very good illustration of the interaction of raising and our treatment of VP extraposition.

It is also well established by several authors that the same interaction (although not so extreme) occurs in clause final extraposed VPs. It's my belief that these are the same processes. Therefore, for the sake of this discussion, I'll refer to both of these processes as *focus raising* following Uszkoreit ([45]) or simply just *raising* as above. (KN99) and (KN100) are extraposed and domain unioned versions of the same structure where *ihr einen Freund vorzustellen* is the complement of *versucht*. However, in (KN101) we see that the indirect object *ihr* has been "raised to" the Mittelfeld while the rest of the VP *einen Freund vorzustellen* has been extraposed.

(KN99) er hat noch nie versucht, ihr einen Freund vorzustellen
 he has yet never tried, her a friend introduce

(KN100) er hat ihr noch nie einen Freund vorzustellen versucht
 he has her yet never a friend introduce tried

(KN101) er hat ihr noch nie versucht, einen Freund vorzustellen
 he has her yet never tried, a friend introduce

Assuming the following syntactic structure for (KN101),

(4.50) er hat ihr noch nie versucht, einen Freund vorzustellen

$$\begin{array}{l} [CP\ er_i \\ \quad [S\ e_i\ ihr_j\ hat \\ \quad \quad [VP_1\ noch\ nie \\ \quad \quad \quad [VP_2\ versucht_2 \\ \quad \quad \quad \quad [VP_3\ e_j\ einen\ Freund\ vorzustellen_3]]]]] \end{array}$$

then the domain of VP_3 can be extraposed in VP_2 's domain as follows.

(4.51) $[VP_2\ versucht_2, [VP_3\ einen\ Freund\ vorzustellen_3]]$

What is even more surprising however is that it is possible to front an object control verb plus the verb it governs without any of its complements as in (KN102) and (KN103).

(KN102) [zu lesen gebeten] hat er ihn es nicht
 to read asked has he him it not
 'He didn't ask him to read it'

- (KN103) [zu lesen empfohlen] hat er ihm alle die Bücher,
 to read recommended has he him (DAT) all those books
 die ...
 which ...
 'He recommended to him to read all those books which ...'

This data adds additional support to the hypothesis that remnant VP topicalisation (and extraposition) is a different process than that involved in either scrambling or unbounded movement. If object control verbs are assumed not to undergo verb raising, clause union or domain union then these examples would not be deriveable without postulating a third type of movement. If domain union is assumed, then there must also be a second type of movement (in addition to unbounded movement) since otherwise the complements of *zu lesen* in both examples would have to appear in topic position along with their head verb. This examination of verb projection topicalisation is meant to be no more than an examination of certain topicalization possibilities. Certainly there are other phenomena such as multiple PP fronting, so-called "split topicalisation" of NPs and separable prefix fronting, to name but a few, which complicate the picture. However, here we content ourselves with examining the nature of verb projection topicalisation and to show, as Netter points out, that what appears in topic position is not a "more or less meaningless rearrangement of substrings, heads and complements"¹⁴

There is other data which also supports the raising hypothesis.

- (4.52) *[zu schenken scheinen] sollte Peter ihr dem Jungen
 to give seem should Peter her (DAT) the boy (DAT)
 das Buch
 the book (ACC)
 'Peter should seem to her to give the boy the book'

The problem here is that *zu schenken scheinen* is not a VP (even with traces). If we consider the related subordinate clause *daß Peter ihr scheinen sollte, dem Jungen das Buch zu schenken* we see that the syntactic structure of both should be

- (4.53) [_T daß
 [_{S[FIN]} sollte
 [_{S[INF]} ihr scheinen
 [_{S[ZU]} Peter dem Jungen das Buch zu schenken]]]]

If this analysis is correct, then *sollte* is a one-place predicate both syntactically and semantically. (Given it's epistemic reading in this sentence, this seems like a valid assumption.) In this case, *scheinen* in addition to taking it's usual S[ZU]

¹⁴Quoted from [28].

complement also takes an optional dative NP complement. Finally, *Peter dem Jungen das Buch zu schenken* is the S[ZU] complement of *scheinen*. The problem is that to derive (4.52) we would have to raise the subject *Peter* in addition to the direct and indirect objects. But this apparently is impossible. What we should expect to be able to do is derive a sentence like (4.54).

- (4.54) **[Peter zu schenken scheinen] sollte ihr dem Jungen*
 Peter to give seem should her (DAT) the boy (DAT)
das Buch
 the book (ACC)
 'It should seem to her that Peter give the boy the book'

But this is unacceptable as well. Precisely why is unclear to me. However, the pattern of a subject plus a nonfinite verb governed by a raising verb is in fact possible as long as the nonfinite verb is unaccusative.

(KN104) and (KN105) show that an unaccusative nonfinite verb can be fronted with its subject. However, as Netter points out, the governing verb must have a raising reading available and, furthermore, in this construction only the raising reading is possible. But this is predicted by our assumptions. In (KN104) and (KN105), *muß* and *dürfen* must subcategorise for a bare infinitival S, otherwise it could not be topicalised. But we already know that both *müssen* and *dürfen* subcategorise for bare infinitival Ss under their raising reading. Note also that the adverbial *öfters* and the negation element *nicht* only take wide scope over the finite modal raising verb in agreement with our earlier observations concerning the interaction of adverbials, *nicht* and topicalised verb projections.

- (KN104) ein Häftling entspringen muß hier öfters
 an prisoner escape must here frequently
 (a) 'it must be frequently the case, that a prisoner escapes'
 (b)*'a prisoner frequently has to escape here'

- (KN105) ein Häftling entspringen darf hier nicht
 a prisoner escape may here not
 (a) 'it must not be the case that a prisoner escapes here'
 (b)*'a prisoner does not have permission to escape here'

Notice furthermore that this is also possible with a plural subject with the subject agreeing with the finite verb. (Cf. the discussion in §4-4).

- (KN105') Häftlinge entspringen dürfen hier nicht
 prisoners escape may here not
 (a) 'it must not be the case that prisoners escape here'
 (b)*'prisoner do not have permission to escape here'

Finally, to emphasize the point that examples like (KN104), (KN105) and (KN105') compare examples (4.55a) and (4.55b).

- (4.55) a. [Ein Außenseiter gewinnen] dürfte eigentlich nicht
 an outsider win should actually not
 'An outside shouldn't actually win'
- b. *[Ein Außenseiter gewinnen] wollte schon immer
 an outsider win wanted already always
 'An outside always wanted to win'

The essential difference between the two examples is that the finite verb in (4.55a) is clearly a raising verb whereas the finite verb in (4.55b) is a control verb.

If one accepts the analysis just discussed for such topicalised nonfinite clauses, then the real question is why this phenomena is limited to unaccusative verbs. There is a body of relevant literature (cf. [16] and [46] for overviews) but speaker judgements vary greatly. In any case, further investigation is beyond the scope of this chapter. However, I would like to note that other cases of "subject plus nonfinite verb topicalisation" need not necessarily be explained in terms of this analysis. I.e., my claim is that (KN104) and (KN105) are possible precisely because of the subcategorisation requirements of the epistemic readings of *müssen* and *dürfen* and nothing more. Explanations for the presence or absence of the "definiteness effect" (cf. [26]) or the fact that *lassen* sometimes allows such constructions for some speakers (cf. [46]), may be explainable in terms of interaction of other aspects of grammar.

(KN106)-(KN108) show some data similar to that of the scope of negation section indicating that *nicht* in the postverbal field can take scope over the finite verb or any verb in the verb cluster but not over a verb in topic position. The converse is also true. If *nicht* occurs in a topic VP then it cannot take wide scope over the finite verb or any of the verbs in the main verb cluster. The converse follows immediately from our assumptions since there is no way that *nicht* could end up in topic position if it modified one of the other verbs in syntactic structure.

- (KN106) [die Königin anstarren] kann er nicht
 the queen stare at can he not
 (a) 'he is unable to stare at the queen'
 (b)*'he is able not to stare at the queen'

- (KN107) [anstarren] muß er die Königin nicht können
 stare at must he the queen not can
 (a) 'he doesn't need to be able to stare at the queen'
 (b) 'he must be unable to stare at the queen'
 (c)*'he must be able not to stare at the queen'

- (KN108) [anstarren können] muß er die Königin nicht
 stare at can must he the queen not
 (a) 'he doesn't need to be able to stare at the queen'
 (b)*'he must be unable to stare at the queen'
 (c)*'he must be able not to stare at the queen'

Unfortunately, I can see no straightforward way to predict the generalisation that *nicht* in the postverbal field can take scope over the finite verb or any verb in the verb cluster but not over a verb in topic position. So far I have said nothing which would prevent the argument of *nicht* from being extracted to topic position. Short of stipulation, I see nothing which blocks this at the moment. This is one point where the account very much depends on one's theory of unbounded dependencies. For example, we have required that *nicht* obligatorily domain unions with its argument's domain. If we assume a trace based approach to extraction however, this domain will simply be the empty domain and union will just return *nicht*'s domain. One possible solution then is to require that *nicht* cannot have an empty complement domain. Further work would be necessary to determine whether this was empirically correct. For example if *nicht* were a clitic like *n't* in English then this would follow automatically.

On the other hand, if gap introduction is an operation on argument lists, then perhaps the failure to strand *nicht* is a consequence of the account. For example, one might assume that only heads allow their arguments to be extracted thus preventing stranding of determiners, *nicht*, modifiers and specifiers in general. Again, more empirical investigation must be done before the issue can be decided.

Finally, I would like to turn to an interesting consequence of the interaction of the treatments of scope of negation, extraposition and raising.

- (4.56) er hat das Buch versucht zu lesen
 he has the book tried to read
 'he has tried to read the book'

This type of example involving two verbs, one of which is a V[ZU] following the verb that governs it while its complement(s) appear to the left of the verb cluster is very common. The question is whether the VP *das Buch zu lesen* has been domain unioned and *versucht zu lesen* is simply an alternative ordering possibility or whether *das Buch* has been raised to the finite clause and the remnant VP *zu lesen* extraposed clause-finally. These examples have always been slightly troublesome since the account presented here would effectively have to stipulate that *versucht zu lesen* was an alternative order if the VP[ZU] was domain unioned without any additional, independent empirical support. However, consider (4.57) and (4.58).

- (4.57) er hat das Buch nicht zu lesen versucht
 he has the book not to read tried
 (a) 'he hasn't tried to read the book'
 (b) 'he has tried not to read the book'

- (4.58) er hat das Buch nicht versucht zu lesen
 he has the book not tried to read
 (a) 'he hasn't tried to read the book'
 (b)*'he has tried not to read the book'

In (4.57), *versucht* is not a control verb with an object NP complement so we know that the VP *das Buch zu lesen* has been domain unioned. Thus we get both the wide and narrow scope readings for *nicht*. However, in (4.58), we get only the wide reading. Therefore, *zu lesen* must be an extraposed VP since if it was domain unioned we would get both narrow and wide scope as in (4.57). Since *das Buch* is the object of *zu lesen* it must have been raised to the finite clause. So, with the help of the very robust generalisation about extraposed VPs and the scope of *nicht* we are able to establish that examples like (4.56) are instances of raising plus extraposition. This is the only case where an intonation break cannot optionally be inserted before an extraposed VP. (It is also the only case where German orthography does not require a comma before an extraposed VP.) Undoubtedly, this is because a single extraposed verb is very "light" prosodically. Notice that the addition of just the object clitic *es* is enough to optionally allow the intonation break for most speakers (given intonational support).

- (4.59) er hat versucht es zu lesen
 he has tried it to read
 'he has tried to read it'

Notice however that even with the clitic *es*, most speakers accept the following form of raising.

- (4.60) er hat es versucht zu lesen
 he has it tried to read
 'he has tried to read it'

That *es* can be raised to the finite clause is surprising since we might expect that elements which are essentially clitics cannot escape from their domain. However, this seems to cause no difficulty for the majority of German speakers. In passing, I should note that clause final sequences of the type *versucht zu lesen* do not appear to involve extraposition in Zurich German. Apparently, both wide and narrow scope readings are available for similar examples (Kathrin Cooper, p.c.).

4-10 Intraposition of VPs in the Mittelfeld

In this section we will examine VPs *intraposed* in the Mittelfeld. Such VPs are called intraposed because they occur in the Mittelfeld separated from the verb cluster by other intervening constituents. This is also called intraposition in analogy with the term *extraposition* although extraposition is only a descriptive term to us since extraposed VPs are actually in VP domain internal position and not external at all.

4-10.1 The Empirical Facts

Consider the following examples.

(4.61) daß er₁ sie₂ ihn₃ wiederzusehen₃ gebeten₂ hat₁
 that he her him to see again asked has
 ‘that he has asked her to see him again’

(4.62) daß ihn₂ niemand₁ wiederzusehen₂ hoffte₁
 that him nobody to see again hoped
 ‘that nobody hoped to see him again’

(4.63) daß es₃ ihn₂ niemand₁ zu lesen₃ gebeten₂ hat₁
 that it him nobody to read asked has
 ‘that nobody has asked him to read it’

According to one view of German clause structure, the substring *ihn wiederzusehen* in example (4.61) is a VP situated in the Mittelfeld, giving a bracketing such as that in (4.64a). Word order variations amongst Mittelfeld NPs, as in (4.62) and (4.63) where object NPs from the VP precede an indefinite pronominal subject, are taken to arise by local movement of NPs out of the VP to be adjoined to VP or S, so that (4.63) would be analysed as in (4.64b).

(4.64) a. daß [S er₁ [VP sie₂ [VP ihn₃ wiederzusehen₃] gebeten₂] hat₁]
 b. daß [S es₃ ihn₂ [S niemand₁ [VP [VP t₂ [VP t₃ zu lesen₃] gebeten₂] hat₁]]]

However, consider the following examples:

(4.65) daß er nur am Samstag Bücher zu lesen versucht hat
 that he only on Sunday books to read tried has
 (a) ‘that he tried to read books only on Sunday’ (b) ‘that he
 only tried to read books on Sunday’

(4.66) daß er₁ [VP Bücher zu lesen] [PP nur am Samstag] versucht hat₁
 that he books to read only on Sunday tried has
 ‘that he has tried to read books only on Sunday’

In (4.65), the PP adjunct *nur am Samstag* can take scope over either *Bücher zu lesen* or over the finite verb *versucht*. This scope ambiguity is eliminated in (4.66) where the PP can only modify *versucht*. In (4.66) the PP adjunct *nur am Samstag* intervenes between the verb *versucht* and its VP complement *Bücher zu lesen*, i.e., so that the two verbs *versucht* and *zu lesen* are no longer (string)

adjacent. The presence of such intervening non-verbal material appears to have a number of consequences. Examples (4.67) and (4.68) show that this VP is a barrier to 'local' movement. Local movement to either the left or right out of the VP is impossible. Note that (4.67) is not ungrammatical simply because a VP precedes an NP, as example (4.69) shows. (4.70) shows that such constructions are ungrammatical even when the scrambled NP is a pronoun. (The latter example is important for showing that even the Wackernagel position cannot license scrambling of pronominal NPs out of the VP.)

(4.67) *daß er₁ [VP t₂ zu lesen] Bücher₂ [PP nur am Samstag] versucht hat₁

(4.68) *daß Bücher₂ er₁ [VP t₂ zu lesen] [PP nur am Samstag] versucht hat₁

(4.69) daß er [VP diese Bücher zu lesen] nur der Maria versprach
that he this book to read only to Maria promised
'that he promised to read this book only to Maria'

(4.70) *daß es₂ niemand₁ [VP t₂ zu lesen] trotzdem versuchte₁
that it nobody to read nevertheless tried
'that nevertheless nobody tried to read it'

Note however that although such "intraposed" VPs are barriers to local 'scrambling' movement, they are not barriers to wh-extraction, as examples (4.71) and (4.72) show.¹⁵

(4.71) Bücher₂ die er₁ [VP t₂ zu lesen] [PP nur am Samstag]
books that he to read only on Sunday
versucht hat₁
tried has
'books that he has tried to read only on Sunday'

(4.72) Was₂ hat₁ er₁ [VP t₂ zu lesen₂] [PP nur am Samstag]
What has he to read only on Sunday
versucht?
tried
'What he has tried to read only on Sundays?'

This data presents a problem for accounts that assume VP bracketing in coherent constructions (as in (4.61)) since they fail to distinguish coherent and intraposed constructions and so provide no explanation for why intraposed VPs are barriers

¹⁵Some speakers do not tolerate unbounded extraction as in these examples.

to local movement. Note that a simple barrier to movement account of (4.67) and (4.68) will not explain the asymmetry with respect to unbounded movement and Mittelfeld 'scrambling'. Several accounts of this data conflate local and unbounded movement with the result that they either overgenerate or undergenerate uniformly. This includes GB scrambling analyses of all types and also LFG analyses based on functional uncertainty.

Finally, here are three additional examples involving scrambled nonlexical NPs which one would normally expect to be grammatical. Notice in particular, that the descriptive ordering generalisation that definite NPs precede indefinite NPs in the Mittelfeld is obeyed.

- (4.73) *daß [die Bücher]₂ [ein Mann]₁ [v_P t₂ zu lesen]
 that the book a man to read
 [p_P nur am Samstag] versucht hat₁
 only on Sunday tried has
 'that a man has tried to read the book only on Sunday'

- (4.74) *daß sie₃ [das Buch]₂ [ein Mann]₁ [v_P t₂ zu lesen]
 that her the book a man to read
 [p_P nur am Samstag] t₃ überredete hat₁
 only on Sunday persuaded has
 'that a man persuaded her to read the book only on Sunday'

- (4.75) *daß ihn₃ [das Buch]₂ [ein Mann]₁ [v_P t₂ zu lesen]
 that him the book a man to read
 [p_P im Park] t₃ gebeten hat₁
 in park asked has
 'that a man asked him to read the book in the park'

4-10.2 A Domain Union Analysis

A domain union analysis for the problem of intraposed VPs in the Mittelfeld is truly trivial. VPs are either domain unioned or not. If domain unioned, the VP forms a verb cluster with the finite verb in subordinate clauses. If it is not domain unioned, then it can either appear to the left or to the right of the finite verb. If it occurs to the right, it appears in extraposed position. If it appears to the left then it can appear in intraposed position. I.e., the LP constraints allow the VP to appear to the left of nonpronominal NPs and PPs. (I have no evidence and doubt highly that intraposed VPs can appear to the left of pronominal NPs in the Mittelfeld.) Since an intraposed VP is not domain unioned, the head verb of the intraposed VP and its complements must be adjacent, i.e., scrambling is not allowed as the data indicates. Let us consider the analyses of the preceding data then.

The syntactic structure of (4.66) is (4.76).

- (4.76) [s [c daß] [s er [vP [pP nur am Samstag] [vP [vP Bücher zu lesen] versucht]] hat]]

The domain structure of *Bücher zu lesen versucht* is (4.77).

- (4.77) [vP [vP Bücher zu lesen] versucht]

That is, the VP *Bücher zu lesen* is an element of it's mother's domain and is not domain unioned into it as described before. The domain of *Bücher zu lesen nur am Samstag versucht* is (4.78).

- (4.78) [vP [vP Bücher zu lesen] [pP nur am Samstag] versucht]

The VP *Bücher zu lesen versucht* is domain unioned into its mother's domain allowing the adjunct PP to occur between the head verb *versucht* and the non-domain-unioned VP *Bücher zu lesen*. The domain of *er Bücher zu lesen nur am Samstag versucht hat* is (4.79).

- (4.79) [s er₁ [vP Bücher zu lesen]₂ [pP nur am Samstag]₂ versucht₂ hat₁]

That is, the domain of the VP *Bücher zu lesen nur am Samstag versucht* is domain unioned into the finite clause's domain, allowing the complement VP *Bücher zu lesen*, the adjunct PP *nur am Samstag* and the verb *versucht* to be elements of the finite clause's domain. *Bücher zu lesen* then appears to the left of the adjunct PP *nur am Samstag*.

The domain of (4.69) is (4.80).

- (4.80) [s er₁ [vP diese Bücher zu lesen]₁ [NP nur der Maria]₁ versprach₁]

This time all of the complements occur non-domain-unioned in the finite clause's domain and the VP *diese Bücher zu lesen* appears to the left of the indirect object complement *nur der Maria* of the finite verb *versprach*.

The domain of (4.71) is (4.81).

- (4.81) [NP Bücher₃ [S[REL] [NP die] [s er₁ [vP t₃ zu lesen]₂ [pP nur am Samstag]₂ versucht₂ hat₁]]]]

This is analogous to (4.79).

The domain of (4.72) is (4.82).

$$(4.82) \quad [s [NP \text{Wat}_3] [s \text{hat}_1 \text{er}_1 [VP \text{t}_3 \text{zu lesen}]_2 [PP \text{nur am Samstag}]_2 \text{versucht}_2]]$$

This is similar to (4.81).

One comment is necessary regarding (4.71) and (4.72). Since an intraposed VP is an ordinary non-domain-unioned VP like any other, it is not an extraction island. Therefore, we predict that both *wh*-extraction and relativisation out of an intraposed VP are possible as the data indicates.

4-10.3 Intraposed VPs and Argument Inheritance

Pollard ([29]) and Hinrichs and Nakazawa ([17]) have both proposed the use of *argument inheritance* to deal with the word order facts of German. In this section, I will briefly show how the use of argument inheritance inevitably leads to generation of the ill-formed examples of the previous section.

Informally, argument inheritance is the inheriting of some or all of the subcategorisation requirements of a complement by the head that subcategorises for the complement. This nearly always takes the form of a verb “inheriting” some of the subcategorisation requirements of its complement VP. In P&S2, this means that a rule schema such as 4.84 is needed.

$$(4.83) \quad \begin{array}{c} \boxed{0} \\ \text{word} \end{array} \left[\begin{array}{l} \text{SYNSEM|LOC} \left[\begin{array}{l} \text{SUBJECT } \{\boxed{1}\} \\ \text{SUBCAT } \{\boxed{2}, \dots, \boxed{m}\} \cup \{\boxed{m+1}, \dots, \boxed{n}\} \end{array} \right] \end{array} \right]$$

$$(4.84) \quad \left[\begin{array}{l} \text{SYNSEM|LOC} \left[\begin{array}{l} \text{SUBJECT } \{\boxed{1}\} \\ \text{SUBCAT } \{\boxed{m+1}, \dots, \boxed{n}\} \end{array} \right] \\ \text{DTRS} \left[\begin{array}{l} \text{HEAD-DTR } \boxed{0} \\ \text{COMP-DTRS } \{\boxed{2}, \dots, \boxed{m}\} \end{array} \right] \end{array} \right]$$

This schema states that one possible rule configuration is for the first $m - 1$ complements of the **SUBCAT** set to be syntactically saturated and for the other $n - m - 1$ subcategorised elements to be inherited to the mother phrasal sign's **SUBCAT** set. On its own, this rule schema won't accomplish anything. Lexical signs also need to be specified which are compatible with the rule schema. The following sign shows an argument inheriting version of the modal verb *wird* which

obligatorily inherits all of the complement requirements of its complement VP (which will effectively be a verb).¹⁶

$$(4.85) \quad \boxed{2} \text{ VP[BSE]} \left[\text{SYNSEM|LOC} \left[\begin{array}{l} \text{SUBJECT } \{\boxed{1}\} \\ \text{SUBCAT } \boxed{3} \end{array} \right] \right]$$

$$\left[\begin{array}{l} \text{PHON } \langle \text{wird} \rangle \\ \text{SYNSEM|LOC|SUBCAT } \{\boxed{1}\text{NP[NOM]}, \boxed{2}\text{VP[BSE]}\} \cup \boxed{3} \end{array} \right]$$

This lexical entry (and others like it) allow all the complements of embedded verbs to become complements of the finite verb and thus to appear as sisters to the finite verb. Since they are all sisters, they can then occur in any order modulo LP statements. There is one problem with this approach which is relevant to the discussion here which has been pointed out by Carl Pollard ([30]) and Karel Oliva ([29]). Consider the sentence *Wird Hans gehen können?*. By use of the inheritance schema this allows the following possible syntactic structure.

$$(4.86) \quad \left[\begin{array}{l} \text{s} \left[\begin{array}{l} \text{v} \left[\begin{array}{l} \text{SUBCAT } \{\text{VP[BSE], NP[NOM], VP[BSE]}\} \\ \text{wird} \end{array} \right] \\ \text{NP[NOM] Hans} \\ \text{VP[BSE] gehen} \\ \text{VP[BSE, SUBCAT } \{\text{VP[BSE]}\}] \text{ können} \end{array} \right] \end{array} \right]$$

However, this means that the ungrammatical V2 clause *Können wird Hans gehen?* is predicted to be possible via the standard filler-gap schema and SLASH feature conventions. Oliva notes that Pollard proposes to disallow argument inheritance of VP complements to solve this problem.

It is always dangerous to speculate about another theory's potential accounts of some hitherto unexplained data but I will examine what seem to be the two most obvious methods that the argument inheritance approach might use to explain intraposed VPs. The first explanation is that intraposed VPs are saturated VPs which just happen not to be adjacent to the verb cluster. However, such an explanation will generate all of the ill-formed examples of the previous section because there is nothing to prevent argument inheritance of the intraposed VP's complements thus resulting in a "scrambled" intraposed VP.

The second hypothetical solution is to allow inheritance of saturated VPs. Since the intraposed VP would itself be an inherited complement, there is nothing in the rule schemas or lexical entries to allow inheritance out of the VP. Therefore, the account would not overgenerate as discussed above. However, we have already

¹⁶VP here is meant to indicate any verb which has none, some or all of its subcategorisation requirements satisfied except for the value of the SUBJECT feature. Therefore, "partial VPs" can be generated and *wird* can inherit none, some or all of its VP complement's subcategorisation requirements.

(4.88) provides further examples involving sensory, causative, subject control and object control verbs.

- (4.88)
- a. [Lesen sehen] wird Peter den Jungen das Buch
 read see will Peter the boy the book
 'Peter will see the boy read the book'
 - b. [Lesen lassen] wird Peter den Jungen das Buch
 read let will Peter the boy the book
 'Peter will let the boy read the book'
 - c. [Zu lesen versprechen] wird Peter ihr das Buch
 to read promise will Peter her the book
 'Peter will promise her to read the book'
 - d.??[Zu lesen überreden] wird Peter sie das Buch
 to read persuade will Peter her the book
 'Peter will persuade her to read the book'

A problem arises for constructing such examples using *scheinen* as the paradigm instance of a raising verb, as in (4.89e). However, consider (4.89f) where *gepflegt* is given a raising interpretation.

- (4.89)
- e. *[Zu schenken scheinen] sollte Peter ihr dem Jungen das
 to give seem should Peter to her the boy the
 Buch
 book
 'Peter should seem to her to give the boy the book'
 - f. ?[zu lesen gepflegt] hatte er damals nur Romane
 to read used had he then only novels
 'Peter only used to read novels'

Thus, it appears that sensory verbs, causatives, EQUI-verbs and raising verbs (with the exception of *scheinen*) can be fronted with dependent verb, and without their complements, whereas object control verbs do not appear to allow this possibility.

We also saw in (4-9) that it is (sometimes) possible to front a verb cluster with some of the arguments of the verbs in the cluster. Consider the examples in

(4.90):

- (4.90)
- a. [das Buch lesen sehen] wird Peter den Jungen
the book read see will Peter the boy
'Peter will see the boy read the book'
 - b. ihm einen Freund vorstellen sehen hat er sie
him a friend present see has he her
'He saw her present a friend to him'
 - c. *dem Peter vorstellen sehen hat er die Maria einen Freund
the Peter present see has he the Maria a friend
'He saw Mary present a friend to Peter'
 - d. ?einen Freund vorstellen sehen hat er die Maria dem Peter

In each of these examples, the sensory verb *sehen* has been topicalised with its VP complement. In (4.90a) and (4.90b), the VP complement has been fronted intact, whilst in (4.90c) and (4.90d), an accusative direct object or indirect object NP from the VP has left behind, respectively. At a quick glance, the data is rather confusing. On the basis of (4.90a) we might expect that a complement can be topicalised with the verb sequence if it is definite since this would contribute to the topic effect of the topic position. This is partly supported by (4.90d) where the fronted indefinite *einem Freund* makes the example marginal. However, (4.90c) is completely ungrammatical with a definite indirect object. On the other hand, the definite indirect object pronoun *ihm* fronted with the indefinite direct object in (4.90b) is perfectly acceptable. Corresponding examples for causatives, EQUI-verbs and object control verbs (modulo the fact that *versprechen* takes a dative NP complement) are shown in (4.91) to (4.93).

- (4.91)
- a. [das Buch lesen lassen] wird Peter den Jungen
the book read let will Peter the boy
'Peter will let the boy read the book'
 - b. ihr ein Buch schenken lassen hat er den Peter
her a book give let has he the Peter
'He has let Peter give her a book'
 - c. *ihr schenken lassen hat er ihn ein buch
 - d. ?ein Buch schenken lassen hat er ihn ihr
'He has let him give her a book'

- (4.92) a. [das Buch zu lesen versprechen] wird Peter dem Jungen
 the book to read promise will Peter the boy
 'Peter will promise the boy to read the book'
- b. ihr ein Buch zu schenken versprechen wird Peter dem Jungen
- c. *ihr zu schenken versprechen wird Peter dem Jungen das Buch
- d. *ein Buch zu schenken versprechen wird Peter ihm ihr
- (4.93) a. [das Buch zu lesen überreden] wird Peter den Jungen
 the book to read persuade will Peter the boy
 'Peter will persuade the boy to read the book'
- b. ihr ein Buch zu schenken überreden wird Peter den Jungen
- c. *ihr zu schenken überreden wird Peter den Jungen das Buch
- d. **ein Buch zu schenken überreden wird Peter ihm ihr

We can see that in all cases, fronting of the governing verb without its NP complement, but with its VP complement intact (as in the (a) and (b) examples) is acceptable but it is apparently impossible to take along the (definite) indirect object of the VP complement whilst stranding its direct object. Furthermore, the (d) examples are much worse for the subject and object control verbs *versprechen* and *überreden*.

This data presents several problems. First, acceptability differs according to construction type within any of the four sets of examples (4.90)-(4.93). Why is it that topicalisation of the two verbs with the direct object is in general much more acceptable than topicalisation of the verbs with the indirect object? den Besten and Webelhuth [11] provide some insights based on partial VP projections but the matter is far from clear especially as there is a fair degree of speaker variation for these types of sentences. Second, although we can make generalisations across the four sets of examples, what is very clear is that the possibilities of stranding complements become much more marked with the control verbs (especially with object control verbs). This is perhaps not too surprising for object control verbs, since many accounts claim that they are not restructuring verbs. However, we have presented considerable evidence that they are in fact restructuring verbs. It is rather surprising though that (4.92d) should be found unacceptable, for example.

The problem for the argument inheritance account of partial VP fronting is that all of the examples presented above should be completely grammatical. There is no difference taken into consideration between the completely productive rule schemas which govern verb cluster formation for example and the speaker, dialect and language dependent facts governing partial VP fronting. As indicated in the comments above, not even an appeal to topic-focus structure can explain the facts.

4-12 Raising

Uszkoreit ([45]) presents additional data which bears on the empirical issues concerning raising.

(U365) presents a simple example of raising from a single extraposed VP to the finite clause leaving the VP to consist of just *zu* plus the infinitive.

- (U365) Letztes Jahr hatte Peter [das große Haus]_i der Stadt
 last year had Peter the big house the city
 versprochen e_i zu reparieren
 promised to repair
 ‘Last year Peter had promised the city to repair the big house’

(U366) is an important example since it lends support to my previous claim that examples like (U365) are in fact examples of VP extraposition plus raising and not just an alternative verb cluster order with the VP having been domain unioned. (U366) has a structure similar to (U365) except that the extraposed VP also contains the adverbial PP *bis morgen*.

- (U366) Darum hatte ich [dieses Fahrrad]_i den Kindern versprochen e_i
 therefore had I this bicycle the children promised
 bis morgen zu reparieren
 until tomorrow to repair
 ‘Therefore I had promised the children to try to repair this
 bicycle by tomorrow’

Uszkoreit also presents (U367) to show that raising can cooccur with *wh*-movement.

- (U367) Aus welchem Grund hast du [das große Haus]_i der Stadt
 for what reason have you the big house the city
 versprochen e_i zu reparieren
 promised to repair
 ‘For what reason did you promise the city to renovate the big
 house?’

However, (U368) shows that when there is both a *wh*-movement and raising, then

the nested dependency is strongly preferred to the crossing dependency.

- (U368) a.??[In diesen Kasten]₁ hatte er [dem gleichen Affen]₂ versucht e_2
 in this box had he the same monkey tried
 beizubringen, e_1 seine Bananenschalen zu werfen
 to bring his banana peels to throw
 'He had tried to teach the same monkey to throw his banana
 peels into this box'
- b. [dem gleichen Affen]₁ hatte er [In diesen Kasten]₂ versucht e_1
 the same monkey had he in this box tried
 beizubringen, e_2 seine Bananenschalen zu werfen
 to bring his banana peels to throw

Uszkoreit presents (U369) to show that raising is a "long distance dependency" since it can move over two VPs. However, I would rather say that it is a middle distance dependency since it is strictly clause bounded in contrast to *wh*-movement for example. It would not be possible to move the NP *dieses Fahrrad* into the main clause. The top-level simplified domain tree for the subordinate clause in (U369) is given in Fig. 4-16.

- (U369) Ich kann dir das Rad nicht geben, weil ich
 I can you that bike not give, because I
 [dieses Fahrrad]; den Kindern versprochen habe zu versuchen e_i
 this bike the children promised have to try
 bis morgen zu reparieren
 until tomorrow to repair
 'I cannot give you the bike because I promised the children to
 try to repair this bicycle by tomorrow'

Uszkoreit also repeats (U366) as (U370) and the variant (U371) to indicate that the LP constraints also apply to the domain that the raised NP ends up in. Thus the variant (U371) is to be preferred since it obeys the ordering constraint that dative NPs precede accusative NPs. This is important for our account since it indicates that there is no dedicated "landing site" or "filler site" for raised elements as there are in unbounded dependencies. Raised elements end up as part of the finite clause's domain and their position depends on the other elements in the domain and the LP constraints.

- (U370) Darum hatte ich [dieses Fahrrad]; den Kindern versprochen e_i
 therefore had I this bicycle the children promised
 bis morgen zu reparieren
 until tomorrow to repair
 'Therefore I had promised the children to try to repair this
 bicycle by tomorrow'

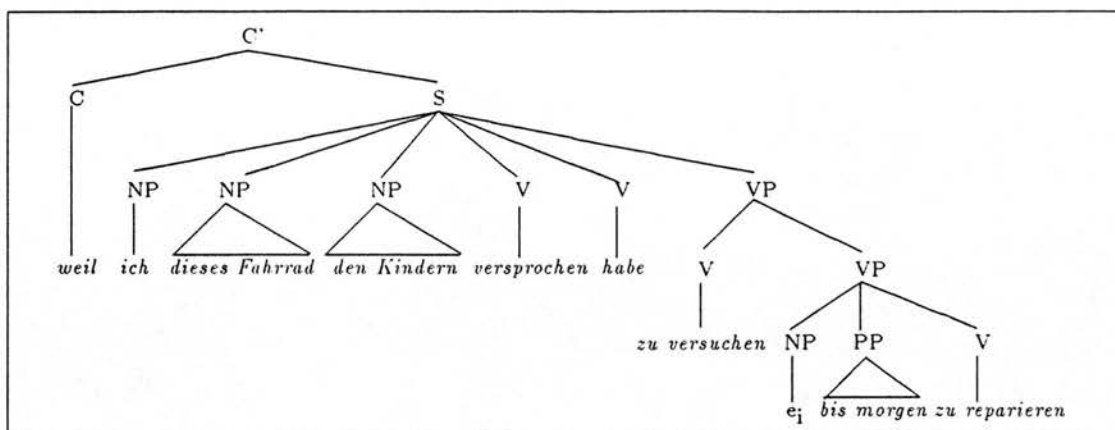


Figure 4-16: Domain tree for subordinate clause of (U369)

- (U371) Darum hatte ich den Kindern [dieses Fahrrad]_i versprochen e_i
 therefore had I the children this bicycle promised
 bis morgen zu reparieren
 until tomorrow to repair
 'Therefore I had promised the children to try to repair this
 bicycle by tomorrow'

(U374) and (U375) indicate that raising is less acceptable if there is an NP in the "landing domain" which has the same case as the raised NP. This is to be expected if raising is a not fully productive process which depends on issues such as pragmatics and topic-focus structure. Allowing multiple occurrences of the same case due to raising makes it more difficult to work out what the dependencies are.

- (U374) ?Ich hatte ihm darum diesen Kindern versprochen zu helfen
 I had him therefore these children promised to help
 'I had him therefore these children promised to help'

- (U375) ?Ich hatte sie deshalb dieses Buch gebeten bis morgen
 I had her therefore this book asked until tomorrow
 zurückzubringen
 to bring back
 'I had her therefore asked her/them to bring this book back by
 tomorrow'

Finally, (U377) shows the raising of two NPs from the same VP to the finite clause. Again notice that the ordering constraints are obeyed. This example is important for third reasons. Again, first it indicates there is no dedicated landing site for raised elements. Second, multiple elements may be raised and they may

give rise to crossing dependencies as opposed to unbounded dependencies which must be nested and third, there are no identifiable gap sites. This is indicated by the disjunction on the coindexing of the gaps.

- (U377) Dann hatte er [den Bestohlenen]₁ [die gleichen Bücher]₂
 then had he the theft victims the same books
 versucht $e_{1/2}$ $e_{1/2}$ zu Schleuderpreisen zurückzuerkaufen
 tried to dumping prices to sell back
 ‘Then he had tried to sell the same books to the theft victims
 again at dumping prices’

4-13 Conclusion

In this chapter, we set out to investigate a number of phenomena in German which taken together typically lead to inconsistencies in configurational approaches to German grammar. As the approach applied here is not configurationally based but rather based on a very regular syntactic and semantic functor-argument structure which is mapped to word order domains which allow discontinuity, it is not surprising that this account has had some success in explaining some of the apparent anomalies found in the data considered.

However, it is clear that more empirical work is needed to determine whether this approach is correct or not. Furthermore, I believe it is essential to investigate this issue in a cross-dialectal and cross-linguistic context to determine precisely what types of variation are available. Monolingual studies may blind us to the generalisations to be found by considering several related languages.

We were also forced to make a minor assumption; namely, that communicative function sometimes plays a role in grammar. To be more specific, sometimes grammar reflects the fact that without some particular idiosyncrasy, it would be impossible to express certain types of propositions. Such an assumption seems at odds with the attempt to provide rigorous formalisation that theories like GPSG or HPSG strive towards but nonetheless empirical evidence seems to demand it occasionally.

Although this chapter is hardly more than a sketch of an analysis of a subset of “the minimum that any self-respecting theory of German should cover”, it is nonetheless encouraging that it covers Netter’s survey of problem data to the extent that it does. Clearly there are generalisations that are being missed (such as the similarity in behaviour of the unannounced passive and unaccusative *lassen* constructions) which need to be examined much more closely. However, I hope that I have shown that there is at least some evidence to support further investigation along these lines.

Chapter 5

A Comparison of German, Dutch, Zurich German and English

*The Lord has promised good to me
His word my hope secures
He will my shield and portion be
As long as life endures
Amazing Grace, Traditional*

5-1 An Analysis of Dutch

5-1.1 Standard Dutch

To account for Dutch, we have to alter Rules 1-3 and change one of the LP statements. Perhaps surprisingly, Rules 1-3 are precisely of the form they take for English (modified of course for the changes to the feature system in P&S1 that we have adopted in this paper). I list them here for convenience.

(5.1) Rule 1 (Dutch)

$$\text{a. } \left[\begin{array}{l} \text{SYN|LOC|ARGS } \langle \rangle \\ \text{DTRS } \left[\begin{array}{l} \text{FUN-DTR|SYN|LOC|LEX -} \\ \text{ARG-DTRS } \langle [] \rangle \end{array} \right] \end{array} \right]$$

$$\text{b. } [\text{ARGS} \langle \rangle] \rightarrow \text{F}[\text{LEX -}], \text{A}$$

(5.2) Rule 2 (Dutch)

$$a. \left[\begin{array}{l} \text{SYN|LOC|ARGS} \langle [] \rangle \\ \text{DTRS|FUN-DTR|SYN|LOC} \left[\begin{array}{l} \text{HEAD|INV} - \\ \text{LEX} + \end{array} \right] \end{array} \right]$$

$$b. [\text{ARGS} \langle [] \rangle] \rightarrow \text{F}[\text{INV} -, \text{LEX} +], \text{A}^*$$

(5.3) Rule 3 (Dutch)

$$a. \left[\begin{array}{l} \text{SYN|LOC|ARGS} \langle \rangle \\ \text{DTRS|FUN-DTR|SYN|LOC} \left[\begin{array}{l} \text{HEAD|INV} + \\ \text{LEX} + \end{array} \right] \end{array} \right]$$

$$b. [\text{ARGS} \langle \rangle] \rightarrow \text{F}[\text{INV} +, \text{LEX} +], \text{A}^*$$

Before we consider the effect of Rules 1-3 we need to state the modification to (3.13) necessary for Dutch. We replace (3.13) with (5.4).

$$(5.4) [\text{DTRS|HEAD-DTR} \square \text{V}[\text{INV} -]] \Rightarrow [\text{DOM} \square \preceq \text{V}]$$

Here we assume that in Dutch, unlike German and like English, the structure [s NP[NOM] VP] is grammatical. Therefore, we need to explain the differences in the structure of V2 clauses and subordinate clauses. In German, we have assumed that in finite clauses, the finite verb is a sister of *all* of its complements in both V2 clauses and subordinate clauses. In Dutch, we assume that the finite verb can be a daughter of a finite VP as in English. Rule 1 allows a nonlexical head (e.g., a VP) to combine with its single unsaturated argument (e.g., the subject). Rule 2 allows a noninverted ([INV -]) lexical head (e.g., a verb) to combine with all of its complements except the first one (e.g., the subject). Rule 3 allows an inverted ([INV +]), lexical head to combine with all of its complements at one time creating an inverted phrase.

We'll now explain subordinate clause order. Nonfinite verbs combine with all of their complements except the first (the subject) by Rule 2 to form nonfinite VPs. Nonfinite, nonextraposed VPs are unioned into governing VPs (as in German). As in German, all verbs must be preceded by all NPs. However, Dutch serialises verbs in the opposite order to German. That is, governing verbs must precede governed verbs. (5.4) forces this order. Finally, subject NPs combine with finite VPs by Rule 1 to form Ss. (Noninverted, finite VPs are not unioned into S.) An LP statement similar to the one for English that requires that complements precede nonlexical heads seems to be required in Dutch as well to force the subject NP to precede the finite VP. This produces the characteristic facts about Dutch

subordinate clause order, namely, that subject NPs may not be interchanged with object NPs (5.5), that objects can in fact sometimes interchange (5.6) and that the canonical order in the verb cluster is that governing verbs precede governed verbs (5.7).

- (5.5) dat mijn vader mijn moeder zag
 that my father my mother saw
 (a) 'that my father saw my mother'
 (b)*'that my mother saw my father'
- (5.6) a. dat hij alleen zijn vader zulke dingen durft te vertellen
 that he only his father such things dares to tell
 'that he only dares to tell his father such things'
 b. dat hij zulke dingen alleen zijn vader durft te vertellen
- (5.7) dat Piet Jan Marie zag helpen zwemmen
 that Pete John Marie saw help swim
 'that Pete saw John help Marie swim'

Of course, there are exceptions to each of these three generalisations. First, passives of ditransitive verbs (as in (5.8)) and unaccusative clauses (as in (5.9) and (5.10)) sometimes allow subject inversion. This indicates that for this class of sentences without an agentive subject, finite subordinate clause structure is as it is in German.

- (5.8) a. dat het boek mijn vader geschonken is
 that the book my father given is
 'that the book has been given to my father'
 b. dat mijn vader het boek geschonken is
 that my father the book given is
 'that my father has been given the book'
- (5.9) dat hem de fouten opgevallen zijn
 that him (ACC) the mistakes (NOM) noticed are
 'that he noticed the mistakes'
- (5.10) dat hem de fouten geirriteerd hebben
 that him (ACC) the mistakes (NOM) irritated have
 'that the mistakes irritated him'

Object swap seems to be limited to cases where the semantics or pragmatics of the sentence makes it clear which object fulfills which object role. This is

evidence for LP constraints of precisely the same type as in German. If LP constraints are ordered with respect to preference along different dimensions, then pairs of constituents which are ambiguous with respect to preferential features will necessarily be given an interpretation that is consistent with the strongest LP constraint. Since Dutch lacks the rich case system of German, this is usually the case. Therefore, the word order appears to be very fixed. However, pragmatic information can be sufficiently strong to allow semi-free order as in German. The consequence of this line of reasoning is that the characteristic “cross-serial” order of most Dutch serial clauses is not proof that Dutch is an indexed language or some other type of context-sensitive language.

We’ll now explain the post-topic structure of V2 clauses in Dutch. This structure is very similar to the structure of subject-auxiliary inversion clauses in English. Basically, an inverted ([INV +]), lexical head (only a verb) can combine with all of its complements at one time. Then, (3.14) requires that the finite verb be clause-initial. The other LP statements then apply as in subordinate clauses. The difference between English and Dutch is that, whereas in English only auxiliaries can be specified [INV +], in Dutch any finite verb can be so specified. The key difference between subordinate and V2 clauses in Dutch then is that in subordinate clauses there is a finite VP whereas in V2 clauses there is no finite VP.

As stated at the beginning, the syntax rules for Dutch are precisely the same as those for English. (A quick check of P&S1 will verify that this is true modulo the changes to the feature system.) There are six major differences between Dutch and English. First, verbs in English do not union their VP complements whereas those in Dutch do. Second, the ordering constraint (5.4) is eliminated for English. Head verbs never end up in the same domain with any of their governed verbs. Third, head verbs precede their NP complements in English unlike German and Dutch. English is lexical-head-initial whereas Dutch and German are head-final in INV – constructions. Fourth, only modals and auxiliaries can be INV + in English whereas any verb can be INV + in Dutch and German. Fifth, there is no process of NP complement raising as there is in Dutch and German. Sixth, main clauses in English are *usually* not V2. They are usually “SVO” clauses of the form [_s NP[NOM] VP[FIN]]. This is just the same as Dutch subordinate clause structure however (modulo the differences noted here). However, English does have the V2 structure in *wh*-questions and “negative adverbial” sentences of the form “Never have I seen such a crazy linguist”, among others.

5–1.2 The Groningen Dialect

In this section, we examine some data collected by Ineke Schuurman ([39]) which she presents to argue for the occurrence of nested VPs. The analysis presented here bears some similarities to her analysis but differs in a few substantive ways. The basic assumptions underlying the account here are that the unmarked order of non *te*-infinitival verbs whose VP domains have been unioned is $V_n \dots V_1$ for n verbs. For example, a three verb sequence occurs in the order $V_3 V_2 V_1$ as in

(IS4a). Notice that the other five permutational possibilities (assuming a verb cluster) in (ISb) are all ungrammatical.

- (IS4)
- a. dat zai plietsie kommen₃ loaten₂ wol₁
 that she police come let wanted
 that she wanted to send for the police
- b. *dat zai plietsie kommen₃ wol₁ loaten₂
 *dat zai plietsie loaten₂ wol₁ kommen₃
 *dat zai plietsie loaten₂ kommen₃ wol₁
 *dat zai plietsie wol₁ kommen₃ loaten₂
 *dat zai plietsie wol₁ loaten₂ kommen₃

The second assumption is that extraposition of *te*-infinitival VPs is possible as in Standard Dutch and German. In (IS5), there is a choice of two analyses. The first is that this is an instance of clause union following the Standard Dutch pattern. The second is that this is an instance of extraposition of the VP *dei truië te braaiën* followed by raising of the NP *de truië* to the finite clause. The latter pattern is very common in both German and Dutch so this is not surprising. The former is to be considered suspect since it would require LP constraints more specific than those which characterise the order in (IS4).

- (IS5)
- dat zai veur t leutje jong dei truië perbaaiert₁ te
 that she for the little boy that sweater tries to
 braaiën₂
 knit
 that she tries to knit that sweater for the little boy

The hypothesis that the second analysis is the more accurate of the two is strengthened by (IS6) and (IS7).

- (IS6)
- dat e koubaaisten toun hening₁ te begunnen₂ te vouern₃
 that he cows then went to start to feed
 that he started to feed the cows at that moment
- (IS7)
- dat zai dei truië perbaaiern₂ wol₁ te braaiën₃
 that she that sweater try wanted to knit
 that she wanted to try to knit that sweater

(IS6) would be given the domain analysis

[_S [_{NP} e] [_{VP} [_{NP} koubaaisten]₁ toun hening [_{VP} te begunnen [_{VP} t₁ te vouern]]]].

That is, *te begonnen* is the head of an extraposed VP which contains a recursively extraposed VP *koubaaisten te vouern* which has *koubaaisten* raised to the finite VP. There is little to decide between the two analyses on just the basis of (IS5) and (IS6) although both patterns are exhibited in both Dutch and German. However, (IS7) provides more conclusive evidence. In this case, the “V₂” *perbaaiern* appears to the left of the finite verb *wol* (V₁) as expected and the verb *te braaiern* is clause final with its complement *dei truie perbaaiern* raised to the finite VP. Since both ordering patterns are present the cleanest theory would be one which supposes that the anomalous word order pattern in (IS7) (which by the way is inconsistent with (IS4)) is a result of extraposition of *dei truie te braaiern* and raising of *dei truie*.

Given these remarks, the data in (ISSa) and (ISSb) would appear to undermine the entire analysis presented so far. In both (ISSa) and (ISSb), there is a verb cluster with the characteristic 3-2-1 order *perbaaiern loaten wol* which is followed by the extraposed verb clusters *te leren zwemmen* and *zwemmen te leren* with raising of the complement *heur kinder* to the finite VP. The possibility of both orders (a) and (b) would seem to be problematic. However, the very possibility of both orders indicates quite clearly that the two sequences are not part of the verb cluster and so they must be an extraposed VP. Furthermore, (IS9) indicates quite clearly that the order in the verb cluster is 3-2-1 invariant. What remains to be explained is the fact that both *te leren zwemmen* and *zwemmen te leren* are possible. Ultimately, we must conclude that the verb which *leren* immediately governs can appear to either side of it *in a verb cluster*. We can find supporting evidence for this from Zurich German which exhibits the same behaviour. Cf. §(5-2). This predicts that the verb immediately governed by *leren* can appear to either side of it in a simple finite verb cluster as well.

- (IS8) a. dat Elsie meester heur kinder perbaaiern₃ loaten₂ wol₁ te
 that Elsie teacher her children try let wanted to
 leren₄ zwemmen₅
 teach swim
 that Elsie wanted to make the teacher try to teach her children
 to swim
- b. dat Elsie meester heur kinder perbaaiern₃ loaten₂ wol₁
 zwemmen₅ te leren₄
- (IS9) dat Elsie meester heur kinder perbaaiern₃ loaten₂ wol₁ te leren₄
 zwemmen₅
 *dat Elsie meester heur kinder perbaaiern₃ wol₁ loaten₂ te leren₄
 zwemmen₅
 *dat Elsie meester heur kinder wol₁ loaten₂ perbaaiern₃ te leren₄
 zwemmen₅
 etc.

(IS10) both supports the hypothesis concerning *leren* and its immediately governed verb and presents yet another puzzle. (IS10) is an example of the so-called

“unannounced passive”. The NP *heur kinder* is the direct object of *leren* and so must be raised to the finite VP. Similarly, the “agentive” *deur*-phrase *deur Kneels* is also raised to the finite VP.

- (IS10) a. dat Elsie heur kinder deur Kneels perbaaiern₂ wol₁ te
 that Elsie her children by Kneels try wanted to
 loaten₃ leren₄ zwemmen₅
 let teach swim
 that Elsie wanted to try to have her children taught to swim by
 Kneels
- b. dat Elsie heur kinder deur Kneels perbaaiern₂ wol₁ te loaten₃
 zwemmen₅ leren₄
 dat Elsie heur kinder deur Kneels perbaaiern₂ wol₁ leren₄
 zwemmen₅ te loaten₃
 dat Elsie heur kinder deur Kneels perbaaiern₂ wol₁ zwemmen₅
 leren₄ te loaten₃
 *dat Elsie heur kinder deur Kneels perbaaiern₂ wol₁ zwemmen₅ te
 loaten₃ leren₄
 **dat Elsie heur kinder deur Kneels perbaaiern₂ wol₁ leren₄ te
 loaten₃ zwemmen₅

perbaaiern wol is the verb cluster (with characteristic 2-1 order) and the sequences composed of *te loaten*, *zwemmen* and *leren* comprise an extraposed VP. Of the six possible permutations of these verbs, only four are acceptable, namely those in which *leren* and *zwemmen* are adjacent. This is not surprising given that the three extraposed verbs form a verb cluster. Again, we see that *leren* and *zwemmen* can come in any order, thus lending support to the claim that they are unordered with respect to each other. However, how can we explain the fact that *te loaten* can be either cluster-initial or cluster-final? The first obvious explanation is that *te loaten* can either precede or follow any verb it governs (at least in an extraposed clause). (IS14) is evidence that this may be the case. Here there is an extraposed VP *te kunnen stoten* which exhibits the order 3-4 (or 1-2 relative to the extraposed VP). Presumably (IS14) can be paraphrased also as *dat hij dat gewicht zal blijken stoten te kunnen*. (Notice, in addition that *zal blijken* exhibits the 1-2 order which is characteristic of Standard Dutch. That there may be differences in the ordering possibilities of modals is exhibited once again by Zurich German. Cf. §(5-2).)

- (IS14) dat hij dat gewicht zal blijken te kunnen stoten
 that he that weight will turn out to can press
 that he will turn out to be able to press that weight

Alternatively we can conclude that *te loaten* actually extraposes its bare infinitival VP complement in the two cases where it appears first in the string of three verbs.

However, I have no data to support such an analysis. Whichever is the case, the account can characterise the facts. Further empirical investigation is necessary to decide which analysis is the correct one.

Schuurman also presents other examples concerning the so-called “pattern reading” of verb clusters containing incorporated nominal elements but I won’t discuss them here since they concern issues which are tangential at best to issues of word order variation.

5–2 An Analysis of Zurich German

5–2.1 Basic Clause Order

For Zurich German (henceforth Zh), Rules 1-3 for German are appropriate. However, (3.12) must be replaced. It appears that the correct generalisation for at least the Zurich dialect is that NP complements need only precede the verb that they depend on but are unordered with respect to all other verbs and follow the same ordering restrictions as German with respect to each other. In (KC88:105) ([7, ex. 105]), the two NPs *sini Chind* and *Mediziin* can appear in either order but *sini Chind* \prec *laat* and *Mediziin* \prec *schtudiere* as required by the generalisation. These are the only six possibilities (where *das er* is initial).

- (KC88:105_a). *das er sini Chind laat Mediziin schtudiere*
 that he his kids lets medicine study
 ‘that he lets his kids study medicine’
- b. *das er sini Chind Mediziin laat schtudiere*
 c. *das er sini Chind Mediziin schtudiere laat*
 d. *das er Mediziin sini Chind laat schtudiere*
 e. *das er Mediziin sini Chind schtudiere laat*
 f. *das er Mediziin schtudiere sini Chind laat*

Therefore, for Zh we must add the following LP constraint which requires that noninverted head verbs must be preceded by all their NP complements.

$$(5.11) \forall \square. [([\text{DTRS} | \text{HEAD-DTR} \ \square] \vee [\text{ARGS} \ \square] \wedge \square \in \square) \Rightarrow [\text{DOM} \ \square \prec \square]]$$

This statement should be read as “every element of a sign’s head daughter verb’s argument list must precede the verb in the sign’s domain”. Or, in other words, all NP arguments must precede their head verb.¹ Notice that the original LP constraint for German NP \prec v implies (5.11). In other words, the Zh constraint is more general.

¹The semantics of the relational dependency \in and the universal quantifier \forall is defined in [34]. It is sufficient here to say that they are classically defined.

To illustrate the point further, (KC88:A:1) ([7, App A, ex. 2]) replaces *laat* 'lets' with *wil la* 'wants to let' and holds the order of the three verbs fixed as *wil* < *la* < *schtudiere*. There are only eight possibilities (with initial *das er*) if the constraint on NP-verb order is maintained and they are all acceptable.

- (KC88:A:1)
- a. *das er sini Chind wil la Mediziin schtudiere*
that he his kids wants let medicine study
'that he wants to let his kids study medicine'
 - b. *das er sini Chind wil Mediziin la schtudiere*
 - c. *das er sini Chind Mediziin wil la schtudiere*
 - d. *das er wil sini Chind la Mediziin schtudiere*
 - e. *das er wil sini Chind Mediziin la schtudiere*
 - f. *das er wil Mediziin sini Chind la schtudiere*
 - g. *das er Mediziin sini Chind wil la schtudiere*
 - h. *das er Mediziin wil sini Chind la schtudiere*

Notice that in (KC88:105), the verbs may appear in both possible orders. This suggests that verbs are unordered with respect to each other in Zh. For the class of verbs that take a bare infinitival VP (with or without an additional NP object complement), this appears to be the correct generalisation. Cooper ([7]) notes that there are 30 possible orderings of *wil sini Chind la Mediziin schtudiere* which follow the NP-verb generalisation and all but three of them are acceptable.² We should note that there are two forms of the causative. *la* is the unstressed "short" form and *laa* is the stressed citation form. The inescapable conclusion to be drawn from the data is that *la* is used if it precedes the verb it governs and *laa* is used if it follows the verb it governs. The "distance" from the NP complement it governs or the verb it governs does not seem to make any difference. This is clearly a case which motivates lexical specification of direction of government. In any case, the German verb ordering constraint (3.13) has to be dropped for Zh, or at the very least, weakened substantially. Cooper also notes that some speakers allow *la* to precede its NP object complement but apparently the stressed form *laa* does not.

²Cooper notes that in these three cases the subordinate clause ends in the sequence *sini Chind laa* which means that a wellformed subordinate clause can be formed if this sequence is dropped. On these grounds she argues quite reasonably that the hearer garden paths on these sentences, analysing the sequence before *sini Chind laa* as a subordinate clause and is thus unable to backtrack. She also provides the thirty admissible permutations of *das ich mues em Vatter hülfe s Gschiiir abwäsche* (KC:B:0) ([7, App. B, ex. 0]) and points out that all of them are acceptable except the three which end in the sequence *em Vatter hülfe*.

- (KC:B:0) *das ich mues em Vatter hülfe s Gschiiir abwäsche*
that I must the father help the dishes wash up
'that I must help father wash up the dishes'

Therefore, in these three cases the same problem arises, namely, that the subordinate clause without the sequence *em Vatter hülfe* is perfectly wellformed and the hearer garden paths. Thus, it does seem that the generalisations are correct.

5-2.2 *z*-VP Extraposition

In this section I will examine extraposition of so-called *z*-VPs in Zh. *z* plays the same role in Zh that *zu* does in German and *te* does in Dutch. This section is based on Cooper's ([8]) account. I will present some basic data from her paper and suggest that there is a simple domain union treatment which accounts for the data in a straightforward way.

(KC90:16a) is an example of recursive VP extraposition in German and (KC90:16b) is the corresponding Zh translation where the German past participle *versprochen* governs the *v*[**zu**] *zu probieren* which in turn governs the *v*[**zu**] *zu erreichen* which are the heads of the two extraposed *VP*[**zu**]s *zu probieren den Hans zu erreichen* and *den Hans zu erreichen* respectively. (KC90:16b) is the corresponding grammatical Zh word for word translation.

- (KC90:16)a. Er hat versprochen zu probieren den Hans zu erreichen
 he has promised to try the Hans to reach
 'He has promised to try to reach Hans'
- b. Er hāt verschproche z probiere de Hans z erreiche

This data indicates that recursive extraposed VPs consisting of a head verb and its complements behave exactly as in German. However, there are two problems.

First, *z* is "missing" in some extraposed VPs when *zu* cannot be missing in the corresponding German examples. Cooper calls this the "missing *z*" problem. Compare (KC90:17) and (KC90:18).

- (KC90:17) Er hat versprochen [*VP* den Hans zu erreichen zu probieren]
 he has promised the Hans to reach to try
 'He has promised to try to reach Hans'

- (KC90:18) Er hāt verschproche [*VP* de Hans probiere z erreiche]
 he has promised the Hans try to reach

We know from (KC90:16) that *verschproche* takes a *VP*[**z**] complement and so *probiere* should occur with *z*. However, this *z* is missing in (KC90:18). To consider this problem, we need to make some assumptions about domain structure. I assume that the structure of (KC90:17) is as indicated in the example. In particular, the syntactic VP *den Hans zu erreichen* is unioned with the domain of the indicated VP that *zu probieren* is the head of. I also assume that the structure of (KC90:18) is (KC90:18d).³

- (KC90:18)d. [_{CP} er [_{S[FIN]} hāt verschproche [_{VP}[**zu**] de Hans probiere z erreiche]]]

³A presentation of the justification of the structure in (KC90:18d) is unfortunately beyond the scope of this paper.

er is the subject occurring in topic position, *hät* is the inverted finite auxiliary appearing clause-initially, *verschproche* is not extraposed and crucially, for the analysis, *de Hans probiere z erreiche* is an extraposed VP. Furthermore, the domain of the VP *de Hans (z) erreiche* is unioned into the domain of the VP headed by *(z) probiere*. This means that the syntactic structure and the domain structure of (KC90:17) and (KC90:18) are the same. The only difference is that a *z* is missing in the Zh example and *probiere* precedes *erreiche* unlike the German example where *zu probieren* governs to the left when its complement is not extraposed.

(KC90:19a) shows that adding the missing *z* makes the example ungrammatical.

(KC90:19)a. *Er hät verschproche de Hans z probiere z erreiche

b. *Er hät verschproche de Hans erreiche z probiere

c.?*Er hät verschproche de Hans z erreiche z probiere

Furthermore, swapping *erreiche* and *probiere* as in (KC90:19b) is ungrammatical. This appears to be due to an ordering constraint which requires that the verb governed by *probiere* follows it. Compare (KC90:25).

(KC90:25)a. Er hät wele de Hans probiere z erreiche

he has wanted the Hans try to reach

'he wanted to try to reach Hans'

b. *Er hät wele de Hans erreiche z probiere

Regardless of whether *probiere* is extraposed or not in (KC90:25a), it is ungrammatical for *erreiche* to precede *probiere* as in (KC90:25b). (KC90:19c) is similar to the German example (KC90:17). It is ungrammatical as well.⁴ (KC90:19c) would be expected to be ungrammatical since *probiere* must govern to the right as indicated above.

The second problem is that *z* can appear on the "wrong" verb. Instead of occurring on the head of a *VP[z]* it may occur on some verb that is governed by the head of the VP as in (KC90:21). (KC90:21a) is expected since *verschproche* governs *laa* and *verschproche* takes a *VP[z]* which *laa* is the head of. We know that *schtudiere* may precede *laa* (as in (KC90:21a)) and that *la* may precede *schtudiere* (as in (KC90:21b)). However, in (KC90:21b), the *z* appears on the wrong verb

⁴Cooper found one speaker who accepted this example. She attributes this to interference from Standard German.

schtudiere. Cooper calls this the “misplaced *z*” problem.

- (KC90:21)a. Er hät verschproche d Chind schtudiere z laa
 he has promised the kids study to let
 ‘He promised to let the kids study’
- b. Er hät verschproche d Chind la z schtudiere
 he has promised the kids let to study
 ‘He promised to let the kids study’

She also cites the examples in (KC90:22) taken from a Zurich radio station. The prepositions *um* in (KC90:22a) and *ohni* in (KC90:22b) and (KC90:22c) subcategorise for a VP[Z].

- (KC90:22)a. Um Gerächtigkeit chöne z haa, mues mer ...
 for justice can to have must one ...
 ‘in order to be able to have justice one must ...’
- b. ...ohni s Schtüürrad mit bedne Händ müese z verlaa
 without the wheel with both hands must to leave
 chönd si rede
 can you talk
 ‘you can phone without having to let go of the steering wheel
 with both hands’
- c. ...ohni de Telefonhörer i de Hand müese z haa
 without the receiver in the hand must to have
 ‘without having to hold the receiver in your hand’

In (KC90:22a), *z* appears on *haa* instead of *chöne* as it should, in (KC90:22b), on *verlaa* instead of *müese* and in (KC90:22c), on *haa* instead of *müese*.

Cooper also cites relevant examples in her thesis. She cites (KC88:50a) and (KC88:50b) from [15] which they claim are grammatical.

- (KC88:50)a. das er wil aagää en Arie z chöne singe
 that he will pretend an aria to can sing
 ‘that he will pretend to be able to sing an aria’
- b. das er wil aagää, z chöne singe

Cooper claims that these sentences are completely ungrammatical because the *z* should appear on *singe* and not *chöne* in both cases. However, (KC88:51a) and (KC88:51b) where *z* appears on the final verb in the extraposed VPs are grammatical.

- (KC88:51)a. das er aagää wil, en Arie chöne z singe
- b. das er aagää wil, chöne en Arie z singe

(KC88:51b) shows that *z* marks the last verb in an extraposed VP and not necessarily the last verb in a continuous verb cluster. Cooper also gives examples (KC88:52).

(KC88:52)a. das er en Arie wil aagää chöne z singe

b. das er wil aagää, chöne en Arie z singe

(KC88:52a) should be analysed as the verb cluster *wil aagää* followed by the extraposed VP *chöne z singe* with *en Arie* raised to the finite clause. (KC88:52b) is just like (KC88:51b) except that *aagää* and *wil* have swapped positions.

The analysis that I would like to suggest theory-neutrally (which Cooper dismisses in [8]) is that if a verb subcategorises for a $\text{VP}[z]$ and the VP is extraposed then *z* marks the last V in the extraposed VP's domain. This means that if several VPs have been unioned together but the head of the extraposed VP is not the last verb in the domain then *z* will not occur on it. It also means that if domain union in an extraposed VP should give rise to more than one *z* in an extraposed VP (as in the German example (KC90:17)) then only one *z* will appear. However, if the extraposed VP itself governs a recursively extraposed *z*-VP, then *z* will appear on the last verb in its domain and so on as in (KC90:16b) which is assigned the domain structure (KC90:16d).

(KC90:16d) [_{CP} er [_S hät versproche [_{VP[z]} z probiere [_{VP[z]} de Hans z erreiche]]]]

This is sufficient to explain all of the data presented and much more but it does depend on (descriptively) distinguishing between cases of "verb raising", "verb projection raising" and extraposition very carefully. Making this distinction in Zh can be very difficult because of the amount of order freedom.

Although it would seem to be easy to explain this data in terms of a domain union account which says that the final verb in an extraposed VP is *z*-marked, it is rather difficult because of the construction rules. For a start, it is not the case that the occurrence of *z* is construction dependent, i.e., that it depends completely on the fact that the VP is realized in extraposed position. First, some verbs take nonextraposed $\text{VP}[z]$ complements (KC91:8) ([8, ex. 8]) and second some verbs

appear to extrapose VP[INF]s (KC91:7) ([8, ex. 8]).⁵

- (KC91:8) a. das er*hät vill z tue / vill z tue hät
 that he has much to do / much to do has
 'that he has a lot to do'
- b. das er*isch z beduure / z beduure isch
 that he is to pity / to pity is
 'that he is to be pitied'
- (KC91:7) a. das er nöd wil [sini Chind schtudiere laa]
 that he not wants his kids study let
 'that he doesn't want to let his kids study'
- b. das er sini Chind laat [Mediziin schtudiere]
 that he his kids lets medicine study
 'that he lets his kids study medicine'
- c. das er mich gseet [s Gschiir abwäsche]
 that he me sees the dishes wash up
 'that he sees me wash up the dishes'
- d. das er em Vatter hilft [s Gschiir abwäsche]
 that he the father helps the dishes wash up
 'that he helps father wash up the dishes'
- e. das er wird [schpöter aachoo]
 that he will later arrive
 'that he will arrive later'

There are basically three options that I see. All of them depend on the idea that the *z* can "float off" of its lexical head to the last verb in an extraposed VP's domain. Thus the *z* will only appear in the domain of the VP[*z*]. This requires a treatment of morphosyntax which I am unprepared to give here to explain how the *z* gets incorporated into separable prefix verbs like *biilegge* to give *biizlegge* when the account makes it clear that *biilegge* is what comes out of the lexicon. This in itself is enough to treat the "misplaced *z*" problem. However, it does not handle the "missing *z*" problem because multiple *z*'s may appear before the final verb. Barring deletion or identification of all these *z*'s (which would be in violation of the domain construction principles) this problem is irreparable.

The second option is an elaboration of the first. It says that the head verb of the extraposed VP does indeed come out of the lexicon with the *z* and that it floats

⁵It is arguable that all of these cases involve domain union or perhaps verb projection raising. In any case, none of the German counterparts of these verbs allow extraposition. Furthermore, (KC91:7a) and (KC91:7e) seem like particularly convincing examples of extraposition. In fact, (KC91:7a) supports an optional intonation break before the putative extraposed VP which would be unexpected if *sini Chind schtudiere laa* was not an extraposed VP. (Cooper, p.c.)

off to the last V but in addition, every verb which is the head of a nonextraposed VP in the government chain is required not to be a $v[z]$. Those which would be $v[z]$ if extraposed become $v[INF]$ s instead. This option does cover both the missing and misplaced z problems. On face value, this appears to be descriptively necessary, since the verb forms of the missing z verbs which appear in the surface string are all $v[INF]$ s when they would be $v[z]$ s if they appeared as the head of a topicalised VP.

A further, third refinement is possible which claims that the z is just like the English *to*, that is, it is a functor from $VP[INF]$ s to $v[z]$ s. However z domain unions its complement $VP[INF]$ and then floats to the last verb and enforces the government requirement above. This at least means that there is no need to explain how the z floats off of its head verb although we still have to explain how the z undergoes incorporation with separable prefix verbs.

Some account of morphosyntactic processes is required for the treatment of V2 given in German, Dutch and Zh. In all three cases, we assume that the finite verb appears clause-initially. Obviously, in the case of separable prefix verbs, this is insufficient. To take just one example, in (5.12) ([45, ex. 184]), the separable prefix verb *anrufen* consists of a separable prefix *an* and an infinitival stem *rufen*. In V2 and V1 clauses ((5.12c) and (5.12d) respectively) the inflected stem appears in domain initial position. So something still needs to be said about the morphosyntax of separable prefix verbs at least.

- (5.12)
- a. Peter wird Paul anrufen
Peter will Paul call up
'Peter will call Paul up'
 - b. weil Peter Paul anruft
because Peter Paul calls up
'because Peter calls up Paul'
 - c. Peter ruft Paul an
Peter calls Paul up
'Peter calls Paul'
 - d. Ruft Peter Paul an?
calls Peter Paul up
'Does Peter call Paul'

One possibility is to give lexical entries domains. Then the domain of *anruft* might consist of two elements, one for *an* and one for *ruft* (where arbitrary categorial information could be associated with each element). This two element domain could then either undergo domain union (in the case of V2 and V1 clauses) or appear as an element of its mothers domain continuously (i.e., nonunioned). Different LP constraints might then apply to the different elements of the lexical domains.

This might help us explain the baffling placement of the prefix *aan* immediately

before the verb cluster and after the negation *niet* in the Dutch example (5.13) and its effect on the scope of *niet* with respect to the three verbs.

- (5.13) a. dat een lijfwacht de koningin niet moet kunnen aankijken
 that a bodyguard the queen not must can look at
 (a) 'that a bodyguard must not be able to look at the queen'
 (b)? 'that a bodyguard must be unable to look at the queen'
 (c)* 'that a bodyguard must be able not to look the queen'
- b. dat een lijfwacht de koningin niet aan moet kunnen kijken
 that a bodyguard the queen not at must can look
 (a) 'that a bodyguard must not be able to look at the queen'
 (b) 'that a bodyguard must be unable to look at the queen'
 (c) 'that a bodyguard must be able not to look the queen'

At this point, Dowty's use of *attachment* ([12]) starts to look very attractive for explaining the "integrity" of morphologically complex lexical items in some positions and their discontinuity in other positions. All of this is really just an argument for sublexical structure which will have to be addressed anyway.

Finally, although a statement of the government requirement that all verbs which would normally be *v[z]*s in the domain of an extraposed VP must be *v[INF]*s except for the head verb is easy to state, it is messy to formalise in the formalism of HPSG. It is certainly possible to "program" the features to make it work, but the results are not elegant. Rather, it seems preferable to derive the phenomenon from more general principles, especially as there is relevant data (although impoverished) from German as in (5.14) ([43, p380]) and (5.15) ([43, p444]). See [43] and [8] for discussion.

- (5.14) a. ohne ihn haben sehen zu können
 without him have see to can
 'without having been able to see him'
- b. *ohne ihn zu haben sehen können
- (5.15) Er scheint ihn haben sehen zu können
 he seems him have see to can
 'He seems to have been able to see him'

5-3 Comparative Clause Structure

In this section I will briefly indicate how the main clause and subordinate clause structures of German, Dutch and English compare and where the difference and similarities between them originate from the point of view of the account given in the preceding chapters.

German main clauses are verb-second (V2) clauses (Figure 5-1). They consist of a topic position followed by an inverted head-initial clause. This inverted clause does not have an articulated [_S NP[NOM] VP] structure. Instead, Rule 3 (German) creates an inverted clause which consists of the finite daughter and all of its complement. The topic can be filled by any major category (XP) by an unbounded dependency in the clause. German subordinate clauses (Figure 5-2) consist of a complementiser followed by a noninverted head-final clause. As in V2 clauses, there is no NP-VP structure.

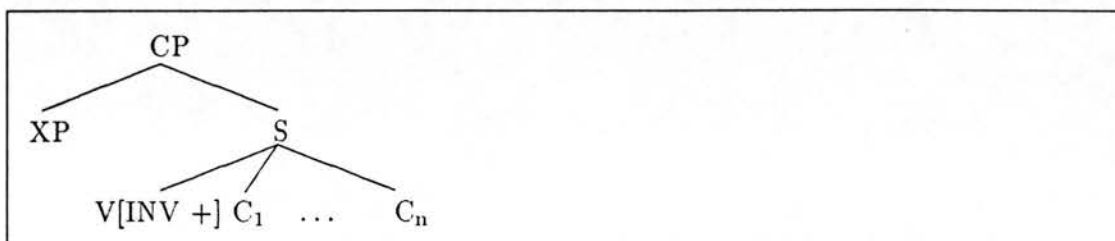


Figure 5-1: German V2 clauses

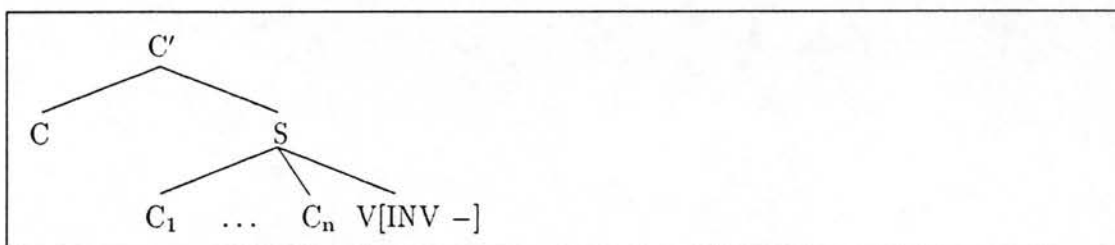


Figure 5-2: German Subordinate clauses

Dutch V2 clauses (Figure 5-3) are like German V2 clauses. They are formed as a topic followed by an inverted head-initial clause. Rule 3 (Dutch) allows the formation of an inverted head-initial clause without an NP-VP structure. However, Dutch subordinate clauses (Figure 5-4) do show an NP-VP structure. The VP is nevertheless noninverted and head-final.

English so-called “subject-aux inversion” clauses (Figure 5-5) are just like the head-initial clauses of German and Dutch. They occur in yes-no questions (V1) and as constituents of V2 constructions. Although normally declarative sentences are not V2 clauses, V2 clauses do appear as wh-questions and negative adverbial sentences as mentioned earlier. Unlike German and Dutch, English main and subordinate clauses have the same structure. This so-called “SVO” structure (Figure 5-6) is exactly the same as the constituent clause in Dutch subordinate clauses except that noninverted VPs are head-initial in English unlike the head-final Dutch VPs.

Therefore, the same structures are available to both English and Dutch. Whereas

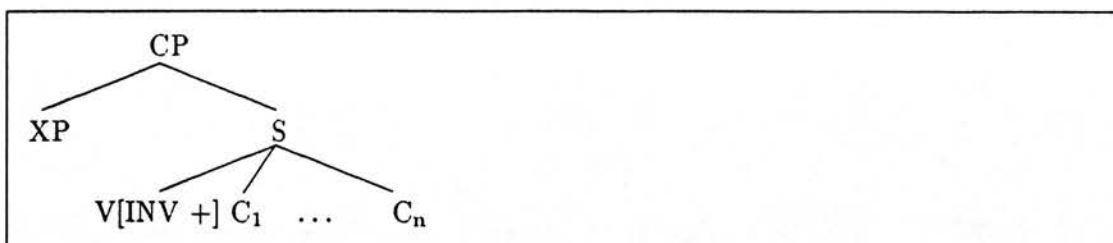


Figure 5-3: Dutch V2 clauses

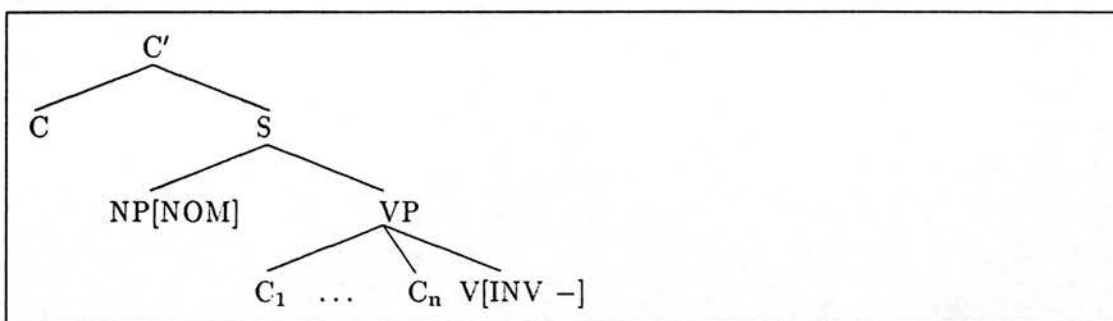


Figure 5-4: Dutch Subordinate clauses

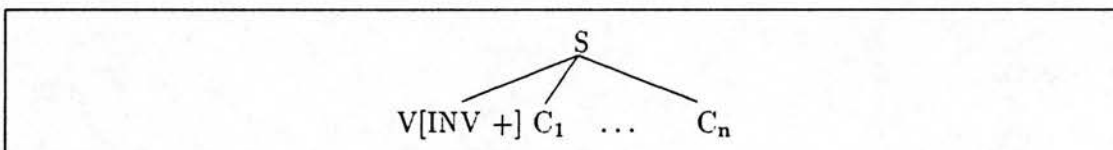


Figure 5-5: English Subject-Aux inversion clauses

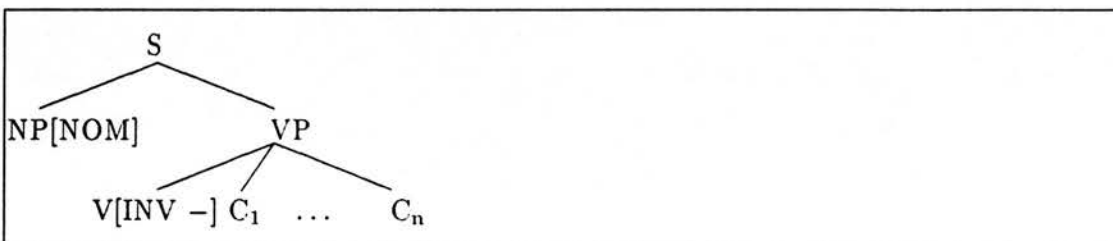


Figure 5-6: SVO clauses

Dutch employs V2 clauses as main clauses, English uses SVO clauses, which are constituents of Dutch subordinate clauses, for both main clauses and subordinate clauses. On the other hand, English only uses V2 clauses for *wh*-questions and negative adverbial clauses. Finally, the only major difference between Dutch and English on the one hand and German on the other is that the SVO structure is unavailable in German altogether. This is usually referred to by saying that German is “nonconfigurational” or that it has a subject-inclusive VP.

Clearly, the outline of V2 clauses here is inspired by the standard GB analysis, namely that the topic fills the specifier of the CP (or complementiser phrase) position. In fact, the category label I have assigned to V2 clauses is CP since I essentially agree that the topic is in the [Spec, CP] position. Unlike GB analyses however, I do not assume that the finite verb fills COMP position. Rather, it is just initial in the constituent finite clause domain. I will simply say that the movement to COMP analysis is not justified by this approach and I am aware of convincing dialect data where both Spec and C(omp) can be filled simultaneously. Therefore, COMP is either empty or not present in V2 clauses. Given the treatment of the complementiser as a functor, it is reasonable to assume that the COMP position is not “topologically obligatory”. On the other hand, the Spec position is topologically obligatory in V2 clauses. This uncontroversial claim is based on contrasts such as the behaviour of impersonal passive constructions in German such as (5.16).

- (5.16) a. daß wurde getanzt
 that was danced
 ‘that there was dancing’
- b. es wurde getanzt
 it (EXPL) was danced
 ‘there was dancing’

By itself, *wurde getanzt* is a full-fledge finite clause and as such can appear as the argument of the complementiser *daß* in (5.16a). However, in (5.16b) there is nothing in *wurde getanzt* to fill the topic position so it is filled with the expletive pronoun *es*. That is, topic position must be filled. Given the analysis presented here, the only way that this can be the case is if it is topologically obligatory, or in other words, that it is assigned by a rule of syntax.⁶

⁶As far as I can see, there is no way to formulate the Head-Filler rule so that both fillers arising from unbounded dependencies and the expletive *es* can appear in topic position. A separate rule would seem to be needed for the expletive *es* case which is similar to the Head-Filler rule.

Chapter 6

Conclusions

6-1 A Selective Summary

This work had three primary goals: (1) to provide a treatment of bounded discontinuous constituency and word order in general and semi-free word order in particular, (2) to provide an alternative account of cross-linguistic word order (particularly in West Germanic) to the *Principles and Parameters* approach of Government and Binding Theory and (3) to formalise the account in a single homogeneous logical formalism which is not based on rewrite rules or other formal language theoretic machinery. Each of these goals has been achieved.

The treatment of bounded discontinuous constituency is based on the use of word order domains and word order domain union as the primary relation between word order domains. The treatment of semi-free word order is based on the use of linear precedence constraints over those word order domains. Furthermore, a theory of linear precedence constraints based on partially ordered sets has been presented which deals with degrees of word order freedom in a novel way.

An alternative account of cross-linguistic word order has also been presented. This account is based on the notions of *theory* and *subtheory*. A theory is a subtheory of another if it contains all of its axioms (and possibly more). The grammars of closely related languages are then taken to be subtheories of the theory of the language family. An example of cross-linguistic variation is presented in the grammar fragments of German, Dutch, Zurich German and English. In this case, the theory of the "language family" is that of universal grammar.

The presentation of universal grammar and the grammar fragments is presented in a two-dimensional variant of the language \mathcal{L}^+ which is formalised in the appendix. The formalisation of such a language not based on formal language theoretic concepts is crucial for enabling the definition of domain union.

In addition, a substantial fragment of German grammar has been presented which covers a wide range of traditionally difficult to account for phenomena. Accompanying the German fragment there are accounts of the syntax and semantics of

quantifiers (and specifiers more generally) complementisers and adjuncts. The treatment of adjuncts includes a discussion of the scope of modification.

6-2 Directions for Future Research

There are many topics which deserve additional attention with respect to the theory of bounded discontinuous constituency and semi-free word order proposed here. I will consider several of them here now.

The process that I have called "raising" and that Uszkoreit calls "focus raising" deserves considerable extra attention. Unlike other researchers, I have assumed that cases of complement raising from extraposed VPs and the formation of "partial VPs" in the context of complex fronting arises from one unified process that I call simply "raising". Raising is a not fully productive and perhaps stylistic process. Its properties are language, dialect and speaker dependent and raising in the context of complex fronting need not have the same properties as raising from extraposed VPs. In fact, in Dutch complex fronting is hardly tolerated at all whereas raising from extraposed VPs is common (especially with intonational support). Much additional research is required to determine exactly what conditions license raising and to determine whether in fact the phenomenon exhibited in complex fronting and raising from extraposed VPs is a single phenomenon.

With respect to German, an obvious phenomenon which needs to be investigated in the context of union of domains is subordinate clause and relative clause extraposition. Two solutions come immediately to mind. The first involves using the attachment operators of Dowty ([12]). All of the prenominal adjuncts of a noun would then be attached to the noun whereas the relative or subordinate clause would not be. Then if the NP is unioned into its mother's domain (i.e., the VP or S) then the relative or subordinate clause would be free to appear in clause final position. The second solution is suggested by Dowty in a comparable discussion for English in his paper [12]. This time, restricted relative clauses are of the schematic category NP/NP. That is, they are NP modifiers and not \bar{N} modifiers. In this case, there is no attachment, but the daughter NP is not domain unioned while the mother NP may be. Since the relative clause is a sister to the daughter NP, domain union results in the daughter NP appearing as a sister element of the mother's domain to the relative clause. The relative clause is then free to occur in clause final position. Obviously, further empirical research is required to determine the best theoretical solution in this case.

One very interesting project would be to give a comprehensive cross-linguistic account of expletives in German, Dutch and English. Much of the difference between German and Dutch can be explained on the basis of the topologically obligatory position of topic position in German and Dutch and the obligatory subject position of Dutch. It would be interesting to try to formulate an account of English expletives that was more in line with corresponding German and Dutch accounts. This is made all the more challenging by data such as clause extraposi-

tion from object position with an expletive (or possibly pronominal) left behind in the Mittelfeld in German where there is no such requirement for subject clause extraposition.

Another topic that requires much additional research is the phenomena that have come to be labelled as “verb projection raising”. As indicated elsewhere in the thesis, we treat this as basically extraposition of non-*zu* or non-*te* infinitival VPs. Again, this seems to be highly language, dialect and speaker dependent. The interesting cases are those such as in Flemish, where, in our terms, it looks like there is VP extraposition plus raising. den Besten and his colleagues (cf. [10]) have proposed that a submaximal V projection has been raised corresponding to Evers’ rule of V-raising. However, this makes several predictions about the structure of VPs which to my knowledge have not been verified yet. The predictions are very similar to those made for binary branching analyses of complex fronting. The crucial question is in what order do complements (and adjuncts) combine with their heads. The data is conflicting because it is possible for example both to front a verb with a modifier while stranding its complements and to raise (or extrapose) a verb with a modifier but no complements. Furthermore, there are asymmetries between the direct and indirect objects, asymmetries which depend on verb type and speaker variation.

One topic which requires a unified treatment is topicalisation in German. The conventional wisdom that appearance in topic position is a good test for constituency in German is just wrong. The converse is true however. If X is a constituent, then X can be fronted. There are too many contradictions to the former however, not least of which, is the fact that multiple adverbials can be fronted. Other counterevidence is the fact that nonfinite verbs with nonagentive subjects can sometimes be fronted. Furthermore, “fronting” or “topicalisation” is perhaps a bad term. In so-called “split topicalisation”, a full NP can appear in topic position while there is a quantifier which takes scope over it in the Mittelfeld. A corresponding English example would be ‘Many books, I have none’. Finally, the fact that the expletive *es* can occur in topic position indicates that the topic need not be filled by movement or its counterparts in nonderivational theories. Rather, there is either a deeper process at work or topic position can be “filled” by a number of different, non-overlapping processes. Again, the data is language, dialect and speaker dependent and much more empirical research is necessary to provide an adequate basis for a theoretical explanation.

The question of whether object control verbs undergo V-raising or clause union (or domain union in our case) is still somewhat equivocal. I have argued that it must undergo domain union based on certain empirical evidence. However, there is a strong preference for object control verbs to act as if they do not. Perhaps one way out of this situation is to observe that there is a preference for object control verbs to attach their direct object unless the direct object is a pronoun, etc. Dowty argues for a similar attachment of the direct object in English to the head verb. Again, more empirical research is required to establish whether there is domain union with a strong ordering preference for the NP not to move or whether some other theory must be explicated to explain those cases in which

the direct object is not adjacent to its head.

Something that is sorely missing in our treatment is a theory of the morphosyntax of separable prefix verbs. A discussion of the numerous possibilities is well beyond the scope of this summary. However, it is clear that this is a crucial issue in both German and Dutch syntax, perhaps even more so in Dutch since there is quite a bit of prefix order freedom. Precisely what the ramifications of the competing proposals are for syntax, morphology and the lexicon need to be studied carefully.

Another dimension which is unfortunately missing from this study is the use of diachronic evidence to support synchronic syntactic theory. I feel that is particularly important in the study of those phenomena, such as verb projection raising, which are not fully productive and which are language, dialect and speaker dependent. Jack Hoeksema has shown the relevance of such study as I have already mentioned elsewhere with respect to verb projection raising in Dutch. The very fact that some aspect of syntax is not fully productive and seems to be highly idiosyncratic almost begs for diachronic investigation since we should expect to find in our synchronic studies, the remnants of past synchronic processes, processes which are now dead.

An obvious shortcoming of the current work is the relatively shallow account of Dutch and Zurich German. It would be highly desirable to study both of these languages in considerably more depth than I have done especially as they exhibit very interesting data for any theory of discontinuous constituency and semi-free word order.

In a similar vein, a thorough analysis of Flemish (perhaps based on van Haegeman's work on West Flemish) is called for. The language is different enough from Dutch that a comparative study of the two languages might reveal deeper insights than study of either alone could provide.

Our account has some interesting consequences for an account of passivisation in German and Dutch. Unfortunately, such a discussion is beyond the scope of this work. This imbalance needs to be redressed.

One minor problem with the theory of reflexivisation presented is that it gives

no explanation of examples of the following kind (taken from [27]):

- (McK33) a James Bond; liess / sah den Spion; mit dem Revolver auf
 James Bond let / saw the spy with the revolver at
 sich; / ihn; zielen
 himself / him aim
 ‘James Bond let / saw the spy aim at him with the revolver’ (R.
 58c) [37, 31]
- b Hans; liess Emma; den Kinderwagen (nicht) neben sich; /
 Hans let Emma the pram (not) beside himself /
 ihm; abstellen
 him leave
 ‘Hans let (didn’t let) Emma leave the pram beside him’

McKay gives a good summary of the problems and analyses several proposals to date. The problems appear to me to be of a more empirical nature than one of supplying an adequate theory. I refer the reader to [27].

Finally, one topic that has been neglected entirely is coordination. An empirically adequate account of coordination in our framework looks like it would have to be based on domain structure. There is some empirical support for this idea. First, if we compare an approach that is based on a flat clause structure with recursive \bar{V} projections, we would not expect to be able to coordinate some suffix of the NPs with a prefix of the Vs as shown schematically below.

$$[X \text{ NP}_1 [\text{NP}_2 \text{ V}_1] \text{ conj} [\text{NP}_3 \text{ V}_2] \text{ V}_3]$$

(Note that the subscripts do not indicate dependency.) However, since the domain structures we hypothesise are entirely “flat” we could predict that any subsequence of the domain could be coordinated and this in fact seems to be the case in both German and Dutch.

I will not speculate any further on this topic. Suffice it to say that a theory of coordination needs to be developed.

Appendix A

Formalisation of the HPSG Formalism

A-1 Introduction

In this appendix, we will investigate extensions suggested by attempts to eliminate the formal language theoretic component of unification-based grammar formalisms and formalise all dimensions of linguistic structure within a single homogeneous feature value logic.¹ Many of the extensions considered below are suggested in HPSG by Pollard and Sag [31] (P&S1) and [32] (P&S2) and by similar treatments including this thesis. This appendix is an attempt to rigorously formalise some of the notation used in P&S1 and P&S2 and this thesis in a modal setting.

In particular, we will be concerned with formalising *functional* and *relational dependencies*. We will argue that the formal semantics that we give to these types of dependencies is the one which most closely matches the informal, intuitive semantics or use of such dependencies in actual grammatical practice. We will then find that the formal semantics has some pleasant features. Among these features is the ability to reconstruct the *type-token* distinction, or to put it another way, the *extensionality-intensionality* distinction for arbitrary structures and the ability to describe *cyclic* or *nonwellfounded* structures of all types.

§A-2 begins with a quick review of the syntax, semantics and proof theory of the language \mathcal{L} . \mathcal{L} is basically Kasper-Rounds logic ([25]) augmented with classical negation but which also replaces path equations with *variables* to indicate reentrancy. For example, $a : b : c = d : e : f$ is expressed as $(a : b : c : x) \wedge (d : e : f : x)$ where x is a variable. Variables serve exactly the same purpose as “indexes” in two-dimensional representations of feature structures.

In §A-3 we examine the motivation for functional and relational dependencies as found in P&S1 and P&S2 and provide some examples of each type of dependency.

¹This appendix is based on [34, Ch. 4]. For related material, cf. [34].

We then argue that the standard predicate modal logic approach to functions and relations is inadequate to capture the intuitive semantics of functional and relational dependencies as used in the literature. We then present what we feel is the correct formal semantics for these dependencies. Function and relation symbols turn out to be *existential polyadic polymodal* modalities. After briefly presenting some examples, we then present the proof theory for formulas containing function and relation symbols. The proof theory is “compositional” in a pleasant way in the sense that the proof theory for function symbols is the same as that of relation symbols except that one of the axioms for relation symbols is generalised slightly.

§A-4 discusses the fact that the interpretation of relation symbols is *intensional* whereas the interpretation of function symbols is *extensional*. This allows the type-token distinction for arbitrary objects, including sets. §A-5 discusses the fact that the language easily allows the description of cyclic or nonwellfounded structures. Such structures have been proposed within Situation Theory and have become objects of study in their own right in the study of nonwellfounded set theory. (Cf. [1].) In particular, this solves some of the problems that Rounds addressed in [38] by providing a language for describing cyclic structures and reentrancy simultaneously. §A-6 discusses some other aspects of the language’s expressive power that are useful in the context of grammar writing. §A-7 briefly returns to the topic of nonwellfounded structures to consider extensionality and nonwellfounded set theory.

Finally, §A-8 presents a translation from the two-dimensional feature structure notation used in P&S1, P&S2 and this thesis into the language investigated in the previous sections. Although the feature structure notation is convenient for formalising grammars, the one-dimensional, linear notation is more convenient for investigating the logic itself.

A-2 The Language \mathcal{L}

We’ll first briefly present the syntax, semantic and proof theory of the language \mathcal{L} so that we have a standard set of definitions to refer to in the discussion in the rest of this appendix.²

²For proofs of the theorems, cf. [34]. Completeness proofs are also given in [35].

A-2.1 Syntax

Let F be a set of *features* (or *attributes* or *labels*). Let X be a set of *variables*. Let $A \subseteq X$ be a set of *atoms* (or *constants*). Then \mathcal{L} is the smallest set such that

$$\begin{array}{l}
 \top \in \mathcal{L} \\
 \perp \in \mathcal{L} \\
 x \in \mathcal{L} \quad \forall x \in X \\
 f : \phi \in \mathcal{L} \quad \forall f \in F, \phi \in \mathcal{L} \\
 \phi \wedge \psi \in \mathcal{L} \quad \forall \phi, \psi \in \mathcal{L} \\
 \phi \vee \psi \in \mathcal{L} \quad \forall \phi, \psi \in \mathcal{L} \\
 \neg \phi \in \mathcal{L} \quad \forall \phi \in \mathcal{L}
 \end{array}$$

We also define \rightarrow and \leftrightarrow metasyntactically in the usual way. Furthermore, we define $\langle f \rangle \phi =_{\text{def}} f : \phi$ and $[f] \phi =_{\text{def}} \neg \langle f \rangle \neg \phi$.

A-2.2 Semantics

Let W be a set of *worlds* (or *indexes* or *nodes*). Then a *partial functional polymodal frame* is a structure $\mathcal{F} = \langle W, \{R_f | f \in F\} \rangle$ s.t. $R_f : W \rightarrow W$ is a partial function for all $f \in F$. A *valuation* $V : X \rightarrow 2^W$ is a partial function s.t. each variable is assigned a singleton set and the restriction of V to A must be bijective. A *model* for \mathcal{L} is a structure $\langle \mathcal{F}, V \rangle$ where \mathcal{F} is a partial functional polymodal frame, V is a valuation and if $a \in A$ and $V(a) = \{w\}$, then $R_f(w)$ is undefined for all $f \in F$.

Let $\mathcal{M} = \langle \mathcal{F}, V \rangle$ be a model. Then

- (A.1) $\mathcal{M} \models_i \top$
- (A.2) $\mathcal{M} \not\models_i \perp$
- (A.3) $\mathcal{M} \models_i x \Leftrightarrow V(x) = \{i\}$
- (A.4) $\mathcal{M} \models_i f : \phi \Leftrightarrow R_f(i) = j$ and $\mathcal{M} \models_j \phi$
- (A.5) $\mathcal{M} \models_i \phi \wedge \psi \Leftrightarrow \mathcal{M} \models_i \phi$ and $\mathcal{M} \models_i \psi$
- (A.6) $\mathcal{M} \models_i \phi \vee \psi \Leftrightarrow \mathcal{M} \models_i \phi$ or $\mathcal{M} \models_i \psi$
- (A.7) $\mathcal{M} \models_i \neg \phi \Leftrightarrow \mathcal{M} \not\models_i \phi$

A-2.3 Proof Theory

In this section, we briefly present two minimal bases for \mathcal{L} . The first is in terms of an extension of the minimal polymodal logic K_i .

Theorem A.1 $\mathcal{L} = K_i \text{Det}_i \text{NomBiFin}$.

$$\begin{aligned}
K_i &: [i](\phi \rightarrow \psi) \rightarrow ([i]\phi \rightarrow [i]\psi) \text{ for all } i \in F \\
Nec_i &: \frac{\psi}{[i]\psi} \text{ for all } i \in F \\
Det_i &: \langle i \rangle \phi \rightarrow [i]\phi \text{ for all } i \in F \\
Nom &: (\sigma : x) \wedge (\sigma' : (x \wedge \phi)) \leftrightarrow (\sigma : (x \wedge \phi)) \wedge (\sigma' : x) \text{ for all } \sigma, \sigma' \in F^* \\
Bi &: a \wedge b \leftrightarrow \perp \text{ for all } a, b \in A \text{ s.t. } (a \neq b) \\
Fin &: a \wedge f : \phi \leftrightarrow \perp \text{ for all } a \in A \text{ and } f \in F
\end{aligned}$$

The notation $K_i A_1 \dots A_n$ denotes the system consisting of K_i , Nec_i and $A_1 \dots A_n$ and propositional logic.

Consider the following axioms and rule of inference.

$$\begin{aligned}
D^\vee &: f : (\phi \vee \psi) \leftrightarrow (f : \phi) \vee (f : \psi) \text{ for all } f \in F, \forall \phi, \psi \in \mathcal{L} \\
f^\perp &: f : \perp \leftrightarrow \perp \text{ for all } f \in F \\
f^\leftrightarrow &: \frac{\phi \leftrightarrow \psi}{f : \phi \leftrightarrow f : \psi} \text{ for all } f \in F, \forall \phi, \psi \in \mathcal{L} \\
D^\wedge &: f : (\phi \wedge \psi) \leftrightarrow (f : \phi) \wedge (f : \psi) \text{ for all } f \in F, \forall \phi, \psi \in \mathcal{L}
\end{aligned}$$

The following lemma states that there is an alternative basis for \mathcal{L} just in terms of the existential features.

Lemma A.1

$$\begin{aligned}
K_i &\equiv D^\vee \\
Nec_i &\equiv \{f^\perp, f^\leftrightarrow\} \\
Det_i &\equiv D^\wedge
\end{aligned}$$

These two bases show that \mathcal{L} is a *nominal deterministic polymodal logic with bijective, final constants*. In the sequel we will assume that our models have bijective, final constants.

A-3 The Language \mathcal{L}^+

A-3.1 Motivation

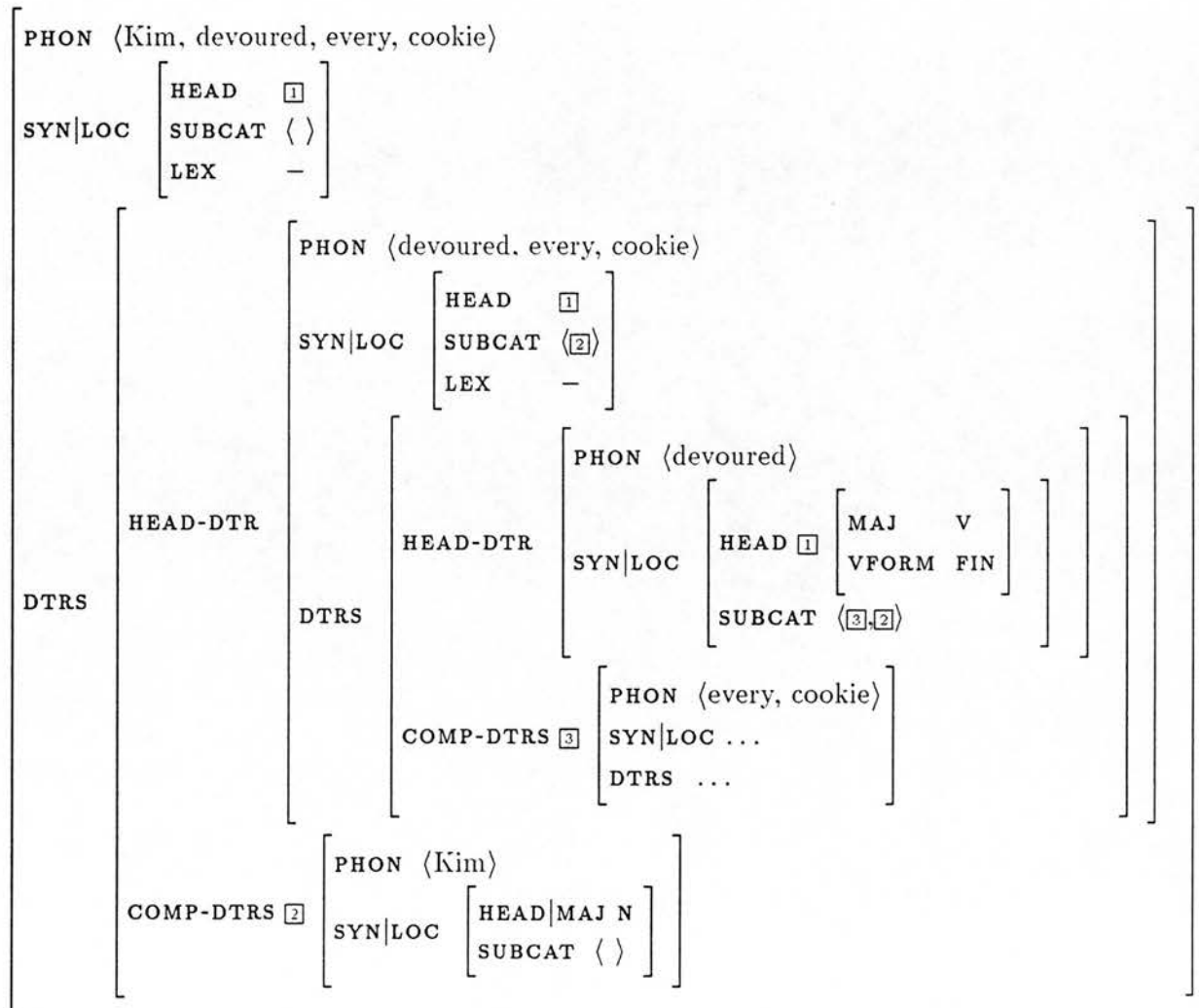
Two major extensions which have been proposed for feature value logics (FVLs) are *functional* and *relational* dependencies. A *functional dependency* exists when the value of a path is a function of the value(s) of one or more other paths. Paradigm examples of functional dependencies are the use of *sets* and *sequences* in HPSG. A set can be analyzed formally as a functional dependency since a set is nothing more than the union of the singleton sets of its elements. For example, $\{a, b, c\} = \{a\} \cup \{b\} \cup \{c\}$ and each union is a functional dependency. Similarly a sequence can be treated as the concatenation of one element sequences of its

elements. E.g., $\langle a, b, c \rangle = \langle a \rangle \circ \langle b \rangle \circ \langle c \rangle$ where \circ is the concatenation operator. Each concatenation is a functional dependency.

A *relational dependency* exists when the values of one or more paths are required to stand in some relation. A paradigm example of a relational dependency is the use of the membership relation (\in) in P&S1. In the following sections we give examples taken from P&S1 of each of these types of dependencies

Sequences

In the following example (P&S1, p. 152, ex. (290)), we see that the value of the PHONOLOGY (PHON) attribute is a sequence of atoms and that the values of the COMPLEMENT-DAUGHTERS (COMP-DTRS) and SUBCATEGORISATION (SUBCAT) attributes are sequences of signs. There are many other uses of sequences in HPSG. In this case, sequences replace formal language theoretic machinery in order to encode the syntax and word order of the sentence *'Kim devoured every cookie'*.



Sets

In this example (P&S1, p. 104, ex. (186)), the value of the **SEMANTICS** attribute for the sentence 'Kim saw Sandy' is shown. The **SEMANTICS** value consists of two parts, the **CONTENT** (**CONT**) and the **INDEXES** (**INDS**). The **INDEXES** attribute takes a set of *indexes* as its value. These indexes consist of a *variable* (**VAR**) and a *restriction* (**REST**) to that variable, thus giving something like a generalized quantifier treatment to noun phrases.

$$\left[\begin{array}{l} \text{CONT} \\ \text{INDS} \end{array} \left[\begin{array}{l} \left[\begin{array}{l} \text{REL} \quad \text{SEE} \\ \text{SEER} \quad \boxed{2} \\ \text{SEEN} \quad \boxed{1} \end{array} \right] \\ \left\{ \left[\begin{array}{l} \text{VAR} \quad \boxed{1} \\ \text{REST} \quad \left[\begin{array}{l} \text{RELN} \quad \text{NAMING} \\ \text{NAME} \quad \text{Sandy} \\ \text{NAMED} \quad \boxed{1} \end{array} \right] \end{array} \right\} , \left\{ \left[\begin{array}{l} \text{VAR} \quad \boxed{2} \\ \text{REST} \quad \left[\begin{array}{l} \text{RELN} \quad \text{NAMING} \\ \text{NAME} \quad \text{Kim} \\ \text{NAMED} \quad \boxed{2} \end{array} \right] \end{array} \right\} \right\} \end{array} \right] \right]$$

Non-algebraic Functions

In this example (P&S1, p. 110, ex. (201), 'Semantics Principle (fourth version)'), the value of the **SEM|CONT** path is a function of the values of the **DTRS|HEAD-DTR** and **SEM|COMP-DTRS** paths. Furthermore, the value of the **SEM|INDICES** path is a function of the value of the **DTRS** path.

$$\left[\begin{array}{l} \text{SEM} \\ \text{DTRS} \end{array} \left[\begin{array}{l} \text{DTRS } \textit{headed-structure} [] \Rightarrow \\ \left[\begin{array}{l} \text{CONT} \quad \text{successively-combine-semantic}(\boxed{1}, \boxed{2}) \\ \text{INDICES} \quad \text{collect-indices}(\boxed{3}) \\ \text{HEAD-DTR|SEM|CONT} \quad \boxed{1} \\ \text{COMP-DTRS} \quad \boxed{2} \end{array} \right] \end{array} \right]$$

Relations

In this example (P&S1, p. 161, ex. (305), 'Rule 4'), the values of the **DTRS|HEAD-DTR|SYN|LOC|HEAD|ADJUNCTS** and the **DTRS|ADJ-DTR|SYN** paths stand in a set membership relation.

$$\left[\begin{array}{c} \text{DTRS} \\ \left[\begin{array}{c} \text{HEAD-DTR|SYN|LOC} \\ \text{ADJ-DTR|SYN } \square \end{array} \right] \left[\begin{array}{c} \text{HEAD|ADJUNCTS } \{\dots, \square, \dots\} \\ \text{LEX } - \end{array} \right] \end{array} \right]$$

A-3.2 Predicate Modal Logic

Given the amount of effort that we have put into investigating \mathcal{L} as a modal logic, one might suppose that predicate modal logic would be a natural way to integrate functional and relational dependencies into \mathcal{L} . So, let's review the syntax and semantics of predicate modal logic briefly to see if it has the right properties.

Let R be a set of *predicate symbols*. Let Z be a set of *individual variables*. Let C be a set of *individual constants*. Let F be a set of *function symbols*. The set of *terms*, T , is the smallest set such that $c \in T$ for all $c \in C$, $z \in T$ for all $z \in Z$ and $f(t_1, \dots, t_n) \in T$ for all $t_1, \dots, t_n \in T$ and n -ary $f \in F$. $g(t_1, \dots, t_n)$ is an *atomic formula* for all $t_1, \dots, t_n \in T$ and n -ary $g \in R$. Finally, the set of formulas is the smallest set such that ϕ is a formula if ϕ is an atomic formula, $\neg\phi$, $\square\phi$ and $\diamond\phi$ are formulas if ϕ is a formula and $\phi \wedge \psi$ and $\phi \vee \psi$ are formulas if ϕ and ψ are formulas.

Semantically, things are slightly different than in propositional modal logic. A *model* is a structure $\langle \mathcal{F}, D, V \rangle$ where \mathcal{F} is a frame in the familiar sense, D is a *domain of individuals* and V is a valuation on both individual variables and predicate and function symbols. Satisfaction for the classical connectives and the modalities \square and \diamond are defined as usual. However, terms do not stand in the satisfaction relation. Instead, they have a *denotation* which is an element of the domain D . The denotation $\llbracket t \rrbracket$ of a term t which is a variable or a constant is $V(t)$. The denotation of an n -ary function symbol f is a function $V(f) : D^n \rightarrow D$. If $t = f(t_1, \dots, t_n)$ for some n -ary $f \in F$ and $t_1, \dots, t_n \in T$, then $\llbracket t \rrbracket = V(f)(\llbracket t_1 \rrbracket, \dots, \llbracket t_n \rrbracket) \in D$. That is, the denotation of the term is the function applied to the denotations of its arguments.

For relation symbols, things are slightly different. The valuation of an n -ary relation symbol g is a relation $V(g) \subseteq D^n \times W$. Then

$$\mathcal{M} \models_w g(t_1, \dots, t_n) \Leftrightarrow \langle \langle \llbracket t_1 \rrbracket, \dots, \llbracket t_n \rrbracket \rangle, w \rangle \in V(g)$$

There is an important point to note about this last definition. The valuation of an n -ary relation symbol g which intuitively relates n -tuples of individuals is an $n + 1$ relation over n individuals and one world. The intuition is very clear and very simple. If worlds are treated as *possible worlds* in the philosophical sense, then it makes sense to talk about a set of n individuals standing in some relation g in world w_1 (say yesterday) but not in a different world w_2 (say today). So, for example, the predication *happy(John)* is true on Friday afternoons at 4:30PM but is false on Monday mornings at 7:00AM.

So, let's see where we are. Terms are not formulas. Function symbols can only take terms as their arguments. Relation symbols also only take terms as their arguments. That is, formulas can never be arguments to either relation or function symbols. Furthermore, a relation symbol plus arguments is an atomic formula. Most importantly, a term denotes an individual (not a world) whereas a formula denotes a truth value when evaluated at a world or it denotes the set of worlds at which it is true. How does this match up with our examples of functional and relational dependencies? Very badly.

If we look at ex. (290) from P&S1 again, we find several problems. First, the value of $\text{DTRS|HEAD-DTR|DTRS|HEAD-DTR|SYN|LOC|SUBCAT}$ is a two element sequence containing the indexes $\boxed{3}$ and $\boxed{2}$. But $\boxed{3}$ and $\boxed{2}$ (since they are just variables) denote worlds. That is, the arguments of the sequence concatenation operator are not terms. Second, we do have genuine terms as the value of the five PHON attributes. However, these terms appear where formulas are supposed to appear. We see the same problems if we look at exs. (186) and (201) from P&S1 again. If we now look at ex. (305) once more, we see that the index $\boxed{1}$ again denotes a world (or truth value) but is also the argument of the implicit membership relation. The basic situation that we are left with is that formulas can appear where terms appear and terms can appear where formulas can appear. It doesn't take very long to conclude that terms should be made first-class formulas and that arbitrary formulas can appear as the arguments to both function symbols and relation symbols.

This of course will come as no surprise to those familiar with formalisms like FUG , HPSG and LFG . Since functional dependencies are functions over the values of paths, they must allow arbitrary formulas as arguments. The same holds for relational dependencies. Syntactically it is very clear what to do. How about semantically? Two possibilities suggest themselves. The first possibility is to *identify* the set of worlds and the set of individuals in the model. Then the syntactic homogeneity between function and relation symbols makes perfect sense semantically. However, there is a less obvious but ultimately preferable possibility. That is to treat function and relation symbols as *polyadic* modal operators, i.e., as modalities which take more than one operand formula. This is the strategy we will follow below although it will only become obvious when we start to look at the proof theory. For this reason, in what follows we will define an interpretation function I for function and relation symbols which is part of the *frame* of a model instead of being separate from the frame. This is because the interpretation function is really an n -ary accessibility relation.

We'll now go on to define the syntax and semantics of \mathcal{L} plus function and relation symbols, or \mathcal{L}^+ as we will call it from now on.

A-3.3 Syntax

Let Rel be a set of *relation symbols*. Let $Fun \subseteq Rel$ be a set of *function symbols*. For each $g \in Rel$, let there be a unique nonnegative integer called its *arity*. Then

\mathcal{L}^+ is the smallest set s.t.

$$\begin{array}{ll}
\top \in \mathcal{L}^+ & \\
\perp \in \mathcal{L}^+ & \\
x \in \mathcal{L}^+ & \forall x \in X \\
f : \phi \in \mathcal{L}^+ & \forall f \in F, \phi \in \mathcal{L}^+ \\
\phi \wedge \psi \in \mathcal{L}^+ & \forall \phi, \psi \in \mathcal{L}^+ \\
\phi \vee \psi \in \mathcal{L}^+ & \forall \phi, \psi \in \mathcal{L}^+ \\
\neg \phi \in \mathcal{L}^+ & \forall \phi \in \mathcal{L}^+ \\
g(\phi_1, \dots, \phi_n) \in \mathcal{L}^+ & \forall n\text{-ary } g \in Rel, \forall \phi_1, \dots, \phi_n \in \mathcal{L}^+
\end{array}$$

(We also allow infix and prefix notation for function symbols where there is no danger of confusion.)

There are really only two comments to make here. First, functional formulas and relational formulas are defined in precisely the same way and can occur in exactly the same syntactic environments. Second, the set of function symbols is a subset of the set of relation symbols. This is because the semantics of function symbols is a special case of the semantics of relation symbols as we will now see.

A-3.4 Semantics

Let W be a set of worlds. Then an *interpretation function* $I : Rel \rightarrow (W^n \times W)$ is a function such that for every n -ary relation symbol $g \in Rel$, $I(g) \subseteq W^n \times W$ and for every n -ary function symbol $f \in Fun$, $I(f) \subseteq W^n \rightarrow W$ is a *partial* function. A *frame* \mathcal{F} is a structure $\langle W, \{R_f | f \in F\}, I \rangle$ such that $\langle W, \{R_f | f \in F\} \rangle$ is a partial functional polymodal frame, if $V(a) = \{w\}$ then $R_f(w)$ is undefined for all $a \in A$ and all $f \in F$ and I is an interpretation function. A model for \mathcal{L}^+ is a structure $\langle \mathcal{F}, V \rangle$ where \mathcal{F} is a frame and V is a valuation. Then for all n -ary $g \in Rel$

$$\begin{aligned}
\mathcal{M} \models_w g(\phi_1, \dots, \phi_n) &\Leftrightarrow \\
&\exists w_1, \dots, w_n. (\mathcal{M} \models_{w_1} \phi_1 \wedge \dots \wedge \mathcal{M} \models_{w_n} \phi_n \wedge I(g)((w_1, \dots, w_n), w))
\end{aligned}$$

and for all n -ary $f \in Fun$

$$\begin{aligned}
\mathcal{M} \models_w f(\phi_1, \dots, \phi_n) &\Leftrightarrow \\
&\exists w_1, \dots, w_n. (\mathcal{M} \models_{w_1} \phi_1 \wedge \dots \wedge \mathcal{M} \models_{w_n} \phi_n \wedge I(f)(w_1, \dots, w_n) = w)
\end{aligned}$$

A few comments are in order here. First, just like in predicate modal logic, the denotation of an n -ary relation symbol is an $n + 1$ relation where the $n + 1$ -th argument is the ‘‘actual’’ world, i.e., the world at which the relational formula is being evaluated. This has several happy consequences, two of which we will discuss here. First, it greatly simplifies the proof theory compared to a semantics which requires that n worlds stand in some relation at some world w iff they

stand in that relation in every world. Second, it allows the semantics of function symbols to be a special case of the semantics of relation symbols. In the case of function symbols, satisfaction is further restricted so that the interpretation of the function symbol is actually a partial function from the n -tuple of argument worlds to the actual world. This means that the proof theory of function symbols is a slight superset of the semantics of relation symbols. It also means that we can profitably trade on the fact that function symbols are partial functional and relation symbols aren't. We will see much more of this in the following sections.

A-3.5 Some examples

We'll now look at some examples of how functional and relational dependencies can be translated in an elegant way into \mathcal{L}^+ .

Functions

Example 1 (Sequence descriptions) Let \circ be a binary function symbol and ϵ a constant such that

$$\begin{aligned}\epsilon \circ a &\leftrightarrow a \leftrightarrow a \circ \epsilon \\ (a \circ b) \circ c &\leftrightarrow a \circ (b \circ c)\end{aligned}$$

for a , b and c metavariables over constants. Then $\langle A, \circ, \epsilon \rangle$ is a monoid (where A is the set of constants).

Then we can translate the following sequence description

$$\left[\text{PHON } \langle \text{Maggie, Thatcher, is, finally, gone} \rangle \right]$$

into \mathcal{L}^+ as

$$\text{phon} : (\text{Maggie} \circ \text{Thatcher} \circ \text{is} \circ \text{finally} \circ \text{gone})$$

Example 2 (Set descriptions) Let \cup and \cap be binary operators, \setminus be a unary operator and 1 and 0 be constants such that the usual Boolean equivalences hold. Then $\langle A, \cap, \cup, \setminus, 1, 0 \rangle$ is a Boolean algebra (where A is the set of constants).

Then we can translate the following set description

$$\left[\text{SEM|INDS } \{\boxed{1}, \boxed{2}, \boxed{3}\} \right]$$

into \mathcal{L}^+ as

$$\text{sem} : \text{inds} : (x \cup y \cup z)$$

Relations

The original ‘Rule 4’ from P&S1 is stated as follows where the ellipsis dots in the set brackets indicate that the value indexed by index \square is an element of the set.

Original ‘Rule 4’:

$$\left[\begin{array}{c} \text{DTRS} \\ \left[\begin{array}{c} \text{HEAD-DTR|SYN|LOC} \\ \text{ADJ-DTR|SYN } \square \end{array} \right] \left[\begin{array}{c} \text{HEAD|ADJUNCTS } \{\dots, \square, \dots\} \\ \text{LEX } - \end{array} \right] \end{array} \right]$$

With relation symbols, we can do better. Let in be a binary relation symbol s.t. $in(x) \leftrightarrow x \cup \top$. Then ‘Rule 4’ can be expressed as followed.

‘Rule 4’ expressed in \mathcal{L}^+ with in :

$$\begin{aligned} dtrs : (& \text{head-dtr} : \text{syn} : \text{loc} : (\text{head} : \text{adjuncts} : (y \wedge in(x)) \wedge \\ & \text{lex} : -) \wedge \\ & \text{adj-dtr} : \text{syn} : x) \end{aligned}$$

A-3.6 Proof theory

We’ll now discuss the proof theory of \mathcal{L}^+ . Before we can present the axiomatisation, we need to define some terminology. The following should be familiar from the discussion of predicate modal logic.

Let the set of *terms* T be the smallest set s.t.

$$\begin{aligned} x \in T & \quad \forall x \in X \\ f(t_1, \dots, t_n) \in T & \quad \forall f \in Fun, \forall t_1, \dots, t_n \in T \end{aligned}$$

Definition A.1 Let $\llbracket t \rrbracket$ be the denotation of a term t . Then

$$\begin{aligned} \llbracket x \rrbracket &= V(x) & \forall x \in X \\ \llbracket f(t_1, \dots, t_n) \rrbracket &= I(f)(\llbracket t_1 \rrbracket, \dots, \llbracket t_n \rrbracket) & \forall f \in Fun, \forall t_1, \dots, t_n \in T \end{aligned}$$

Lemma A.2

$$\mathcal{M} \models_i t \Leftrightarrow \llbracket t \rrbracket = \{i\}$$

The set of *generalized paths* P is the smallest set s.t.

$$\begin{aligned} \epsilon \in P \\ f : p \in P & \quad \forall f \in F, \forall p \in P \\ g(\dots, \overset{i}{p}, \dots) \in P & \quad \forall g \in Rel, \forall p \in P, \forall \phi_1, \dots, \phi_{i-1}, \phi_{i+1}, \dots, \phi_n \in \mathcal{L}^+ \end{aligned}$$

Notation The notation $g(\dots, p^i, \dots)$ indicates that p is in the i th argument position of g .

If $p \in P$ is a generalized path, then $p^\pi(\phi)$ denotes a formula formed by substituting the formula ϕ into p at the "location" in p encoded by the "position" π .³

With these definitions and notational conventions we can now state a minimal basis for \mathcal{L}^+ .

Theorem A.2 *Let $g \in Rel$, $\phi, \psi, \phi_1, \dots, \phi_n \in \mathcal{L}$, $p_1, p_2 \in P$, $x \in X$ and $t \in T$. Then the following are a minimal basis for \mathcal{L}^+ .*

$$(A.8) \quad g(\dots, \phi \vee \psi, \dots) \leftrightarrow g(\dots, \phi, \dots) \vee g(\dots, \psi, \dots)$$

$$(A.9) \quad g(\dots, \perp, \dots) \leftrightarrow \perp$$

$$(A.10) \quad \frac{\phi \leftrightarrow \psi}{g(\dots, \phi, \dots) \leftrightarrow g(\dots, \psi, \dots)}$$

$$(A.11) \quad p_1^\pi(x \wedge \phi) \wedge p_2^{\pi'}(x) \leftrightarrow p_1^\pi(x) \wedge p_2^{\pi'}(x \wedge \phi)$$

$$(A.12) \quad p_1^\pi(t \wedge \phi) \wedge p_2^{\pi'}(t) \leftrightarrow p_1^\pi(t) \wedge p_2^{\pi'}(t \wedge \phi)$$

Notice first of all that (A.8), (A.9) and (A.10) are similar to the nondeterministic fragment of \mathcal{L} . This is because \mathcal{L}^+ without terms is just the minimal normal polyadic modal logic K^n . In fact, the same types of interderivable results hold. (A.8) is equivalent to a form of K and (A.9) and (A.10) are equivalent to a form of Necessitation. (A.10) allows us to extend substitution of equivalents to relation symbols. (A.11) is a generalized version of the variable schema which takes account of the fact that variables can appear embedded in the arguments of relation symbols. (A.8)–(A.11) are sufficient to axiomatise \mathcal{L} plus relation symbols only (\mathcal{L}^R), i.e., no function symbols. If we add function symbols, then (A.8)–(A.10) remain but (A.11) is replaced by (A.12) which is a generalisation of it. (A.12) covers the fact that terms also denote single worlds.

(A.8)–(A.12) is complete for \mathcal{L}^+ but they are not particularly useful since there isn't an axiom for every connective. So, we state two more axioms below (without proof) which are theorem schemata of \mathcal{L}^+ .

Theorem A.3 *(A.13) and (A.14) are theorem schemata of \mathcal{L}^+ .*

$$(A.13) \quad \phi \wedge \psi, \dots \rightarrow g(\dots, \phi, \dots) \wedge g(\dots, \psi, \dots)$$

$$(A.14) \quad \phi_1, \dots, \phi_n \rightarrow g(\neg\phi_1, \top, \dots, \top) \vee \dots \vee g(\top, \dots, \top, \neg\phi_n) \vee \neg g(\top, \dots, \top)$$

³The particular encoding used is irrelevant so long as the position of every subformula in a formula is encoded uniquely. For an example of such an encoding, cf. [41].

Example of (A.11) In order to demystify (A.11) and (A.12), here is a simple example. Let $f, g \in F$, $h, j, k \in Rel$, $x \in X$ and $\phi, \psi, \chi \in \mathcal{L}^+$.

$$\begin{aligned} f : g : h(f : g : (x \wedge \phi), \psi) \quad \wedge \quad h(j(k(x, \psi, \chi))) \\ \leftrightarrow \\ f : g : h(f : g : (x), \psi) \quad \wedge \quad h(j(k(x \wedge \phi, \psi, \chi))) \end{aligned}$$

The left conjunct in the “top” half of the biconditional has a generalised path along the features $f : g$, then “through” the relation h and then along the first argument path $f : g$ to the formula $x \wedge \phi$. The second conjunct also has a generalised path through the first arguments of the relation symbols h, j and k to the formula x . Therefore, by (A.11) (and (A.12)) the ϕ which is conjoined to variable x in the left conjunct can be moved over and conjoined to the variable x in the right conjunct. The “bottom” half of the biconditional shows exactly this. In other words, (A.11) is just like the variable schema except that it lets you look inside relation arguments as well as along paths for formulas conjoined to variables. (A.12) is just like (A.11) except that it allows the same thing to be done with terms and not just variables.

A-4 Intensionality and Extensionality

Consider (A.15) and (A.16).

$$(A.15) \quad \left[\begin{array}{c} \mathbf{F} \\ \mathbf{G} \end{array} \left[\begin{array}{c} \mathbf{G} \ 1 \\ \mathbf{H} \ 2 \\ \mathbf{G} \ 1 \\ \mathbf{H} \ 2 \end{array} \right] \right]$$

$$(A.16) \quad \left[\begin{array}{c} \mathbf{F} \\ \mathbf{G} \end{array} \left[\begin{array}{c} \boxed{1} \left[\begin{array}{c} \mathbf{G} \ 1 \\ \mathbf{H} \ 2 \end{array} \right] \\ \left[\begin{array}{c} \mathbf{G} \ 1 \\ \mathbf{H} \ 2 \end{array} \right] \end{array} \right] \right]$$

If we consider these feature structures to be objects themselves instead of descriptions of objects then the values of \mathbf{F} and \mathbf{G} are *token identical* in (A.16), i.e., they must be the same object whereas in (A.15), the values of \mathbf{F} and \mathbf{G} are isomorphic but not the same object. We say that they are *type identical*.

That is, standard feature structure notation is *intensional*. If it were *extensional*, then two feature structures would be isomorphic iff they were the same object.

I.e., (A.15) and (A.16) would be equivalent. Intensionality allows the possibility of isomorphic but not equal objects.⁴

Classical set theory (e.g. **ZFC**) is extensional. Two sets are equal iff they contain the same elements. However, since FVLs (including \mathcal{L}) are intensional with respect to feature structures it is reasonable to consider the intensionality/extensionality distinction with respect to any type of object in one's ontology. For example, we might want to allow the possibility that two sets which contain the same elements are not the same set or that two sequences containing the same elements are not the same sequence. Furthermore, we might want to allow isomorphic terms from an arbitrary term algebra to be distinct. We will now see that \mathcal{L}^+ allows the extensionality/intensionality distinction with respect to every function and relation symbol. To make this more concrete, let's consider an example.

Consider a binary function symbol \circ axiomatised by A (the associativity axiom) and a binary relation symbol \bullet also axiomatised by A. I.e.,

$$\begin{aligned} A: a \circ (b \circ c) &\leftrightarrow (a \circ b) \circ c \\ A: a \bullet (b \bullet c) &\leftrightarrow (a \bullet b) \bullet c \end{aligned}$$

Then

$$\vdash f : (x \circ y) \wedge f : \phi \wedge g : (x \circ y) \rightarrow g : \phi$$

by (A.12) but

$$\not\vdash f : (x \bullet y) \wedge f : \phi \wedge g : (x \bullet y) \rightarrow g : \phi$$

I.e., the two occurrences of the term $x \circ y$ are necessarily true at only one and the same world but the two occurrences of the relational formula $x \bullet y$ may be true at different worlds (i.e., be satisfied at different worlds). Therefore, we can effectively have two "versions" of the same algebraic operators, one extensional (a function symbol; \circ in this case) and one intensional (a relation symbol; \bullet in this case).

Furthermore, we can use relation symbols in conjunction with variables to enforce (token) identity and without variables to express simple type identity. For example, using the standard terminology, in

$$f : ((x \bullet y) \wedge z) \wedge g : ((x \bullet y) \wedge z)$$

⁴My use of the terms *extensional* and *intensional* might seem a little strange to someone who is only used to thinking about extensionality and intensionality in terms of set theory. However, I'm trying to describe a more general distinction which is essentially structural.

we would say that the two $x \bullet y$ are *token-identical* since they are conjoined with the same variable z whereas in

$$f : (x \bullet y) \wedge g : (x \bullet y)$$

the two $x \bullet y$ are only *type-identical*.

If this seems a little mysterious, it translates into feature structure notation as follows:

$$\left[\begin{array}{l} \mathbf{F} \ \boxed{1} (x \bullet y) \\ \mathbf{G} \ \boxed{1} (x \bullet y) \end{array} \right]$$

vs.

$$\left[\begin{array}{l} \mathbf{F} \ (x \bullet y) \\ \mathbf{G} \ (x \bullet y) \end{array} \right]$$

In summary, we can consider relation symbols to be “intensional” function symbols and function symbols to be “extensional” function symbols.

There are two final comments. First, most of the discussion above treats the formulas or *descriptions* as if they were the structures themselves, e.g., as if relation symbols were algebraic operators in intensional set theory. This may seem a contradiction of the *description/object distinction* but it isn't. The description/object (or syntax/semantics) distinction is mostly a technical one designed to solve problems about the “meaning” of disjunction, etc. But there are many model structures and many possible semantics that can be given to \mathcal{L}^+ which have the same proof theory. Ultimately, it is the proof theory that matters. We can use the language to axiomatise and reason about any mathematical domain which it is strong enough to talk about. We will do more of this in the next section and in §A-7 where we will see how \mathcal{L}^+ can talk about nonwellfounded structures and, in particular, nonwellfounded set theory.

Second, one might object that the formalism is stronger than the kinds of formalisms that linguists want. However, I tried to argue very carefully in §A-3.2 that the formal semantics defined here is precisely what the informal semantics requires. The properties discussed in this and the following sections follow automatically from that semantics.

A-5 Nonwellfounded Structures

\mathcal{L}^+ can also describe *cyclic* or *nonwellfounded* structures of all types. E.g.,

$$\begin{aligned} x \wedge (f : g : x) \\ x \wedge (x \circ y \circ z) \\ x \wedge in(x, y) \end{aligned}$$

The first of these is a cyclic feature structure. The second is a sequence which contains itself as its first element. The third is a relation whose first argument is the variable whose denotation is the world at which the relation is required to be true.

In general, nonwellfounded structures are always possible. The proof theory does not derive any inconsistencies from them. To selectively eliminate various types of nonwellfoundedness, we can add axiom schemas of the following types since \mathcal{L}^+ is strongly complete.

$$\begin{aligned} x \wedge \sigma : x &\leftrightarrow \perp \quad \forall \sigma \in F^+ \\ x \wedge (x \circ y) &\leftrightarrow \perp \\ x \wedge (x \bullet y) &\leftrightarrow \perp \\ x \wedge (x \cup y) &\leftrightarrow \perp \end{aligned}$$

The first of these axiom schemas eliminates cyclic feature structures, the second cyclic sequences, the third cyclic “intensional” sequences and the fourth gives wellfounded set theory.

A-6 Other Issues

In this section, we briefly discuss some other aspects of the expressive power of \mathcal{L}^+ and some of its more unusual features.

One perhaps surprising feature of the language is that formulas of the form $(f : \phi) \wedge (a \circ b)$ are welldefined and satisfiable. Some authors have argued for precisely this type of expressiveness. (Cf. e.g., [38].) At any rate, if different operators (i.e., function symbols) are meant to be inconsistent then axiom schema have to be added to rule out such formulas. This leads naturally to considering multi-sorted and order-sorted versions of \mathcal{L}^+ . There is no technical problem with this at all. Functions and relation symbols would be part of a signature specifying the sorts of their arguments, variables would be sorted and the set of worlds in a frame would also be sorted. This is undoubtedly an interesting line of research and there is much relevant computer science literature available. Unfortunately, it is beyond the scope of this appendix to consider such sorted versions of \mathcal{L}^+ .

One issue that hasn't been discussed at all but which has been implicit throughout is the case of nullary function and relation symbols. We have mentioned repeatedly that variables are just special propositional variables. Such propositional variables are just the *sorts* of many other feature value formalisms like Smolka's Feature Logic ([42]). It would have been trivial to add propositional variables since they don't change the proof theory or complicate the class of model structures. However, we haven't bothered because nullary relation symbols act just like propositional variables. Basically, they can be satisfied at more than one world. Nullary function symbols on the other hand can be satisfied at only one world and so they act just like variables. This means that variables could be eliminated from \mathcal{L}^+ . This should not be surprising given the fact that the term schema subsumes the variable schema anyway.

The use of nullary relations makes it very easy to express HPSG-style *type implications*. E.g., consider the following type implication from P&S1.

$$\text{phrasal-sign}[\] \Rightarrow \left[\begin{array}{l} \text{SYN|LOC|SUBCAT } \boxed{2} \\ \text{DTRS } \left[\begin{array}{l} \text{HEAD-DTR|SYN|LOC|SUBCAT } \text{append}(\boxed{1},\boxed{2}) \\ \text{COMP-DTRS } \boxed{1} \end{array} \right] \end{array} \right]$$

This can be translated very simply into \mathcal{L}^+ by using a nullary relation symbol *phrasal-sign* without the need for an elaborate type calculus.

$$\begin{aligned} \text{phrasal-sign} \rightarrow & (\text{syn} : \text{loc} : \text{subcat} : y \wedge \\ & \text{dtrs} : (\text{head-dtr} : \text{syn} : \text{loc} : \text{subcat} : \text{append}(x, y) \wedge \\ & \text{comp-dtrs} : x)) \end{aligned}$$

Finally, relation symbols allow the definition of what might be called "multiple value functions" or possibly "nondeterministic" functions. That is, relations which treat the actual world w at which the relation must be true as an "output variable". However, since a relation can be true at more than one world on the same arguments, different "output values" can be computed. For example, consider the following axioms.

$$\begin{aligned} x \circ \epsilon & \leftrightarrow x \leftrightarrow \epsilon \circ x \\ ((x \circ y) \circ z) \vee (z \circ (x \circ y)) & \leftrightarrow x \circ (y \circ z) \end{aligned}$$

\circ is a binary relation symbol which is nonetheless called the "shuffle operator" in formal language theory. (\circ is meant to be sequence concatenation and ϵ the empty sequence.) What the second axiom says is that three sequences σ_1 , σ_2 and σ stand in the shuffle relation if σ is a sequence of length $|\sigma| = |\sigma_1| + |\sigma_2|$ (where $|\sigma|$ is the length of σ) containing all of the elements of both σ_1 and σ_2 in some order s.t. the original order of the elements in σ_1 and σ_2 are preserved in σ .

For example, let $\sigma_1 = \langle a, b \rangle$ and $\sigma_2 = \langle c, d \rangle$. Then

$$\begin{aligned} \sigma_1 \circ \sigma_2 &\leftrightarrow \langle a, b, c, d \rangle \vee \\ &\langle a, b, c, d \rangle \vee \\ &\langle a, c, b, d \rangle \vee \\ &\langle a, c, d, b \rangle \vee \\ &\langle c, a, b, d \rangle \vee \\ &\langle c, a, d, b \rangle \vee \\ &\langle c, d, a, b \rangle \end{aligned}$$

You will notice that in each of the sequences in the disjunction, a precedes b (as in σ_1) and c precedes d (as in σ_2) and that there are no other possibilities which meet these requirements. The shuffle operator is the “sequence union” or “domain union” operation used to formalise word order variation in German in terms of word order domains (sequences of constituents) in the body of the thesis.

To give an artificially trivial example, consider a grammar for a totally nonconfigurational language with only binary rules. Let the **DOMAIN** attribute (**DOM**) encode the word order domain of phrasal signs. Then the **DOMAIN** attribute of a phrasal sign will be the shuffle of the two daughter **DOMAIN** sequences.

In feature structure notation this would be

$$\left[\begin{array}{l} \text{DTRS} \left\langle \left[\text{DOM } \boxed{1} \right], \left[\text{DOM } \boxed{2} \right] \right\rangle \\ \text{DOM } \boxed{1} \circ \boxed{2} \end{array} \right]$$

In \mathcal{L}^+ , we would write this as

$$dtrs : ((dom : x) \circ (dom : y)) \wedge dom : (x \circ y)$$

If x and y are two element sequences (as in the example above) then \circ will be as if it were a function which (nondeterministically) produces six values. Such “multiple-value” functions are extremely useful for writing concise, perspicuous grammars.

A-7 Nonwellfounded Sets

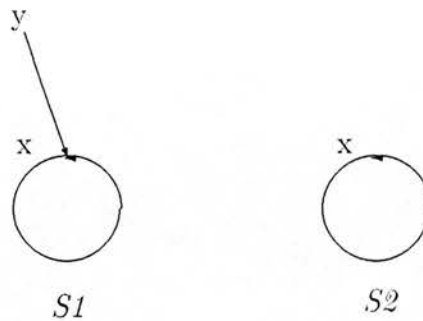
Let $\{\}$ be a ‘wrap-fix’ unary operator which takes a term as its argument and returns a singleton set containing that term.⁵ Then consider the formula $x \wedge \{x\}$.

⁵I am going to have to assume a very basic familiarity with Ch. 1 of [1] in this section.

By the following proof we see that $\vdash x \wedge \{x\} \rightarrow x \wedge \{\{x\}\}$.

1. $x \wedge \{x\}$ hypothesis
2. $x \wedge \{x\} \wedge \{x\}$ idempotence
3. $x \wedge \{x \wedge \{x\}\}$ (A.11)
4. $x \wedge \{x\} \wedge \{\{x\}\}$ (A.13)
5. $x \wedge \{\{x\}\}$ PC

But unlike Aczel's nonwellfounded set theory ([1]) we can't prove $x \wedge \{\{x\}\} \rightarrow x \wedge \{x\}$. We would have to add an axiom corresponding to AFA for this to be valid. AFA is a very strong extensionality axiom. The type of extensionality that we have is somewhat weaker. Consider the following two "pictures" of graphs.



Assume Aczel's pictorial conventions for representing set membership, i.e., a directed arc from node y to node x means that the set x is an element of the set y . Let D be a function from sets into elements of sets. Then, in picture $S1$, $D(y) = x$ and $D(x) = x$. But according to the usual notion of extensionality from classical set theory, two sets are equal iff they contain the same elements. But $D(y) = D(x)$ so we should be able to conclude that $x = y$, or pictorially, that $S1 = S2$. What does this mean informally? It means that a set whose only element is itself is equal to a set whose only element is a set whose only element is itself. That this should be true is not at all intuitively obvious, but this is precisely what classical extensionality requires.

Using $\{\}$ in the same fashion as above, we want to show that $\{x \wedge \{x\}\} \leftrightarrow x \wedge \{x\}$. Can we? In fact, we can by the following proof. It is easier to think about if we consider one direction at a time. Let's first consider the right to left half.

1. $x \wedge \{x\}$ hypothesis
2. $x \wedge \{x\} \wedge \{x\}$ idempotence
3. $x \wedge \{x \wedge \{x\}\}$ (A.11)
4. $\{x \wedge \{x\}\}$ PC

Now let's see the left to right half.

1. $\{x \wedge \{x\}\}$ hypothesis
2. $\{x \wedge x \wedge \{x\}\}$ idempotence
3. $\{x\} \wedge \{x \wedge \{x\}\}$ (A.13)
4. $\{x\} \wedge x \wedge \{\{x\}\}$ (A.12)
5. $x \wedge \{x\}$ PC

So, we see that function symbols have the weak form of extensionality that one would expect even in the context of intensional, nonwellfounded set theory and even in a language which can express identity.

Again, one might object that since the biconditional holds the two formulas $\{x \wedge \{x\}\}$ and $x \wedge \{x\}$ have the same satisfying models and so we really haven't reasoned about $S1$ and $S2$ at all. The thing that must be remembered is that $S1$ and $S2$, although having the appearance of graphs, are not model structures. They are just elements of a mathematical domain that we have modelled in \mathcal{L}^+ . The fact that they have the same satisfying models doesn't matter. In fact, any time we model an algebraic domain in which distinct terms can be equal we will encounter the same situation. This is not an indication that we are doing something wrong.

In summary, we have modelled a form of nonwellfounded set theory with a weak extensionality axiom which is consistent and, rather surprisingly, has the behaviour that should be expected. This is surprising considering the rather humble origins of \mathcal{L}^+ as a formalisation of functional and relational dependencies as found in the unification-based grammar formalism literature.

A-8 The translation of the HPSG formalism into \mathcal{L}^+

In this section, we will consider the translation of the HPSG formalism into \mathcal{L}^+ . This consists of two parts. First we consider the translation of individual formulas of the HPSG formalism into \mathcal{L}^+ in §A-8.1. Second, we separately consider the translation of axioms of the HPSG formalism into \mathcal{L}^+ in §A-8.2.

A-8.1 The translation of the feature structure notation of HPSG into \mathcal{L}^+

We will consider the translation of the HPSG formalism into \mathcal{L}^+ by cases. We define a translation function $\tau : \text{HPSG} \rightarrow \mathcal{L}^+$ as follows.

Let $\tau([\]) = \top$.

Let c be a constant or atom. Then $\tau(c) = c$.

Let s be a sort symbol and a nullary relation symbol of \mathcal{L}^+ and v an arbitrary formula of HPSG. Then

$$\tau(s^v) = s \wedge \tau(v)$$

That is, sorts are translated as nullary relation symbols. (Functional sorts or nullary function symbols are not used in HPSG.)

Let n be a positive integer, v be an arbitrary formula of HPSG and x_1, x_2, \dots be variables of \mathcal{L}^+ . Then

$$\tau(\boxed{x}_n) = x_n$$

and

$$\tau(\boxed{x}_n v) = x_n \wedge \tau(v)$$

Let f, f_1, \dots, f_n be features and v, v_1, \dots, v_n be arbitrary formulas of HPSG. Then

$$\tau\left(\left[\begin{array}{c} f \quad v \end{array} \right]\right) = f : \tau(v)$$

$$\tau\left(\left(\left[\begin{array}{cc} f_1 & v_1 \\ \vdots & \vdots \\ f_n & v_n \end{array} \right]\right)\right) = f_1 : \tau(v_1) \wedge \dots \wedge f_n : \tau(v_n)$$

Let v_1, \dots, v_n be arbitrary formulas of HPSG. Then

$$\tau(v_1 \wedge \dots \wedge v_n) = \tau(v_1) \wedge \dots \wedge \tau(v_n)$$

Let v_1, \dots, v_n be arbitrary formulas of HPSG. Then

$$\tau(v_1 \vee \dots \vee v_n) = \tau(v_1) \vee \dots \vee \tau(v_n)$$

Let v be an arbitrary formula of HPSG. Then

$$\tau(\neg v) = \neg \tau(v)$$

Let g be a relation symbol and v_1, \dots, v_n be arbitrary formulas of HPSG. Then

$$\tau(g(v_1, \dots, v_n)) = g(\tau(v_1), \dots, \tau(v_n))$$

A-8.2 The translation of axioms in HPSG into \mathcal{L}^+

Introduction

One question which arises in the HPSG notation is the treatment of formulas in feature structure notation which are intended to be interpreted as logical *axioms*. For example, consider the following definition of the relation delete repeated here from §3-10.

(A.17) delete

$$\begin{aligned} \text{delete}(\boxed{1}, \langle \boxed{1} \circ \boxed{2} \rangle) &\leftrightarrow \boxed{2} \\ \text{delete}(\boxed{1}, \langle \boxed{2} \circ \boxed{3} \rangle) &\leftrightarrow \\ &\langle \boxed{2} \rangle \circ \text{delete}(\boxed{1}, \boxed{3}) \end{aligned}$$

First, it should be recalled that such definitions are disjunctive. That is, the pair of axioms is satisfied if and only if either one of them is. This means that the translation of the pair of axioms is a disjunction of the translation of each of the axioms as follows.

$$(\text{delete}(x_1, \langle x_1 \rangle \circ x_2) \leftrightarrow x_2) \vee (\text{delete}(x_1, \langle x_2 \rangle \circ x_3) \leftrightarrow \langle x_2 \rangle \circ \text{delete}(x_1, x_3))$$

However, this is not sufficient. As in definite clause logic [20], each of the disjunctions must be interpreted as universally prefixed formulas. Thus, the correct translation is

$$\begin{aligned} \forall x_1, x_2. (\text{delete}(x_1, \langle x_1 \rangle \circ x_2) \leftrightarrow x_2) \vee \\ \forall x_1, x_2, x_3. (\text{delete}(x_1, \langle x_2 \rangle \circ x_3) \leftrightarrow \langle x_2 \rangle \circ \text{delete}(x_1, x_3)) \end{aligned}$$

\mathcal{L}^+ does not contain universal quantifiers though. Therefore, we must augment \mathcal{L}^+ with universal quantifiers to correctly interpret formulas of feature structure notation intended as axioms. We will now define a language \mathcal{L}^{++} which is \mathcal{L}^+ with quantifiers.

Syntax

First, some notation. $\phi(x_1, \dots, x_n)$ indicates that ϕ is a formula with free variables x_1, \dots, x_n . A frequent notational convention will be to refer to a formula $\phi(x)$ and then to $\phi(y)$ which indicates the formula formed by substituting y for every free occurrence of x in ϕ . (I assume that the reader is familiar with the notion of free variable and variable substitution.)

Let X be a countably infinite set of *variables* and F a set of *features*. Let Rel be a set of *relation symbols*. Let $Fun \subseteq Rel$ be a set of *function symbols*. For each

$g \in Rel$, let there be a unique nonnegative integer called its *arity*. Then \mathcal{L}^{++} is the smallest set s.t.

$$\begin{array}{ll}
\top \in \mathcal{L}^{++} & \\
\perp \in \mathcal{L}^{++} & \\
x \in \mathcal{L}^{++} & \forall x \in X \\
f : \phi \in \mathcal{L}^{++} & \forall f \in F, \phi \in \mathcal{L}^{++} \\
\phi \wedge \psi \in \mathcal{L}^{++} & \forall \phi, \psi \in \mathcal{L}^{++} \\
\phi \vee \psi \in \mathcal{L}^{++} & \forall \phi, \psi \in \mathcal{L}^{++} \\
\neg \phi \in \mathcal{L}^{++} & \forall \phi \in \mathcal{L}^{++} \\
g(\phi_1, \dots, \phi_n) \in \mathcal{L}^{++} & \forall n\text{-ary } g \in Rel, \forall \phi_1, \dots, \phi_n \in \mathcal{L}^{++} \\
\exists x. \phi \in \mathcal{L}^{++} & \forall x \in X, \phi \in \mathcal{L}^{++} \\
\forall x. \phi \in \mathcal{L}^{++} & \forall x \in X, \phi \in \mathcal{L}^{++}
\end{array}$$

Semantics

Model structures of \mathcal{L}^{++} are slightly different than those of \mathcal{L}^+ . A model \mathcal{M} is a frame defined for \mathcal{L}^+ .

Let \mathcal{M} be a model and $x, y \in X$. Let $V[w/x](y) = x$ if $y = x$ and $V[w/x](y) = V(y)$ if $y \neq x$.

Then satisfiability of \exists and \forall is defined as follows.

$$\begin{array}{l}
(\mathcal{M}, V) \models_w \exists x. \phi(x) \Leftrightarrow (\mathcal{M}, V[w'/x]) \models_w \phi(x) \text{ for some } w' \in W \\
(\mathcal{M}, V) \models_w \forall x. \phi(x) \Leftrightarrow (\mathcal{M}, V[w'/x]) \models_w \phi(x) \text{ for all } w' \in W
\end{array}$$

Theorem A.4 *The following axioms are sound in \mathcal{L}^{++} .*

$$\begin{array}{l}
Def \exists \quad \exists x. \phi(x) \Leftrightarrow \neg \forall x. \neg \phi(x) \\
Def \forall \quad \forall x. \phi(x) \Leftrightarrow \neg \exists x. \neg \phi(x)
\end{array}$$

Theorem A.5 *The proof theory for \mathcal{L}^+ plus the following schemas are a sound and complete proof theory for \mathcal{L}^{++} .*

$$\begin{array}{l}
\forall 1 \quad \forall x. \phi(x) \rightarrow \phi(y) \\
\forall 2 \quad \frac{\psi \rightarrow \phi(y)}{\psi \rightarrow \forall y \phi(y)} \text{ for } y \text{ not free in } \psi \\
\forall 3 \quad \phi \leftrightarrow \psi \text{ iff } \forall x. \phi \leftrightarrow \forall x. \psi
\end{array}$$

The dual counterparts for \forall are easily interderivable.

Theorem A.6 *The proof theory for \mathcal{L}^+ plus the following schemas are a sound and complete proof theory for \mathcal{L}^{++} .*

$$\begin{array}{l}
\exists 1 \quad \phi(y) \rightarrow \exists x. \phi(x) \\
\exists 2 \quad \frac{\phi(y) \rightarrow \psi}{\exists y \phi(y) \rightarrow \psi} \text{ for } y \text{ not free in } \psi \\
\exists 3 \quad \phi \leftrightarrow \psi \text{ iff } \exists x. \phi \leftrightarrow \exists x. \psi
\end{array}$$

Bibliography

- [1] Peter Aczel. *Non-Well-Founded Sets*. Center for the Study of Language and Information, Stanford, Ca., 1988.
- [2] Renate Bartsch. *The Grammar of Adverbials*. Number 16 in Linguistics Series. North-Holland, Amsterdam, 1976.
- [3] Gunnar Bech. *Studien über das Deutsche Verbum Infinitum, Band I*. Copenhagen, 1955.
- [4] Joan Bresnan, editor. *The Mental Representation of Grammatical Relations*. MIT Press, Cambridge, Massachusetts, 1982.
- [5] Joan Bresnan, Ron Kaplan, Stanley Peters, and Annie Zaenen. Cross-serial dependencies in dutch. *Linguistic Inquiry*, 5:614–619, 1982.
- [6] Noam Chomsky. *Syntactic Structures*. Janus, 1957.
- [7] Kathrin Cooper. Word order in bare infinitival complement constructions in swiss german. Master's thesis, Centre for Cognitive Science, University of Edinburgh. Edinburgh, 1988.
- [8] Kathrin Cooper. Zurich german z and verb raising constructions. In Martin Mellor Elisabet Engdahl, Mike Reape and Richard Cooper, editors, *Parametric Variation in Germanic and Romance: Proceedings from a DYANA Workshop*, pages 76–85, Edinburgh, 1990. Centre for Cognitive Science, University of Edinburgh.
- [9] Haskell B. Curry. Some logical aspects of grammatical structure. structure of language and its mathematical aspects. In Jacobson, editor, *Proceedings of the Twelfth Symposium in Applied Mathematics*, pages 56–68. American Mathematical Society, 1963.
- [10] Hans den Besten and Jerry Edmondson. The verbal complex in continental west germanic. In Werner Abraham, editor, *On the Formal Syntax of the Westgermania*, pages 155–216. Benjamins, Amsterdam, 1983.
- [11] Hans den Besten and Gert Webelhuth. Stranding. Manuscript. Universiteit Amsterdam and University of Massachusetts, 1988.

- [12] David Dowty. Towards a minimalist theory of syntactic structure. In Wietske Sijsma and Arthur van Horck, editors, *Discontinuous Constituency*. Mouton de Gruyter, Berlin, 1992. Also appears in the *Proceedings of the Symposium on Discontinuous Constituency*, The Institute for Language Technology and Artificial Intelligence, Tilburg, The Netherlands. January 25-27, 1990, pp33-72.
- [13] Arnold Evers. *The Transformational Cycle in Dutch and German*. PhD thesis, University of Utrecht, Utrecht, 1975.
- [14] Gerald Gazdar, Ewan Klein, Geoffrey K. Pullum, and Ivan A. Sag. *Generalised Phrase Structure Grammar*. Blackwell and Harvard University Press, Cambridge, England and Cambridge, Massachusetts, 1985.
- [15] Liliane Haegeman and Henk van Riemsdijk. Verb projection raising, scope, and the typology of rules affecting verbs. *Linguistic Inquiry*, 17(3):417-466. 1986.
- [16] Hubert Haider. Topicalization and other puzzles of german syntax. In G. Grewendorf and W. Sternefeld, editors, *Scrambling and Barriers*. Benjamins, Amsterdam, 1987.
- [17] E. Hinrichs and T. Nakazawa. Flipped out: Aux in german. In *The Proceedings of the 25th Regional Meeting of the Chicago Linguistics Society*, Chicago, 1989. University of Chicago.
- [18] Jack Hoeksema. Partiële extrapositie van infinitiefkomplementen in het midlenederlands. *TABU*, 20:156-165, 1990.
- [19] Jack Hoeksema. Suppression of a word-order pattern in westgermanic. Unpublished manuscript, University of Groningen. Paper delivered at ICHL., 1991.
- [20] Markus Höhfeld and Gert Smolka. Definite relations over constraint languages. Technical Report 53, IBM Deutschland, Stuttgart, Germany, October 1988.
- [21] John E. Hopcroft and Jeffrey D. Ullman. *Introduction to Automata Theory, Languages and Computation*. Addison-Wesley, Menlo Park, California, 1979.
- [22] Mark E. Johnson. *Attribute-Value Logic and the Theory of Grammar*. Number 15 in CSLI Lecture Notes. CSLI, Stanford, California, 1987.
- [23] Lauri Karttunen. Radical lexicalism. In Mark Baltin and Anthony Kroch, editors, *Alternative Conceptions of Phrase Structure*, pages 43-65. University of Chicago Press, Chicago, 1989.
- [24] Robert T. Kasper. *Adjuncts in the Mittelfeld*. chapter ??, page ?? University of Chicago Press, Chicago, Illinois. 1992. To appear.

- [25] Robert T. Kasper and William C. Rounds. A logical semantics for feature structures. In *Proceedings of the 24th Annual Meeting of the Association for Computational Linguistics*, pages 257–266, New York, N.Y., 10–13 June 1986. Columbia University.
- [26] Angelika Kratzer. On deriving syntactic differences between german and english. unpublished manuscript., 1984.
- [27] Terence McKay. *Infinitival Complements in German: Lassen, scheinen and the verbs of perception*. Cambridge University Press, 1985.
- [28] Klaus Netter. Clause union phenomena and complex predicates in german. In Klaus Netter and Mike Reape, editors, *Clause Structure and Word Order Variation in Germanic*, number R1.1.B in DYANA Report, pages –. University of Edinburgh, Edinburgh, January 1991.
- [29] Karel Oliva. Remarks on two approaches to word order variation in german. Unpublished manuscript, 1992.
- [30] Carl Pollard. On non-head movement. In Wietske Sijsma and Arthur van Horck, editors, *Discontinuous Constituency*. Mouton de Gruyter, Berlin, 1992. Also in *Proceedings of the Symposium on Discontinuous Constituency*, The Institute for Language Technology and Artificial Intelligence, Tilburg, The Netherlands, January 25-27, 1990, pp125-137.
- [31] Carl Pollard and Ivan Sag. *An Information-Based Approach to Syntax and Semantics: Volume 1 Fundamentals*, volume 13 of *CSLI Lecture Notes Series*. CSLI, Stanford, California, 1987.
- [32] Carl Pollard and Ivan Sag. *An Information-Based Approach to Syntax and Semantics: Volume 2 ???*, volume 1? of *CSLI Lecture Notes Series*. CSLI, Stanford, California, 1992.
- [33] Geoffrey K. Pullum. Free word order and phrase structure rules. In James Pustejovsky and Peter Sells, editors, *Proceedings of the 12th Annual Meeting of the Northeastern Linguistic Society*, pages 209–220. Amherst, 1982. Department of Linguistics, University of Massachusetts.
- [34] Mike Reape. An introduction to the semantics of unification-based grammar formalisms. DYANA Deliverable R3.2.A, Centre for Cognitive Science, University of Edinburgh, Edinburgh, 1991.
- [35] Mike Reape. A feature value logic with intensionality, nonwellfoundedness and functional and relational dependencies. In *Constraint Propagation*. Academic Press, 1993.
- [36] Mike Reape and Mark Hepple. Word order and constituency in west continental germanic. In *Parametric Variation in Germanic and Romance: Preliminary Investigations*, number R1.1.A in DYANA Report, pages –. University of Edinburgh, Edinburgh, January 1990.

- [37] Marga Reis. Reflexivierung in deutschen aci-konstruktionen. ein transformationsgrammatisches dilemma. In *Papier zu Linguistik*, volume 9, pages 5–82. Springer-Sciptor, Kronberg, 1976.
- [38] William C. Rounds. Set values for unification-based grammar formalisms and logic programming. Technical report, CSLI, 1989.
- [39] Ineke Schuurman. Some particular verb-raising phenomena. In Elisabet Engdahl, Mike Reape, Martin Mellor, and Richard Cooper, editors, *Parametric Variation in Germanic and Romance: Proceedings from a DYANA Workshop*, Edinburgh Working Papers in Cognitive Science, pages 199–207, Edinburgh, September 1989. Centre for Cognitive Science, University of Edinburgh.
- [40] Peter Sells. *Lectures on Contemporary Syntactic Theories*. Number 3 in CSLI Lecture Notes. CSLI, Stanford, California, 1985.
- [41] Jörg H. Siekmann. Universal unification. In R.H. Shostak, editor, *Proceedings of the Seventh International Conference on Automated Deduction*. Lecture Notes in Computer Science, pages 1–42, Napa, California, May 1984. Springer-Verlag.
- [42] Gert Smolka. A feature logic with subsorts. Technical Report 33, IBM Deutschland, Stuttgart, Germany, 1988. To appear in the proceedings of the workshop "Unification Formalisms – Syntax, Semantics and Implementation", Titisee, West Germany. April, 1988.
- [43] Wolfgang Sternefeld. *Bausteine Syntaktischen Wissens. Ein Lehrbuch der Generativen Grammatik*. Westdeutscher Verlag, Opladen, 1988.
- [44] Hans Uszkoreit. Linear precedence in discontinuous constituents: Complex fronting in german. In G. Huck and A. Ojeda, editors, *Discontinuous Constituency*, number 20 in Syntax and Semantics, pages 405–25. Academic Press, Orlando, Florida, 1987.
- [45] Hans Uszkoreit. *Word Order and Constituent Structure in German*, volume 8 of CSLI Lecture Notes. CSLI, Stanford, California, 1987.
- [46] Arnim von Stechow. Status government and coherence in german. In G. Grewendorf and W. Sternefeld, editors, *Scrambling and Barriers*. Benjamin, Amsterdam, 1990.
- [47] Hendrik Willem Zeevat. *Aspects of Discourse Semantics and Unification Grammar*. PhD thesis. Universiteit van Amsterdam, Amsterdam, 1991.
- [48] Henk Zeevat, Ewan Klein, and Jonathan Calder. Unification categorial grammar. In Ewan Klein Nick Haddock and Glyn Morrill, editors, *Categorial Grammar. Unification Grammar and Parsing*, number 1 in Edinburgh Working Papers in Cognitive Science. Center for Cognitive Science, University of Edinburgh, Edinburgh, 1987.

- [49] Arnold Zwicky. Concatenation and liberation. In Anne M. Farley et al., editor, *Proceedings of the 22nd Regional Meeting of the Chicago Linguistics Society*, pages 65-74. Chicago, 1986. Chicago Linguistic Society.