Continuous or Discontinuous Constituents?

Stefan Müller
DFKI GmbH
Stuhlsatzenhausweg 3
66123 Saarbrücken
`Stefan.Mueller@dfki.de`

1 Introduction

During the last years, several grammarians have argued for linguistic descriptions of language that use the concept of discontinuous constituents (Reape, 1991, 1992, 1994; Pollard, Kasper and Levine, 1992, 1994; Kathol and Pollard, 1995; Kathol, 1995; Müller, 1995, 1997, 1999a; Richter and Sailer, 1999; Donohue and Sag, 1999; Penn, 1999).

Reape (1991) notes that it is possible to develop grammars based on the concept of discontinuous constituents which span every subset of an input string. The complexity of the paring problem for such grammars is at least exponential in both time and space. As Reape (1991, p. 62) has argued for the processing of grammars with discontinuous constituents and as Carroll (1994) has demonstrated for different parsing strategies for grammars with continuous constituents, such theoretical values are not of much help when it comes to practical systems. In the following I will compare two grammars for German: one that assumes continuous constituents and one that allows for discontinuous constituents. I will show different ways to account for the German data and I will explain why I believe that grammars with discontinuous constituents are more appropriate for systems that analyze German sentences. This claim will be supported by a statistics computed from 10,000 utterances taken from the Verb*mobil* corpus.[1]

2 The Phenomena

2.1 Relatively Free Constituent Order in the *Mittelfeld*

In German, complements and adjuncts of a head can be ordered relatively freely. For instance, with a ditransitive verb like *geben* (*give*), all six permutations of the arguments are possible, provided appropriate context and intonation. In general the possability to permute constituents depends on a broad variety of interacting constraints as for instance animateness, heavyness, definiteness. See for instance (Behagel, 1930; Drach, 1937; Hoberg, 1981; Höhle, 1982; Uszkoreit, 1987).

[†]Thanks to Dan Flickinger, Bob Kasper, Detmar Meurers, Gerald Penn, and anonymous reviewers for discussion and/or comments on earlier versions of this paper.

This work was supported by the Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie (BMBF) (Project Verbmobil, Part 2.4, Grant Number 01 IV 701 V0).

[1]On Verb*mobil* see Burgard (To Appear).

(1) a. Deshalb gab der Mann der Frau das Buch.
therefore gave the man the woman the book
'Therefore the man gave the woman the book.'
b. Deshalb gab der Mann das Buch der Frau.
c. Deshalb gab das Buch der Mann der Frau.
d. Deshalb gab das Buch der Frau der Mann.
e. Deshalb gab der Frau der Mann das Buch.
f. Deshalb gab der Frau das Buch der Mann.

Verbs like *kaufen* (*buy*) take four arguments (see Kunze (1991)), and as Wegener (1985) argued convincingly, some of the so-called "free datives" have to be analyzed as complements as well.

(2) Deshalb kauft Karl von Hans für fünf Mark seiner Frau ein Buch.
therefore buys Karl from Hans for five Marks his wife a book
'Therefore Karl buys a book for his wife from Hans for five Marks.'

In principle, all permutations of the five arguments of *kaufen* are possible. For sentences with five arguments the number of possible permutations is $5! = 120$.

2.2 Verb Position

There are three possible verb positions in German sentences: verb first position (3a), verb second position (3b), and verb last position (3c).

(3) a. Liebt der Mann die Frau?
loves the man the woman
'Does the man love the woman?'
b. Der Mann liebt die Frau.
The man loves the woman.'
c. daß der Mann die Frau liebt.
'that the man loves the woman.'

Verb second sentences are usually analyzed as derived from verb first sentences by the fronting of one constituent.

2.3 The Predicate Complex

Verbs that embed an infinitive without *zu* and verbs that select for participles form a complex with their verbal complement (Hinrichs and Nakazawa, 1989a). Furthermore, some of the verbs that select an infinitive with *zu* form a complex (Kiss, 1995).

(4) a. Hat er den Mann der Frau das Buch von Karl für fünf Mark kaufen lassen?
has he the man the woman the book from Karl for five Mark buy let
'Did he let the man buy the book for the woman from Karl for five marks?'
b. weil es ihm jemand zu lesen versprochen hat.
because it him somebody to read promised has
'since somebody promised him to read it.'

Hinrichs and Nakazawa (1989a) suggested analyzing these verbal complexes via argument attraction, essentially a lexical variant of a functional composition combining the two verbal functors. For (4b) this means that *zu lesen* and *versprochen* form a verbal complex. The complex inherits all arguments of the verbs that are involved in the complex formation, i.e., for the complex *zu lesen versprochen* we have *jemand*, *ihm*, and *es* as arguments. Since all arguments are dependents of the verbal complex, it can be explained why these elements can be permuted in the same way as normal complements of one single verb can be (see the discussion of (1)).[2] Note that the number of arguments of the complex *kaufen lassen hat* is six. In principle, this addition of arguments can be iterated.

(5) weil Hans Cecilia John das Nilpferd füttern helfen läßt.
because Hans Cecilia John the hippo feed help let
'because Hans lets Cecilia help John feed the hippo.'

Füttern is a transitive verb. *Helfen* takes a subject, a dative NP and a verbal complex. The complex *füttern helfen* has three arguments. *Lassen* takes a subject, an accusative object (which is linked to the subject of *füttern helfen*) and a verbal complex. The complex *füttern helfen läßt* has four arguments.

Restricting the number of complements that a verbal complex may take is no less ad hoc than limiting the number of center self-embeddings. Restrictions on the number of arguments should not be part of a competence grammar.

2.4 Extraposition

Extraposition is a dependency, which is nonlocal in nature, even if the the constraints which are imposed on this dependency differ from those found with fronting as the classical nonlocal dependency:

(6) a. Karl hat mir [ein Bild [einer Frau $\_i$]] gegeben,
[die schon lange tot ist]$_i$.
b. Karl hat mir [eine Fälschung [des Bildes [einer Frau $\_i$]]] gegeben,
[die schon lange tot ist]$_i$.
c. Karl hat mir [eine Kopie [einer Fälschung [des Bildes [einer Frau $\_i$]]]] gegeben,
[die schon lange tot ist]$_i$.
'Karl gave me a copy of a forgery of the picture of a woman who has been dead for a long time ago.'

Relative clauses can be extraposed from an arbitrarily deeply embedded NP. The same holds for complement clauses, as I have shown in (Müller, 1999a, Ch. 13.1.1).

The number of extraposed constituents is unlimited in principle, as (7) shows.

(7) Ich habe gearbeitet [an diesem Abend] [in der Kneipe] [als Kellnerin].[3]
I have worked at this evening in the pub as barmaid
'I worked as a barmaid in the pub that evening.'

[2] The possibility to scramble the complements of the embedded verb with the arguments of the matrix verb has been denied for various classes of embedded verbs. However, that basically all permutations are possible becomes clear if one takes the data that was discussed by Bech (1955, p. 136), Bierwisch (1963, p. 125), Jacobs (1991, p. 20), and Haider (1991, p. 5). The permutation of elements with the same morphological case is restricted by performance factors in both simplex and complex clauses. See Kuno (1980, p. 175) for Japanese and Müller (1999a, p. 172–173) for German. A detailed discussion of predicate complex constructions and more references can be found in (Müller, In Preparation).

[3] Spiegel, 23/1997, p. 122

In (7), three PPs are extraposed. The PPs are adjuncts and are as such not subcategorized by the verb, i.e., they are not predictable.[4] See section 3.1.4 for the consequences of this fact for some analyses.

3 Continuous Constituents

3.1 Linguistic Theory

Like GPSG, HPSG divides the grammar into Immediate Dominance and Linear Precedence rules (LP rules). In GPSG, LP statements play the role of constraining order in local trees. While in HPSG larger linearization domains are possible, as is discussed in section 4, most HPSG publications have implicitly adopted the GPSG conception of applying LP rules to local trees only.

3.1.1 Relative Free Constituent Order in the *Mittelfeld*

3.1.1.1 Flat Structures

To account for the constituent order freedom in (1), Uszkoreit (1987) suggested a flat structure. Since all complements in figure 1 are in the same local tree, they can be permuted as long as no LP rule is violated.

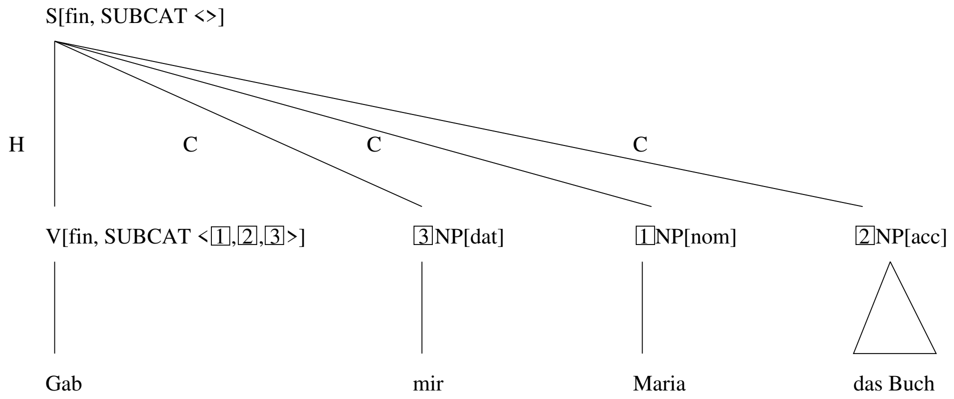

Figure 1: Flat Structure: *Gab mir Maria das Buch?*

The problem with this approach is that the number of rules that is needed is quite big. There have to be rules for intransitive verbs, for transitive verbs, for ditransitive verbs, and for verbs with four arguments. If the verb appears in initial position, it is possible that there is a verbal complex at the right periphery of the clause. In order to account for this, the number of rules has to be increased again. In German, adverbs can be placed anywhere between the complements. The number of adverbs is not restricted. If this has to be reflected in the grammar rules, the number of rules is infinite. Even if one restricts the number of adverbs in an ad hoc way, the set of rules will be huge.

[4]Note that some recent versions of HPSG treat adjuncts as complements. Adjuncts are introduced into the subcat list of their head by a lexical rule (van Noord and Bouma, 1994). With such an approach the problem is shifted to another place and lazy evaluation techniques are needed to process the grammar. Lazy evaluation means to wait with the executation in one branch of the program until enough information is available. Not all systems that are used to process HPSG-like grammars have built in machinary that allows for lazy evaluation.

3.1.1.2 Binary Branching Structures

The problems that one has with flat structures disappear if one uses binary branching structures. A head is combined with one complement at one projection step. At each projection step it is possible to combine the projection with an adjunct. So it is trivial to account for the free appearance of adjuncts in the German *Mittelfeld*.

However, it is not trivial to account for the free ordering of complements. Since *er*, and *das Buch*, and

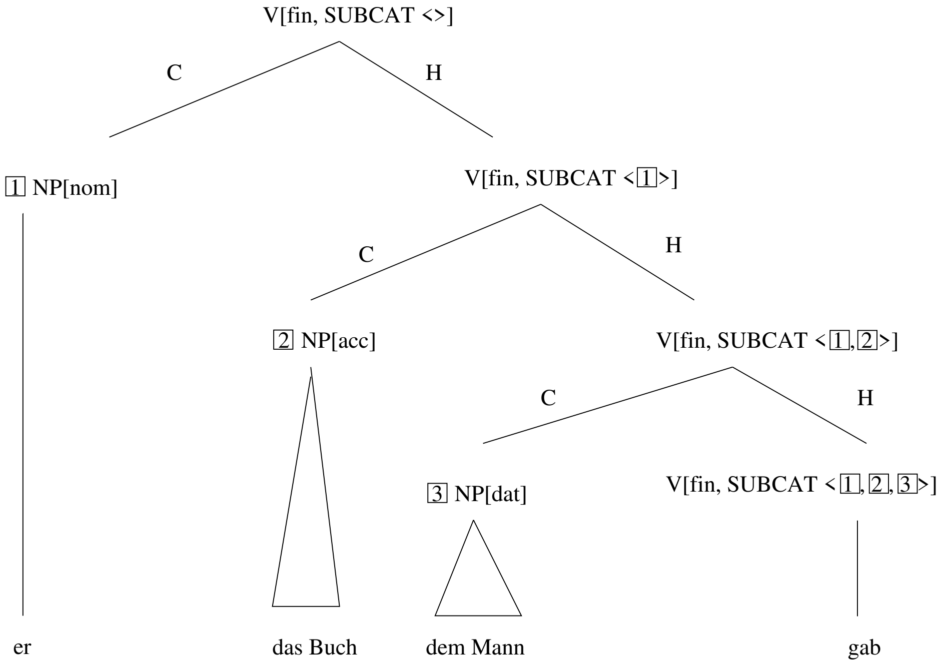

Figure 2: Binary Branching Structure: *er das Buch dem Mann gab.*

dem Mann are not sisters in a local tree, they cannot be permuted freely. There are several solutions to this problem: Gunji (1986), Hinrichs and Nakazawa (1989b), Pollard (1990), and Engelkamp, Erbach and Uszkoreit (1992) suggested using a set rather than a list to represent valence information. In head-complement structures, two elements can be combined if the complement is an element of the subcat set of the head.

The problem with such an approach is that one gets spurious ambiguities for constructions where the head is in the middle. An example would be the head noun in an NP analysis.

(8) a. der Entschluß zu gehen
the decision to go

b. seine Behauptung, daß Gysi ein Spitzel war
his claim that Gysi a spy was
'his claim that Gysi was a spy'

c. das Bild von Maria
the picture of Maria

In sentences like (8) the head noun could be combined with its complement (*zu gehen*, *daß Gysi ein Spitzel war*, *von Maria*) or with the deteminer (*der*, *das*) or subject (*seine*) first. To solve this problem one had to assume either a DP analysis, i.e., an analysis where the determinator and the subject is the head, or introduce a special valence feature for determiners and subjects of deverbal NPs.

Another example of a case where spurious ambiguities arise is the conjunction in coordinated structures, if they are treated as the head of the construction, as suggested by Paritong (1992).

A further problem with the subcat set approach is that one needs a second list for principles and mechanisms that rely on the order of the subcat list, as for instance the Binding Theory, some versions of Case Principles (Heinz and Matiasek, 1994; Müller, To appear), and other prominence related phenomena like *Vorfeldellipse* (so-called 'Topic Drop') (Fries, 1988) and free relative clauses (Müller, 1999b).

An alternative with almost the same problems is a relaxation of the subcat principle. If one allows the subcat principle to take an arbitrary element from the subcat list, it is not possible anylonger to formulate principles that refere to elements at a certain position of the subcat list of a phrase. As Detmar Meurers pointed out, that with this approach principles that rely on the order of the subcat list possibly may be formulated at word level, since there the order information is undisturbed.

Uszkoreit (1986) suggested using a lexical rule that for each verb licences lexical entries with permuted elements in the subcat list for all possible permutations of these elements. This means that at least six lexical entries are licensed for a ditransitive verb like *geben*.[5] This considerably increases the lexical ambiguity. Furthermore the approach has problems with spurious ambiguities that cannot be solved without stipulatations. To see this consider the example in (9).

(9) Dem Mann gab er das Buch.
the man$_{dat}$ gave he$_{nom}$ the book$_{acc}$
'He gave the man the book.'

The fronting of constituents in German is usually analyzed in terms of a nonlocal dependency: A complement of the verb is removed from the subcat list by the saturation by a trace, a unary projection, or by a lexical rule. The complement is introduced into a list (SLASH) and percolated up the tree and then realized to the left of the finite verb. The problem with (9) is that it can be analyzed with two lexical entries that have the subcat lists shown in (10a) and (10b).

(10) a. $\left\langle \text{NP}[\textit{nom}], \text{NP}[\textit{dat}], \text{NP}[\textit{acc}] \right\rangle$
b. $\left\langle \text{NP}[\textit{nom}], \text{NP}[\textit{acc}], \text{NP}[\textit{dat}] \right\rangle$

Only the *nom* and *acc* are realized to the right of the finite verb and *nom* and *acc* have the same order in both orderings in (10). With an analysis of extraction based on lexical rules an order for rule application could be stipulated, i.e., the extraction lexical rule could be restricted to apply before the permutation lexical rule. With the other two approaches one is forced to assume exception features that block the extraction of an element that was permuted by the lexical rule.

Note, that the problem would disappear if permuted elements were extraction islands as is assumed in various GB publications. Then all reordered elements could be marked as islands simultaneously blocking the extraction of the element in the subcat list itself. I have shown in (Müller, 1999a, p. 101) that permuted elements are not islands. Therefore the spurious ambiguities in (9) can only be avoided by ad hoc features.

[5]Uszkoreit (1986) assumes even more than six lexical entries for *geben* since his lexical rule instantiates all features that are relevant for linearization. All these different permutations and instantiations give rise to 18 different lexical entries for a ditransitive verb.

Depending on the way nonlocal dependencies are introduced, this is a problem for the subcat-set approach and for the approach with the relaxed subcat principle also. If the nonlocal dependency is introduced in syntax, i.e., by a trace or by a unary projection, it is not clear at which position in the tree the introduction has to happen. This basically is equivalent to the problem of linearizing traces in an account that assumes flat structures.

3.1.2 Verb Position

3.1.2.1 Flat Structure

If a flat structure is assumed, the verb is in the same local tree as its adjuncts and complements, and it can be serialized either to the left or to the right of them.

3.1.2.2 Binary Branching Structures

If binary branching structures are assumed, the verb position usually is described by head movement analyses (Kiss and Wesche, 1991; Netter, 1992; Müller and Kasper, To Appear).

To see the motivation for such an analysis, consider the structures for (11) in figure 3.

(11) Bringt Peter morgen die Ladung?
brings Peter tomorrow the load
'Does Peter bring the load tomorrow?'

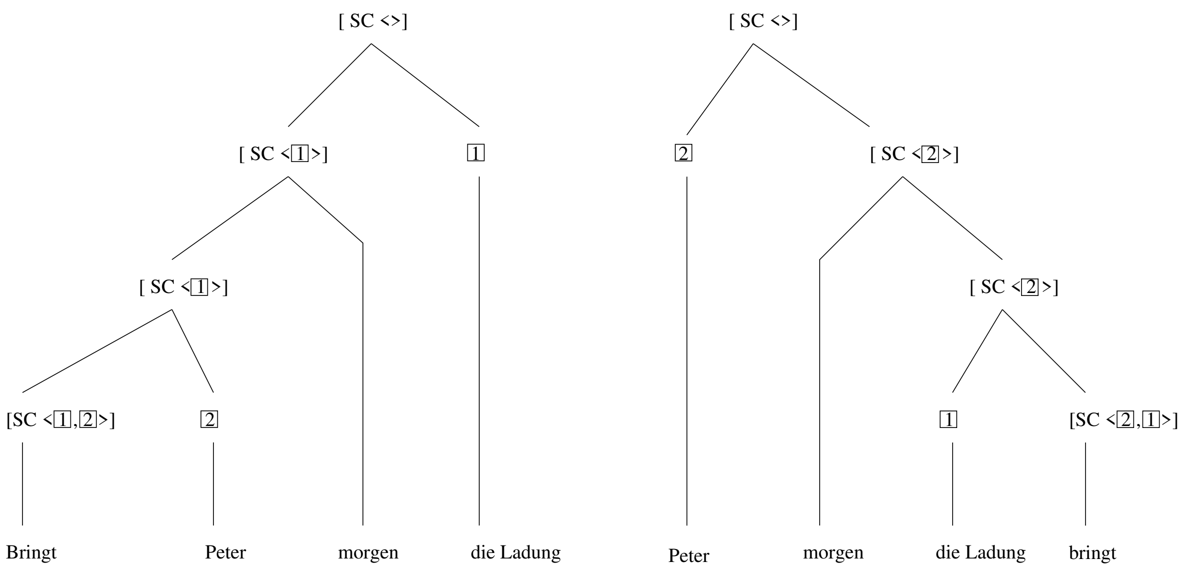

Figure 3: Unmotivated Divergence of Structures for Verb-First and Verb-Last Sentences

Without verb movement one would get two different structures for the sentence, the first structure being the mirror image of the second. The problem with these structures is that adverbs like *morgen* scope over different parts of the tree. Since scope in German is determined from left to right, the different structures in a verb first sentence would make wrong predictions.

(12) a. daß Peter gestern wegen der Konferenz arbeitete.
that Peter yesterday because.of the conference worked

b. daß Peter wegen der Konferenz gestern arbeitete.

c. Arbeitete Peter gestern wegen der Konferenz?

The two sentences in (12a) and (12b) have different readings. The first adjunct always scopes over the second. If (12c) had a structure as is shown in the first tree in figure 3, the reading in (12b) is predicted which is empirically wrong.

To cope with this problem, Netter suggests the analysis shown in figure 4. A verbal trace (13) attracts

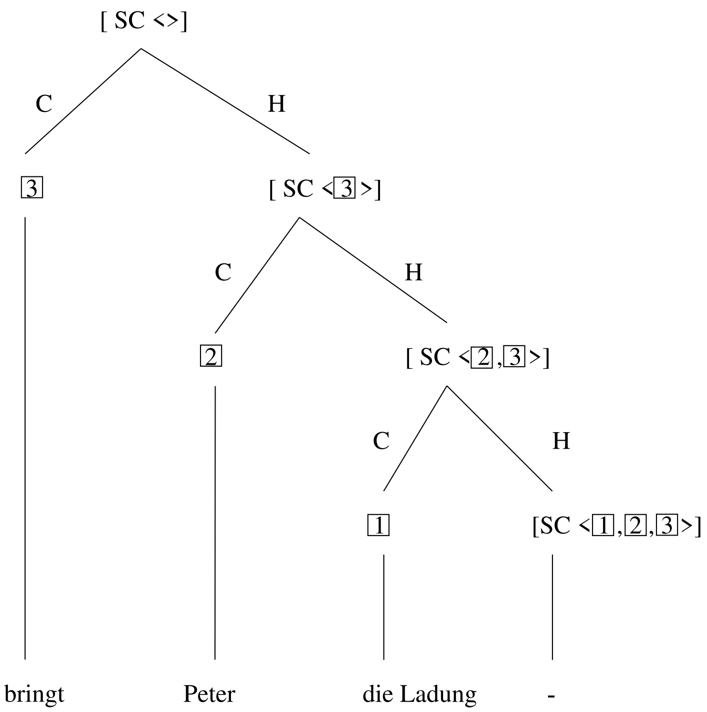

Figure 4: Analysis with Verbal Trace

all arguments of the verb ($\boxed{1}$).[6]

$$(13)\quad \begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \textit{verb} \end{bmatrix} \\ \text{SUBCAT} & \text{V[LEX+, SUBCAT } \boxed{1}\text{, CONT } \boxed{2}\text{]} \oplus \boxed{1} \end{bmatrix} \\ \text{CONT} & \boxed{2} \\ \textit{loc} & \end{bmatrix}$$

The trace functions as the head. It combines with the complements of the verb ($\boxed{1}$ and $\boxed{2}$ in figure 4), and after having done so with the verb itself.

The problem with such head-movement approaches is that the trace is dramatically underspecified. Without further constraints some parsers will build structures that cannot be used in actual analyses

[6] I adapted Netter's trace in a way that the order of elements on the subcat list corresponds to the order that is assumed by Pollard and Sag (1994) and throughout this paper.

Note, that the problem discussed below is not a general problem of argument attraction approaches. The problem arises just in case underspecified valance lists are instantiated by saturation of complements. If the element from which arguments are attracted is combined with its head before any other argument is combined with this head, all valance lists are instantiated.

since these structures do not correspond to linguistic objects. The trace can be combined with determiners and other material that can never be complements of a verb. For a bottom-up parser the search space will be huge. See (Müller, 1999a, p. 180–182) for an example. If the grammar contains other empty elements, the search space will be infinite.

3.1.3 The Predicate Complex

3.1.3.1 Flat Structures

Hinrichs and Nakazawa (1994, p. 11) assume a rule for sentences with a predicate complex that is similar to the following rule.[7]

(14) H[SC <>] $\rightarrow$ H, C^{+}, VC

This rule is an abbreviation of a set of rules (see section 3.2.1.2 below). The C^{+} stands for a set of at least one non-verbal complement. The VC stands for a verbal complex that is built by the verbal complex schema shown in (15).

(15) H $\rightarrow$ H, VC

A sentence with a verbal complex like (4a) is analyzed in the following way: With (15) the verbal complex *kaufen lassen* is built. This verbal complex inherits all arguments from *kaufen* and *lassen*. Then (14) is used to combine *hat* and all complements of *hat* and *kaufen lassen*.

3.1.3.2 Binary Branching Structures

With binary branching structures no additional machinery is necessary. A rule for the verbal complex like the one in (15) is sufficient. The remaining work is done by the verb movement analysis: The verb in first position takes a projection with an empty element that built the verbal complex.

3.1.4 Extraposition

There are several approaches to extraposition in HPSG that treat it as a nonlocal dependency using the same nonlocal mechanism that accounts for extraction to the left[8] (Keller, 1994, 1995; Bouma, 1996). Since extraposition behaves in some respect differently from extraction to the left, a different feature called EXTRA is used for these nonlocal dependencies.

To analyze a sentence like (7) hypotheses are introduced that the verb *gearbeitet* will be combined with adjuncts which are extraposed. This can be done by a trace, by a lexical rule, or by a unary grammar rule. Since the number of extraposed adjuncts is unlimited in principle, a system that parses bottom-up and that introduces such hypotheses would have an infinite search space. Van Noord and Bouma (1994) have shown how a lexical rule based approach can be combined with lazy evaluation techniques in a way that makes such grammars processable.

[7] Hinrichs and Nakazawa (1998, p. 132) assume the rules in (i).

(i) a. H[SC <>] $\rightarrow$ C^{*}, H_{word}
b. H $\rightarrow$ H[VFORM $\neg$ fin], VC

[8] Extraction to the left is usually called topicalization, but for German the term fronting is more appropriate since expletives can be fronted and they are not topics.

3.2 Implementations

Many implemented systems that can process HPSG-like grammars, like for instance, LKB (Copestake, 1999), and PAGE (Uszkoreit et. al., 1994) do not have a linearization component. Grammars that were developed in these systems use a context free phrase structure backbone. An example for such a grammar is the grammar developed by Walter Kasper (semantics) and me (syntax) in the Verb*mobil* project (Müller and Kasper, To Appear).

There are other systems that allow for relational constraints like ALE (Carpenter and Penn, 1996; Penn and Carpenter, 1999), ALEP (Schütz, 1996), and ProFIT (Erbach, 1995, 1998). Examples for grammars in the latter systems are the grammar developed by Meurers (1994) and LS-Gram by Schmidt, Rieder and Theofilidis (1996). As discussed by Meurers (1994, Chapter 4.2) relational constraints can be used to factor out specifications from the individual phrase structure and encode them as part of definite clauses attached to several rules. Apart from using this method to encode generalizations over several rules such as the universal principles of HPSG, it can also be used to express the LP constraints in the form of relational attachments to rather underspecified phrase structure rules. As Meurers points out, the serious disadvantage for processing with such a grammar is that the information in the relational attachments is not available to guide the parsing process. For efficiency reasons, proceeding in this way therefore does not seem to be an option for practical grammar implementations.

3.2.1 Flat Structures

3.2.1.1 Multiplying Out the Rules

If flat structures are used, there are two ways to account for the free constituent order in the *Mittelfeld*. Firstly, a subcat list with fixed order can be used and all possible permutations are represented in the rules. This leads to an enormous amount of phrase structure rules which are shown in (16).[9]

(16) a. H → H[SC <A>], A
  b. H → A, H[SC <A>]
  c. H → H[SC <A>], A, VC
  d. H → H[SC <A, B>], A, B, VC
  e. H → H[SC <A, B>], B, A, VC

[9] A reviewer pointed out to me that ALE provides the possibility to represent the rules in (i) by a single rule of the form in (ii).

(i) a. H → H[SC <>]
  b. H → A, H[SC <A>]
  c. H → A, B, H[SC <A,B>]
  d. H → A, B, C, H[SC <A,B,C>]
  e. …

(ii) H → X, H[SC X]

At the time such a schematic rule is applied in parsing, the number of daughters has to be known though. This essentially makes it necessary that the head is found first to determine the length of the subcat list and thereby the number of daughters involved. Therefore in ALE rules like (16a,c,d) cannot be abbreviated in this way. It is also not possible to collaps rules like (16e) and (16g) since here the arguments are explicitly refered to in order to permute them.

Note that systems which do not provide for relational constraints the number of daughters has to be fixed anyway, since information has to be percolated from the daughters to the mother (for instance nonlocal features and semantic indices).

f. H → A, B, H[SC <A, B>]
g. H → B, A, H[SC <A, B>]

(17) a. H → H[SC <A, B>], A, B
b. H → H[SC <A, B>], B, A
c. H → H[SC <A, B, C>], A, B, C, VC
d. H → H[SC <A, B, C>], A, C, B, VC
e. H → H[SC <A, B, C>], C, A, B, VC
...
f. H → A, B, C, H[SC <A, B, C>]
g. H → A, C, B, H[SC <A, B, C>]
h. H → C, A, B, H[SC <A, B, C>]
...

The first two rules are needed for intransitive verbs with the verb (H) in initial and in final position. The third rule stands for a sentence with an intransitive verb, a verbal complex and a finite verb in initial position. The sentences in (18) are examples.

(18) a. Hat er geschlafen?
has he slept
'Did he sleep?'
b. Hat er schlafen wollen?
has he sleep want
'Did he want to sleep?

The fourth and fifth rules stand for intransitive verbs in sentences where the verbs in the verbal complex add an argument.

(19) a. Hat das Kind ihn schlafen lassen?
'Did the child let him sleep?'
b. Hat ihn das Kind schlafen lassen?

The sixth and the seventh rule stand for sentences with intransitive verbs and a verbal complex in final position that adds one argument .

(20) a. daß das Kind ihn schlafen lassen hat.
'that the child let him sleep.'
b. daß ihn das Kind schlafen lassen hat.

In these rules the verbal complex functions as a normal head. The rule therefore can be used for transitive verbs in final position as well.

(21) a. daß keine Frau diesen Mann liebt.
'that no woman loves this man'
b. daß diesen Mann keine Frau liebt.

For transitive verbs one has to add rules for head initial position (17a–b), for a verbal complex that adds one argument with the head in initial position (17c–d), and rules for sentences with a verbal complex in final position (17f–h). The '...' stand for three more rules with appropriate permutations. So for verbs with two complements we had to add $2! + (2 * 3!)$ rules. For a maximum of *n* complements we get:

(22) $$2 + \sum_{i=1}^{n} i! + 2 * (i+1)!$$

The 2 in (22) stands for the rules (16b–c), which are not part of the recursion. If one assumes a maximum of 5 elements in the subcat list of lexical verbs, one gets 1901 rules.

If rules are processed from left to right, verb-last sentences like (23) introduce an enormous amount of active edges since the arity of the appropriate rule is not known until the verbal complex is found.

(23) weil Hans Cecilia John das Nilpferd füttern helfen läßt.
'because Hans lets Cecilia help John feed the hippo.'

3.2.1.2 A Lexical Rule

If one uses lexical rules that license lexical entries with all permutations of the elements in a subcat list, like the one discussed in section 3.1.1.2, the following 17 phrase structure rules are sufficient if one artificially restricts the length of the subcategorization list to five elements:

(24)
a. H → H[SC <A>], A
b. H → A, H[SC <A>]
c. H → H[SC <A>], A, VC
d. H → H[SC <A, B>], A, B, VC
e. H → A, B, H[SC <A, B>]
f. H → H[SC <A, B>], A, B
g. H → H[SC <A, B, C>], A, B, C, VC
h. H → A, B, C, H[SC <A, B, C>]
i. H → H[SC <A, B, C>], A, B, C
j. H → H[SC <A, B, C, D>], A, B, C, D, VC
k. H → A, B, C, D, H[SC <A, B, C, D>]
l. H → H[SC <A, B, C, D>], A, B, C, D
m. H → H[SC <A, B, C, D, E>], A, B, C, D, E, VC
n. H → A, B, C, D, E, H[SC <A, B, C, D, E>]
o. H → H[SC <A, B, C, D, E>], A, B, C, D, E
p. H → H[SC <A, B, C, D, E, F>], A, B, C, D, E, F, VC
q. H → A, B, C, D, E, F, H[SC <A, B, C, D, E, F>]

The rules in (24) list the daughters in the same order as they appear in the subcat lists of the respective heads. The permutations are accounted for by the lexical rule. So in contrast to (16) and (17) there is no permutation of subcat elements in the grammar rules in (24).

For a maximum of *n* complements we get.

(25) $2 + 3 * n$

If one assumes a maximum of 5 elements in the subcat list of verbs, one gets 17 rules. See Meurers (1994) for such a proposal and for other rules that are necessary to analyze constructions that have not been discussed here. In Meurers' grammar the number of complements + subject is restricted to three, i.e., he has rules that are equivalent to (25a–i).

To my knowledge, there is no HPSG grammar around that uses flat structures and that can be processed with a system without lazy evaluation techniques[10], and that accounts for adjuncts in the *Mittelfeld*.

3.2.2 Binary Branching Structures

In grammars that use binary branching structures (LS-Gram, Verb*mobil*) the description of adjuncts is less problematic, since in such a setup adjuncts can be sisters of every binary tree or terminal node. In LS-Gram the scope facts were not accounted for. If one treats the verb position in a linguistically motivated way, one gets performance problems due to the enormous search space (see section 3.1.2.2).

In both LS-Gram and Verb*mobil* a lexical rule is used that permutes the elements of the subcat list (see section 3.1.1.2). So two binary branching head-complement rules and two predicate complex rules are sufficient.

(26) a. H → X, H[SC <X>]
b. H → H[SC <X>], X

(27) a. H → VC, H[VCOMP <VC>]
b. H → H[VCOMP <VC>], VC

3.2.3 Extraposition

As far as I know, there is no system/grammar around that does not use lazy evaluation and handles extraposition in an adequate way. Despite of the known nonlocality of the phenomenon, extraposition is described by local means in both the LS-Gram and the Verb*mobil* grammar.

4 Discontinuous Constituents

Finally, I want to explain how the mentioned phenomena can be analyzed with discontinuous constituents. In the Babel system can process a grammar as the one described in the following directly.

4.1 Relative Free Constituent Order in the *Mittelfeld*

To parse the sentences in (1), a dominance structure is built that is shown in figure 5 on the next page. The elements that are circled are inserted into a list which is called word order domain (DOM). In this list the elements can appear in any order provided that no LP-constraint is violated.

This is formalized in the following way: Heads contain a description of themselves in their domain list.

[10] See also footnote 4.

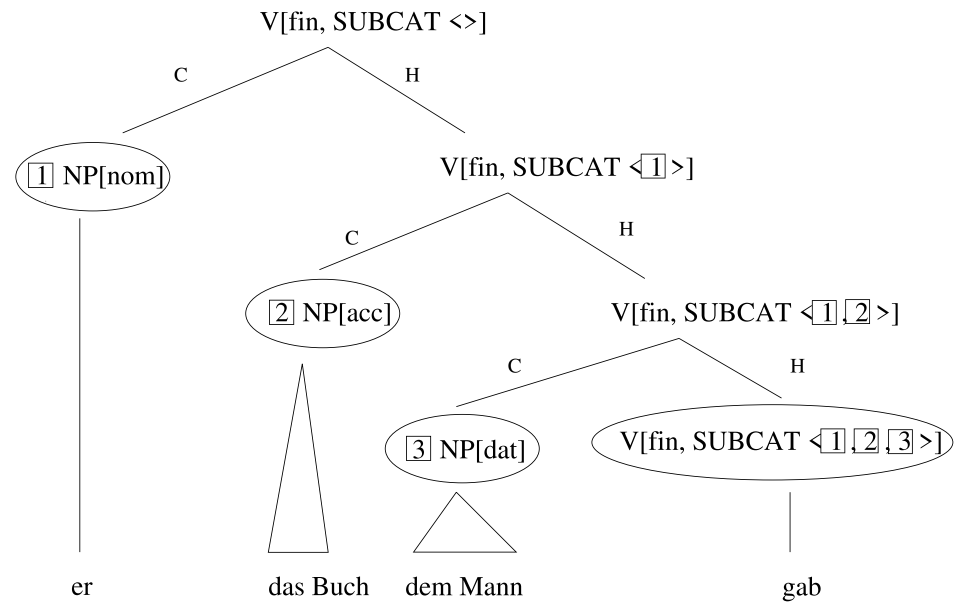

Figure 5: Dominance Structure for: *er das Buch dem Mann gab.*

(28)
$$\begin{bmatrix} \text{PHON} & \boxed{1} \\ \text{SYNSEM} & \boxed{2} \\ \text{DOM} & \left\langle \begin{bmatrix} \text{PHON} & \boxed{1} \\ \text{SYNSEM} & \boxed{2} \\ \textit{lexical-sign} & \end{bmatrix} \right\rangle \\ \textit{lexical-sign} & \end{bmatrix}$$

If a head is combined with complements or adjuncts or other dependent elements, these are inserted into the domain of the head.

(29) Construction of Domains:
$$\begin{bmatrix} \text{DTRS} & \begin{bmatrix} \textit{headed-structure} \end{bmatrix} \\ \textit{phrasal-sign} & \end{bmatrix} \Rightarrow \begin{bmatrix} \text{DTRS} & \begin{bmatrix} \text{HEAD-DTR}|\text{DOM} & \boxed{1} \\ \text{FILLER-DTRS} & \boxed{2} \\ \text{ADJ-DTRS} & \boxed{3} \\ \text{COMP-DTRS} & \boxed{4} \\ \text{CLUSTER-DTRS} & \boxed{5} \end{bmatrix} \\ \text{DOM} & \boxed{1} \bigcirc \boxed{2} \bigcirc \boxed{3} \bigcirc \boxed{4} \bigcirc \boxed{5} \end{bmatrix}$$

The $\bigcirc$ is the *shuffle* relation as used by Reape (1994). The *shuffle* relation holds between three lists A, B, and C, iff C contains all elements of A and B and the order of the elements of A and the order of elements of B is preserved in C. So if a and b are elements of A and a precedes b in A, it has to precede b in C too.

The PHON value of a phrasal sign is the concatenation of the PHON values of its domain elements.

$$(30)\quad \begin{bmatrix} \textit{phrasal-sign} \end{bmatrix} \Rightarrow \begin{bmatrix} \text{PHON} & \boxed{1} \oplus \ldots \oplus \boxed{n} \\ \text{DOM} & \left\langle \begin{bmatrix} \text{PHON} & \boxed{1} \\ \textit{sign} & \end{bmatrix}, \ldots, \begin{bmatrix} \text{PHON} & \boxed{n} \\ \textit{sign} & \end{bmatrix} \right\rangle \end{bmatrix}$$

In (30), $\oplus$ corresponds to the *append* relation.

If a phrase like (1a) is analyzed, it gets exactly the same dominance structure as (1f). The only difference is that the constituents in (1f) are continuous, whereas for (1a) we get the discontinuous constituents *gab der Frau* and *gab der Frau das Buch*. Adverbs are inserted into the domain of their head as well. Their free appearance in the *Mittelfeld* is therefore explained.

4.2 Verb Position

In a domain based approach the verb is in the same linearization domain as its adjuncts and complements. It can be serialized either to their left in verb initial position or to their right in verb final position. The dominance structure is identical in both cases. An example that involves a discontinuous projection is shown in figure 6. The scope facts are explained in a similar way: Only the order of the

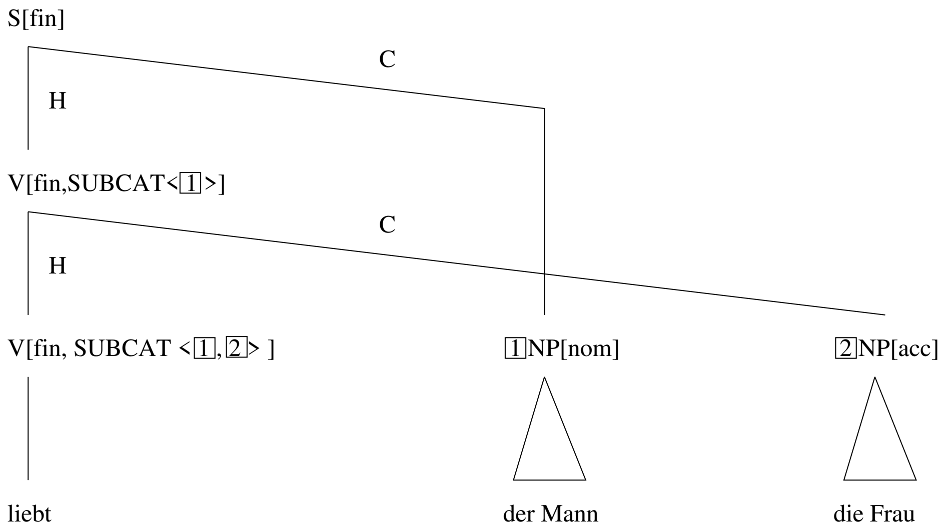

Figure 6: Verb First with Discontinuous Constituents

combinations of adjuncts and heads is important. The order is the same regardless whether the verb is in final or in inital position. Linearization rules ensure that an adjunct always has scope over all adjuncts to its right, i.e., the last adjunct in a linearization domain is combined with the verb it modifies first.

4.3 Extraposition

With discontinuous constituents, extraposition can be accounted for in a What You See Is What You Get manner (WYSIWYG). Such an approach was suggested by Kathol and Pollard (1995), who combine an extraposed element and its antecedent despite their discontinuity. In the process of word order domain formation, the extraposed constituent is inserted in the higher domain separately. No introduction of hypotheses is necessary.

5 Comparison

In the following the Verb*mobil* grammar will be compared with a grammar that uses discontinuous constituents: the grammar of the Babel system, which was developed earlier and which was extended to cover the Verb*mobil* data. Both grammars provide an analysis for approximately 80 % of the grammatical input in the Verb*mobil* domain (appointment scheduling and trip planning). The grammar that uses discontinuous constituents accounts for adjunct extraction, extraposition as a nonlocal dependency, *it* extraposition, complex fronting, free relative clauses, and depictive secondary predicates. All these phenomena are very costly for grammars with continuous constituents.[11]

Figure 7 shows the amount of passive edges that are created during a full parse relative to the utterance length. Both systems parsed the utterances of the CDs 1, 15, and 20 of the Verb*mobil* project. These CDs contain 10.340 utterances.

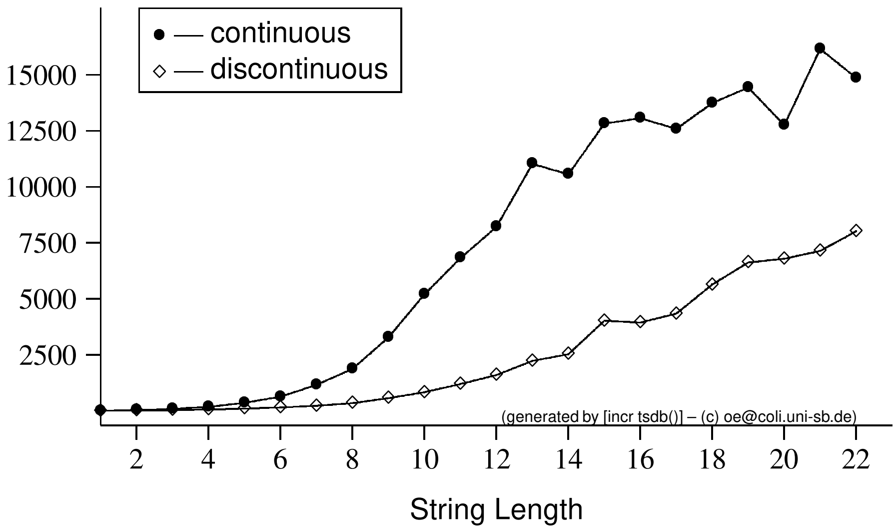

Figure 7: Passive edges generated by the grammars during a full parse

The figure shows that the number of passive edges grows more quickly in a system with continuous constituents. For the curve of the grammar with continuous constituents it is important to know that an edge limit influenced the curve in the region over twelve words. The edge limit was necessary since process sizes of higher then 2 GB were not allowed. For unlimited parses I expect the curve to grow in the same way as it grows up to twelve words. For a twelve-word utterance the number of passive edges in a system with continuous constituents is higher than the number of passive edges for a twenty-two-word sentence in a system with discontinuous constituents.

Comparing these results, one thing has to be mentioned: The size of the feature structures in a system that uses continuous constituents is smaller than the size for structures in a system with discontinuous constituents. The reason for this difference is that in a system with continuous constituents the daughters of a feature structure need not be stored during parsing. In a system with discontinuous

[11] See the electronic version of this document for a detailed discussion of the phenomena including examples from the Verb*mobil* domain and an explanation why these phenomena are expansive. (http://www.dfki.de/~stefan/Pub/e_discont.html).

constituents however, the domain elements have to be stored in each newly constructed edge. Linear precedence constraints can instantiate features of domain elements that are co-indexed elsewhere in the structure. So at least the immediate daughters have to be used during parsing.

6 Summary and Outlook

It has been shown that the domain-based grammar is better suited to parse sentences with a sentence length of up to 22 words. The search space for such sentences is considerably smaller. The Verb*mobil* corpus contains sentences with a length of up to 60 words, but the number of sentences with more than 22 words is not significant (2.3 % of 23,547 utterances). It remains to be seen how different grammars behave when it comes to sentences longer than the ones one finds in spoken language.

An interesting direction for further research is to implement a grammar that can be processed with one system in two different modes: with and without discontinuous constituents. Since the part of the Verb*mobil* grammar that accounts for constituent order does not interfere with other parts of the grammar it can be separated easily and the other linearization module can be plugged in. A system that can process both kinds of grammars is currently under development in Tübingen (Fouvry and Meurers, To Appear). With such a system at hand it becomes possible to compare runtimes which is not possible now since there are to many varying parameters as for instance system specific memory requirements.

Bibliography

Abney, Steven P. and Hinrichs, Erhard W. (eds.). 1995. *Proceedings of the Seventh Conference of the European Chapter of the Association for Computational Linguistics*, Dublin.

Bech, Gunnar. 1955. *Studien über das deutsche Verbum infinitum*. Linguistische Arbeiten, No. 139, Tübingen: Max Niemeyer Verlag, 2nd unchanged edition 1983.

Behagel, Otto. 1930. Von deutscher Wortstellung. *Zeitschrift für Deutschkunde* pages 81–89.

Bierwisch, Manfred. 1963. *Grammatik des deutschen Verbs*. studia grammatica II, Berlin: Akademie Verlag.

Bouma, Gosse. 1996. Extraposition as a Nonlocal Dependency. In Geert-Jan Kruijff, Glynn Morrill and Dick Oehrle (eds.), *Proceedings of Formal Grammar 96*, pages 1–14, Prag, http://www.let.rug.nl/~gosse/papers/. 07.24.97.

Burgard, Rose (ed.). To Appear. *Das Verbmobil-Buch*. Symbolic Computation, Berlin Heidelberg New York: Springer-Verlag.

Carpenter, Bob and Penn, Gerald. 1996. Efficient Parsing of Compiled Typed Attribute Value Logic Grammars. In Harry Bunt and Masaru Tomita (eds.), *Recent Advances in Parsing Technology*, Text, Speech and Language Technology, No. 1, Dordrecht/Boston/London: Kluwer Academic Publishers.

Carroll, John. 1994. Relating Complexity to Practical Performance in Parsing with Wide-coverage Unification Grammars. In James Pustejovsky (ed.), *Proceedings of the Thirty-Second Annual Meeting of the ACL*, pages 287–294, Association for Computational Linguistics, Morristown, NJ, e-Print-Archive: http://xxx.lanl.gov/abs/cmp-lg/9405033. 04.07.99.

Copestake, Ann. 1999. The (New) LKB System, ftp://www-csli.stanford.edu/~aac/newdoc.pdf. 06.24.99.

Donohue, Cathryn and Sag, Ivan A. 1999. Domains in Warlpiri. In *Sixth International Conference on HPSG–Abstracts. 04–06 August 1999*, pages 101–106, Edinburgh.

Drach, Erich. 1937. *Grundgedanken der deutschen Satzlehre*. Darmstadt: Wissenschaftliche Buchgesellschaft.

Engelkamp, Judith, Erbach, Gregor and Uszkoreit, Hans. 1992. Handling Linear Precedence Constraints by Unification. In *Proceedings of the Thirtieth Annual Meeting of the ACL*, pages 201–208, ACL, Newark, DE, erschien auch als CLAUS-Report, Nr. 19, Universität des Saarlandes.

Erbach, Gregor. 1995. ProFIT: Prolog with Features, Inheritance and Templates. In Abney and Hinrichs (1995).

Erbach, Gregor. 1998. *Bottom-Up Earley Deduction for Preference-Driven Natural Language Processing*. Saarbrücken Dissertations in Computational Linguistics and Language Technology, No. 4, Saarbrücken: Deutsches Forschungszentrum für Künstliche Intelligenz und Universität des Saarlandes.

Fouvry, Frederik and Meurers, Detmar. To Appear. Towards a Platform for Linearization Grammars. In Erhard Hinrichs, Detmar Meurers and Shuly Wintner (eds.), *ESSLLI'2000 Workshop on Linguistic Theory and Grammar Implementation*.

Fries, Norbert. 1988. Über das Null-Topik im Deutschen. Forschungsprogramm Sprache und Pragmatik 3, Universität Lund-German. Inst., Lund.

Görz, Günther (ed.). 1992. *Konvens 92. 1. Konferenz „Verarbeitung natürlicher Sprache“. Nürnberg 7.–9. Oktober 1992*, Informatik aktuell, Berlin, Heidelberg, New York, Springer-Verlag.

Gunji, Takao. 1986. Subcategorization and Word Order. In William J. Poser (ed.), *Papers from the Second International Workshop on Japanese Syntax*, pages 1–21, Stanford: Center for the Study of Language and Information.

Haider, Hubert. 1991. Fakultativ kohärente Infinitivkonstruktionen im Deutschen. Arbeitspapiere des Sonderforschungsbereiches 340 No. 17, IBM Deutschland GmbH, Heidelberg.

Heinz, Wolfgang and Matiasek, Johannes. 1994. Argument Structure and Case Assignment in German. In Nerbonne et. al. (1994), pages 199–236.

Hinrichs, Erhard W. and Nakazawa, Tsuneko. 1989a. Flipped out: AUX in German. In *Aspects of German VP Structure*, SfS-Report-01-93, Eberhard-Karls-Universität Tübingen.

Hinrichs, Erhard W. and Nakazawa, Tsuneko. 1989b. Subcategorization and VP Structure in German. In *Aspects of German VP Structure*, SfS-Report-01-93, Eberhard-Karls-Universität Tübingen.

Hinrichs, Erhard W. and Nakazawa, Tsuneko. 1994. Partial-VP and Split-NP Topicalization in German—An HPSG Analysis. In Erhard W. Hinrichs, Walt Detmar Meurers and Tsuneko Nakazawa (eds.), *Partial-VP and Split-NP Topicalization in German—An HPSG Analysis and its Implementation*, Arbeitspapiere des Sonderforschungsbereiches 340, No. 58, Universität Tübingen.

Hinrichs, Erhard W. and Nakazawa, Tsuneko. 1998. Third Construction and VP Extraposition in German: An HPSG Analysis. In Erhard W. Hinrichs, Andreas Kathol and Tsuneko Nakazawa (eds.), *Complex Predicates in Nonderivational Syntax*, volume 30 of *Syntax and Semantics*, pages 115–157, San Diego: Academic Press.

Hoberg, Ursula. 1981. *Die Wortstellung in der geschriebenen deutschen Gegenwartssprache*, volume 10 of *Heutiges Deutsch. Linguistische Grundlagen. Forschungen des Instituts für deutsche Sprache*. München: Max Huber Verlag.

Höhle, Tilman N. 1982. Explikation für „normale Betonung“ und „normale Wortstellung“. In Werner Abraham (ed.), *Satzglieder im Deutschen – Vorschläge zur syntaktischen, semantischen und pragmatischen Fundierung*, Studien zur deutschen Grammatik, No. 15, pages 75–153, Tübingen: Gunter Narr Verlag.

Jacobs, Joachim. 1991. Bewegung als Valenztransfer. SFB 282: Theorie des Lexikons 1, Heinrich Heine Uni/BUGH, Düsseldorf/Wuppertal.

Kathol, Andreas. 1995. *Linearization-Based German Syntax*. Ph. D.thesis, Ohio State University.

Kathol, Andreas and Pollard, Carl J. 1995. Extraposition via Complex Domain Formation. In *Proceed-*

ings of the Thirty-Third Annual Meeting of the ACL, Association for Computational Linguistics, Boston, http://linguistics.berkeley.edu/~kathol/Papers/ACL95.ps.gz. 06.29.99.

Keller, Frank. 1994. Extraposition in HPSG. Verbmobil-Report 30, IBM Deutschland Informationssysteme GmbH, Institut für Logik und Linguistik, Heidelberg.

Keller, Frank. 1995. Towards an Account of Extraposition in HPSG. In *Proceedings of the Seventh Conference of the EACL*, Dublin.

Kiss, Tibor. 1995. *Infinite Komplementation. Neue Studien zum deutschen Verbum infinitum*. Linguistische Arbeiten, No. 333, Tübingen: Max Niemeyer Verlag.

Kiss, Tibor and Wesche, Birgit. 1991. Verb Order and Head Movement. In Otthein Herzog and Claus-Rainer Rollinger (eds.), *Text Understanding in LILOG*, Lecture Notes in Artificial Intelligence, No. 546, pages 216–242, Berlin Heidelberg New York: Springer-Verlag.

Kuno, Susumu. 1980. A Further Note on Tonoike's Intra-Subjectivization Hypothesis. In Yukio Otsu and Ann Farmer (eds.), *Theoretical Issues in Japanese Linguistics*, volume 2 of *MIT Working Papers in Linguistics*, pages 171–184, Cambridge: MIT Working Papers in Linguistics.

Kunze, Jürgen. 1991. *Kasusrelationen und semantische Emphase*. studia grammatica XXXII, Berlin: Akademie Verlag.

Meurers, Walt Detmar. 1994. On Implementing an HPSG theory. In Erhard W. Hinrichs, Walt Detmar Meurers and Tsuneko Nakazawa (eds.), *Partial-VP and Split-NP Topicalization in German – An HPSG Analysis and its Implementation*, Arbeitspapiere des SFB 340, No. 58, Universität Tübingen.

Müller, Stefan. 1995. Scrambling in German – Extraction into the *Mittelfeld*. In Benjamin K. T'sou and Tom B. Y. Lai (eds.), *Proceedings of the tenth Pacific Asia Conference on Language, Information and Computation*, pages 79–83, City University of Hong Kong, http://www.dfki.de/~stefan/Pub/e_scrambling.html. 06.29.2000.

Müller, Stefan. 1997. Yet another Paper about Partial Verb Phrase Fronting in German. Research Report RR-97-07, Deutsches Forschungszentrum für Künstliche Intelligenz, Saarbrücken, a shorter version appeared in *Proceedings of COLING 96*, pages 800–805. http://www.dfki.de/~stefan/Pub/e_pvp.html. 06.29.2000.

Müller, Stefan. 1999a. *Deutsche Syntax deklarativ. Head-Driven Phrase Structure Grammar für das Deutsche*. Linguistische Arbeiten, No. 394, Tübingen: Max Niemeyer Verlag, http://www.dfki.de/~stefan/Pub/e_hpsg.html. 06.29.2000.

Müller, Stefan. 1999b. An HPSG-Analysis for Free Relative Clauses in German. *Grammars* 2(1), 53–105, http://www.dfki.de/~stefan/Pub/e_freeRel.html. 06.29.2000.

Müller, Stefan. In Preparation. Yet Another Theory of Complex Predicates. Ms. Saarbrükken: Deutsches Forschungszentrum für Künstliche Intelligenz. http://www.dfki.de/~stefan/Pub/e_complex.html. 06.29.2000.

Müller, Stefan. To appear. Case in German – Towards an HPSG Analysis. In Tibor Kiss and Detmar Meurers (eds.), *Constraint-Based Approaches to Germanic Syntax*, Studies in Constraint-Based Lexicalism, Stanford: Center for the Study of Language and Information, http://www.dfki.de/~stefan/Pub/e_case.html. 06.29.2000.

Müller, Stefan and Kasper, Walter. To Appear. HPSG analysis of German. In Burgard (To Appear).

Nerbonne, John, Netter, Klaus and Pollard, Carl J. (eds.). 1994. *German in Head-Driven Phrase Structure Grammar*. CSLI Lecture Notes, No. 46, Stanford: Center for the Study of Language and Information.

Netter, Klaus. 1992. On Non-Head Non-Movement. An HPSG Treatment of Finite Verb Position in German. In Görz (1992), pages 218–227.

Paritong, Maike. 1992. Constituent Coordination in HPSG. In Görz (1992), pages 228–237.

Penn, Gerald. 1999. Linearization and *WH*-Extraction: Evidence from Serbo-Croatian. In Robert D. Borsley and Adam Przepiórkowski (eds.), *Slavic in Head-Driven Phrase Structure Grammar*, Studies in Constraint-Based Lexicalism, pages 149–182, Stanford: Center for the Study of Language and Information.

Penn, Gerald and Carpenter, Bob. 1999. ALE for Speech: a Translation Prototype. In *Proceedings of the 6th Conference on Speech Communication and Technology (EUROSPEECH)*, Budapest, Hungary.

Pollard, Carl J. 1990. On Head Non-Movement. In Harry Bunt and Arthur van Horck (eds.), *Discontinuous Constituency*, Natural Language Processing, No. 6, pages 279–305, Berlin, New York: Mouton de Gruyter.

Pollard, Carl J., Kasper, Robert T. and Levine, Robert D. 1992. Linearization Grammar. Research Proposal to the National Science Foundation.

Pollard, Carl J., Kasper, Robert T. and Levine, Robert D. 1994. Studies in Constituent Ordering: Toward a Theory of Linearization in Head-Driven Phrase Structure Grammar, research Proposal to the National Science Foundation. ftp://csli-ftp.stanford.edu/linguistics/sag/linearization-prop.ps.gz. 06.29.99.

Pollard, Carl J. and Sag, Ivan A. 1994. *Head-Driven Phrase Structure Grammar*. Studies in Contemporary Linguistics, Chicago, London: University of Chicago Press.

Reape, Mike. 1991. Word Order Variation in Germanic and Parsing. DYANA Report Deliverable R1.1.C, University of Edinburgh.

Reape, Mike. 1992. *A Formal Theory of Word Order: A Case Study in West Germanic*. Ph. D.thesis, University of Edinburgh.

Reape, Mike. 1994. Domain Union and Word Order Variation in German. In Nerbonne et. al. (1994), pages 151–198.

Richter, Frank and Sailer, Manfred. 1999. Lexicalizing the left periphery of German finite sentences. In Valia Kordoni (ed.), *Tübingen Studies in Head-Driven Phrase Structure Grammar*, Arbeitspapiere des SFB 340, Nr. 132, Volume 1, pages 116–154, Universität Tübingen.

Schmidt, Paul, Rieder, S. and Theofilidis, Axel. 1996. Final Documentation of the German LS-GRAM Lingware. Deliverable DC-WP6e (German), IAI, Saarbrücken.

Schütz, Jörg. 1996. The ALEP Formalism in a Nutshell, http://www.iai.uni-sb.de/alep/alep-docs.html. 06.09.99.

Uszkoreit, Hans. 1986. Linear Precedence in Discontinuous Constituents: Complex Fronting in German. Report No. CSLI-86-47, Center for the Study of Language and Information, Stanford.

Uszkoreit, Hans. 1987. *Word Order and Constituent Structure in German*. CSLI Lecture Notes, No. 8, Stanford: Center for the Study of Language and Information.

Uszkoreit, Hans, Backofen, Rolf, Busemann, Stephan, Diagne, Abdel Kader, Hinkelman, Elizabeth A., Kasper, Walter, Kiefer, Bernd, Krieger, Hans-Ulrich, Netter, Klaus, Neumann, Günter, Oepen, Stephan and Spackman, Stephen P. 1994. DISCO—An HPSG-based NLP System and its Application for Appointment Scheduling. In *Proceedings of COLING 94*, pages 436–440, Kyoto, Japan.

van Noord, Gertjan and Bouma, Gosse. 1994. The Scope of Adjuncts and the Processing of Lexical Rules. In *Proceedings of COLING 94*, pages 250–256, Kyoto Japan.

Wegener, Heide. 1985. *Der Dativ im heutigen Deutsch*. Studien zur deutschen Grammatik, No. 28, Tübingen: Gunter Narr Verlag.