Arguments desperately seeking Interpretation: Parsing German Infinitives

Christopher Laenzlinger
Martin S. Ulmann
Eric Wehrli*
LATL - Department of Linguistics
University of Geneva CH-1206 Geneva
laenzlinger@latl.unige.ch, ulmann@latl.unige.ch, wehrli@latl.unige.ch

Abstract

In this paper we present a GB-parsing system for German and in particular the system's strategy for argument interpretation, which copes with the difficulty that word order is relatively free in German and also that arguments can precede their predicate. In this latter case, the parser makes a provisional interpretation, which is checked when the argument structure of the predicate is available. Moreover, a strategy of argument transfer is used in cases of long-distance scrambling, according to which arguments and adjuncts are attached to the domain of the coherent verb, ECM verb, or raising verb, and transferred to the infinitival complement for interpretation.

1 Introduction

Free word order languages raise difficulties for parsing systems based on phrase-structure rule grammars, where the constituents are ordered. Indeed, to list all the possible orders leads to an increase in the grammar size and a corresponding decrease in performance. There have been several approaches to this problem, notably those based on the ID/LP (immediate dominance/linear precedence) grammars (*cf.* Gazdar *et al.* 1985) or functional unification grammars (*cf.* Karttunnen & Kay 1985). Within the Government and Binding framework, Kashket (1991) presents a parser for Warlpiri, a non-configurational language, where word order and its variation depends mainly on case marking.

Although German is a partially free word order language, we will assume that it has a fixed base word order, which is modified by a set of movement transformations. In this paper, we will present the argument interpretation strategy of our parser for German, which is able to handle the difficulties arising from word order variations, focusing on the treatment of infinitival constructions.[1]

2 The DIPS Parser

2.1 General Properties of DIPS

DIPS (*Deutsches Interaktives Parsing System* 'German Interactive Parsing System') is a large-scale interactive GB-based[2] parsing system. Its architecture is basically similar to that of IPS (Wehrli 1992) and FIPS (Laenzlinger & Wehrli 1991). The parser produces a set of GB S-structures (trees) from an input sentence. These structures are associated with information concerning traces, argument structure, and case features.

The syntactic structure of constituents corresponds to the GB $\overline{X}$-schema. We consider German to be an SOV-language (i.e. objects precede their predicates in their base position). Thus, the $\overline{X}$-schema is parameterized in German as follows: The complement (Compl) precedes the head X^0 for the categories V, A, I, whereas it follows the head for the categories C, D, P, N, Adv.[3] As the specifier (Spec) is always on the left, the $\overline{X}$-schema has the structure given in (1).

(1) $XP \rightarrow$ Spec $\overline{X}$
$\overline{X} \rightarrow X^0$ Compl, if X^0={C^0, D^0, P^0, N^0, Adv^0 }
$\overline{X} \rightarrow$ Compl X^0, if X^0={V^0, A^0, I^0}

On the basis of this schema, the clause structure in German has the general representation given in

*Thanks to Scott Fergusson for comments. This work has been supported in part by a grant from the FNRS, grant no 11-33731.92

[1] There are other proposals to deal with infinitival constructions in German: Netter 1986 discusses an LFG approach, and Rambow 1994 uses a variation of TAG.

[2] *Cf.* Chomsky & Lasnik 1992, Haegeman 1994 for a presentation of Government and Binding Theory (GB), and Berwick *et al.* 1991, Wehrli 1988 for a possible implementation of the theory.

[3] We assume Abney's 1987 DP-hypothesis, according to which the head of a noun phrase is the determiner (D^0).

Figure 1.

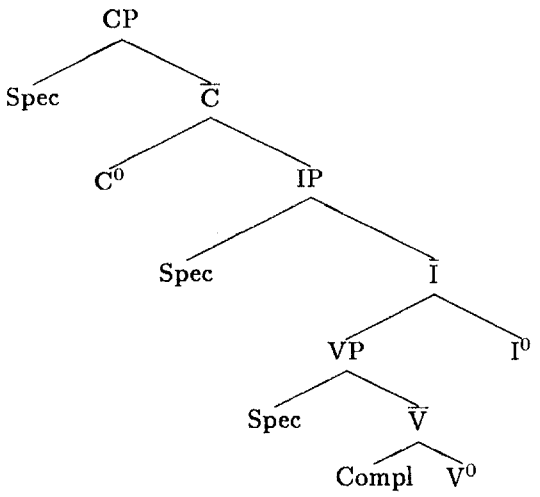

Figure 1: Structure of a German clause

2.2 General Parsing Strategy

The parsing strategy is data-triggered (mainly bottom-up), proceeds from left to right, and treats alternatives in parallel by using a chart (*cf.* Kay 1980/1986 and Kaplan 1973). The analysis of a sentence proceeds at two levels: the lexical level and the syntactic level. The lexical analysis looks up the words in the lexicon; each lexical item (word) projects the node corresponding to its category; thus, the lexical features are transferred to the syntactic level in accordance with the Projection Principle (Chomsky 1981). The projected node is then inserted into the chart as an edge. The syntactic analysis builds all possible structures by making use of cross-category projections (similar to Grimshaw's 1991 "extented projections") and attachments, which are further filtered by grammatical constraints; structure building is incremental, as the current constituent is immediately integrated into the existing hypotheses.

A cross-category projection creates a new constituent with the same start and end vertex in the chart as the subconstituent from which it is projected. This kind of projection is limited to some categories and triggered by intrinsic features. For instance, an infinitival verb projects the structure in Figure 1 from VP to CP.

Attachment combines the current constituent with the constituents which immediately precede this current constituent in the chart. Attachments can be divided into two different types of combination:

1. A constituent of the left context is attached to the current constituent (left attachment).
2. The current constituent is attached to a constituent of the left context (right attachment).

In order to keep track of where a constituent can be attached in the structure, a list of active nodes specifies the potential attachment sites; this list is systematically updated. Attachments are further constrained as follows:

- Formal attachments are restricted to adjacent constituents and are licensed by lexical properties such as selection or agreement (*e.g.* auxiliary-verb selection, determiner-noun agreement).
- Non-formal attachments concern thematic complements and are licensed by subcategorization and theta properties.

The second type of attachment requires a specific argument interpretation strategy (AIS) to establish the link between the argument and the predicate which subcategorizes it.

2.3 The Argument Interpretation Strategy

The aim of the AIS is to match the arguments with the subcategorization properties (argument structure) of the predicate, and thus to establish an interpretation, which corresponds to the assignment of the thematic roles. The argument structure of a verb (predicate) is provided by the lexicon and specifies the number and type of arguments that the predicate can take; while there can be more than one argument structure for a verb at the lexical level, there is only one argument table for a (verb) node at the syntactic level, which contains the arguments of the clause. This argument table is matched with the corresponding argument structures, which has the effect of filtering the inappropriate argument structures.

The AIS has to deal with two types of difficulties: first, the predicate (with its argument structure) is not always available at the time the argument is attached; second, the large number of possible word orders in German makes the argument's grammatical function difficult to determine.

The argument structure is only available if the main verb (predicate) occurs in C^0, that is the second position in the clause (verb second with the main verb), and thus at most one argument precedes the verb. In this case, a final interpretation of the arguments is established immediately (at the moment of attachment); the arguments are inserted into the definitive argument table of the clause and interpreted by being matched with the argument structure of the verb (theta assignment); if more than one interpretation is possible, different hypotheses are considered in parallel. If the verb follows the arguments, they are also inserted into the argument table, with a provisional interpretation.[4] The matching between the argu-

[4] This strategy seems to have psycholinguistic sup-

ment table and the argument structure eventually takes place at the time the main verb is attached.

The task of identifying the grammatical function of an argument is complicated by the large number of possible word orders, which results from the interaction of three syntactic processes: verb second, scrambling, and extraposition. The verb second constraint requires that the tensed verb occupies the second position of the main clause; for the first position, however, a large number of constituents (XP) is possible, such as the subject, an object, an adjunct, an empty operator. Scrambling is a process that modifies the order of clause-internal arguments and adjuncts under some constraints (*cf.* for instance, Uszkoreit 1987). Extraposition is the occurrence of prepositional or sentential complements or adjuncts after the verb in its base position V^0. Thus, the grammatical function of an argument depends not only on its position, but also on case and agreement information and (scrambling) ordering constraints.

The interpretation module works as follows: The first step is to check whether there are arguments to be interpreted. If so, it is further checked whether the main verb is available, with the argument structures. In case it is not available, the new argument is inserted into the provisional argument table (and its interpretation can be checked only later, when the argument structure is available). If it is available, the new argument is matched with the argument structures; if there is a provisional argument table instead of one argument, the matching is effected for each argument in turn. Thus, the list of argument structures is filtered and a list of new argument tables is returned. For each of these argument tables, it is checked whether its arguments obey the ordering constraints. If so, the new structure is completed and for each argument that is not in its base position, a chain is created to link the argument with that position, in which a trace is inserted.

Let us illustrate how the analysis proceeds on the basis of the sentence in (2).

(2) Die Kinder haben diesen Bericht gelesen.
'the children have this report read'
the children have read this report.

When the parser reads the verb *haben*, the general clause structure (*cf.* Figure 1) is projected from VP to CP, triggered by the tensed verb, which is placed in C^0 leaving a head trace in V^0 and in I^0. Then, the first constituent *die Kinder* is attached as the specifier of the CP.

As this first constituent is morphologically ambiguous between nominative and accusative, it can be interpreted a priori as a subject or as a direct object. For the hypothesis of *haben* as a main verb or of a verb with particle, the argument structure is available: *diesen Bericht* is the direct object and *die Kinder* the subject. This hypothesis, however, fails when the parser arrives at the participle *gelesen*. For the hypothesis of *haben* as an auxiliary, *die Kinder* is inserted into the provisional argument table as the subject or the direct object of a forthcoming verb and *diesen Bericht* is inserted as direct object. When the past participle is read, the arguments are matched with the argument structure of *gelesen*: *die Kinder* as subject and *diesen Bericht* as direct object. A trace is inserted into the specifier of IP for the subject, and another trace into the complement of VP for the direct object, as illustrated in (3).

(3) $[_{CP}\ [_{DP}$ die Kinder $]_j [_{\bar{C}}$ haben$_i$ $[_{IP}\ [_{DP}$ $t]_j$ $[_{DP}$ diesen Bericht $]_k$ $[_{\bar{I}}\ [_{VP}\ [_{VP}\ [_{DP}$ $t]_k$ gelesen $]\ t_i\ ]]\ t'_i\ ]]]$

In the following section, we will show how the AIS works in the case of infinitival constructions.

3 The Treatment of Infinitives

3.1 Different Infinitival Structures

German displays two types of infinitives: infinitives introduced by the conjunction *zu* and infinitives without *zu*.

3.1.1 Infinitives without *zu*

Infinitives without *zu* occur as the complement of modal verbs (*e.g.* *müssen* 'must') and exceptional case marking (ECM) verbs (*e.g.* *sehen* 'see', *lassen* 'let/make'). Modals are treated on a par with auxiliaries, *i.e.* they are taken to select an infinitival VP as complement and are not associated with an argument table. In compound tenses, the infinitival form of the modal is usually used instead of its past participle form; in example (4a), the infinitive *wollen* substitutes for the participle *gewollt*. This phenomenon is called *infinitivus pro participio* (IPP) or *Ersatzinfinitiv*. If the verb selecting the IPP is in its base position, the order of the verbs differs from the usual one: auxiliaries that would be at the right of the IPP immediately precede the final predicates, as illustrated in example (4b), where *hätte* precedes *besuchen wollen*.

(4)a. Das Kind hat die alte Frau besuchen wollen.
'the child had the old woman visit want'
The child wanted to visit the old woman.

b. Wenn das Kind die alte Frau hätte besuchen wollen...
'if the child the old woman would-have visit want'
If the child had wanted to visit the old woman...

From a structural point of view, this reordering can be analyzed as verb raising (VR): the verbs

port: German speakers assign an interpretation to arguments even before the predicate is available (*cf.* Bader & Lasser 1993).

which whould precede the uppermost final auxiliary (without VR) are attached to the right of the auxiliary head (right-adjoined position), forming head chains with their base positions, as represented in (5).

(5) [$_{VP}$ [$_{VP}$ [$_{VP}$ t_i] t_j] hätte besuchen$_i$ wollen$_j$]

The phenomena of IPP and verb raising also occur with ECM verbs, as example (6) shows. Unlike modals, ECM verbs are analyzed as taking an infinitival CP as complement and assign (accusative or dative) case to the subject of the infinitival clause, e.g. the accusative case to *ihn* in (6).

(6) Nachdem ihn die Polizei hatte fliehen sehen...
'after him the police had escape see'
After the police had seen him escape...

Furthermore, the subject of the infinitival clause (*ihn*) can be attached to a position higher than the subject of the main clause as a result of scrambling.

3.1.2 Infinitives with *zu*

The subject of infinitival clauses with *zu* is an empty constituent. In control constructions, the subject is a null pronoun PRO, which can be coreferential with (controlled by) the subject (example (7a)) or the object (example (7b)) of the upper clause according to the lexical property of the main verb. In raising constructions, the subject of the infinitive is a trace coindexed with the subject of the higher clause (example (7c)).

(7)a. Er$_i$ behauptete, [$_{CP}$ PRO$_i$ sie gesehen zu haben].
'he claimed PRO her seen to have'
he claimed to have seen her.

b. Sie hat ihm$_i$ t_j erlaubt, [$_{CP}$ PRO$_i$ das Buch anzusehen]$_j$.
'she has allowed him, PRO the book to-look-at'
she allowed him to look at the book.

c. Sie$_i$ schien [$_{CP}$ t_i ihn gesehen zu haben].
'she seemed him seen to have'
She seemed to have seen him.

The infinitival clause can be extraposed in control constructions (7b), but not in raising construction.

Among subject-control verbs, there is a class of verbs, called 'coherent verbs', which form a clause union with their infinitival complement (by restructuring). As a consequence, arguments and adjuncts attached to the upper clause can be interpreted with respect to the infinitival clause.

(8) Gestern hat sie$_i$ der Professor versucht [$_{CP}$ t_i zu küssen].
'yesterday has her the professor tried to kiss'
Yesterday the professor tried to kiss her.

In the example (8), the pronoun *sie* is the direct object of the infinitival clause, although it is attached to the main clause.

3.2 Treatment of Infinitival Particularities

3.2.1 Verb Raising

The main problem with VR is that the verbs occur on the right of the uppermost final auxiliary, while their maximal VP constituents remain on the left and contain a head trace. As a solution to this problem, we propose attaching the structure that contains the verb to the left and extracting all of the heads, which are adjoined to the right of the upper verb. Take for instance the VP in (9a) and the complex VP in (9b); the latter is attached as the complement of the former, i.e. to its left. However, to account for the surface word order, the heads *besuchen* and *wollen* must be extracted and attached to the right of *hätte*, as shown in (9c).

(9)a. [$_{VP}$ hätte]

b. [$_{VP}$ [$_{VP}$ besuchen] wollen]

c. [$_{VP}$ [$_{VP}$ [$_{VP}$ t_i] t_j] hätte besuchen$_i$ wollen$_j$]

This solution also works for verb raising in ECM constructions, although the verbal head of the infinitival clause is deeper in the structure.

3.2.2 Control

An infinitive with *zu* projects a non-finite clause (CP) to which an empty subject is added (Spec IP). If the infinitival clause is a complement of a control verb, the empty subject must be coindexed with the controlling argument (lexically specified). As illustrated in (10), an infinitival clause can precede or follow its controller. Therefore, the coindexation only applies when both arguments (controller and controllee) are available (the infinitival CP and the indirect object in (10)).

(10)a. Die Mutter erlaubte ihrer Tochter$_i$ nicht, [$_{CP}$ PRO$_i$ ins Theater zu gehen].
'The mother allowed her daughter not to the theatre to go'
The mother did not allow her daughter to go to the theatre.

b. [$_{CP}$ PRO$_i$ Ins Theater zu gehen] erlaubte die Mutter ihrer Tochter$_i$ nicht.
'to the theatre to go allowed the mother her daughter not'
The mother did not allow her daughter to go to the theatre.

3.2.3 Argument restructuring

Since restructuring allows arguments and adjuncts to be attached to the clause containing a

coherent verb, while being interpreted with respect to the infinitival clause, the AIS needs to be extended. The first modification concerns the matching procedure: An argument that may be interpreted with respect to the infinitival complement is left in the argument table of the coherent verb — for this reading, no matching takes place and this argument is marked as 'uninterpreted'. The second modification occurs after argument trace insertion. At this stage, it is checked whether there are arguments marked as 'uninterpreted' and whether the infinitival complement is available. If both of these conditions are fulfilled, the uninterpreted arguments are transferred from the argument table of the main verb to a provisional argument table, which is matched with the predicate of the infinitival complement.

Consider example (8): the pronoun *sie* and the nominative DP *der Professor* are attached to the main clause. When the parser reads *versucht*, it interprets the DP unambiguously as the subject. For the pronoun, two analyses are taken into account (in parallel). On the one hand, it can be the direct object of the main verb (*Gestern hat sie der Professor versucht* 'Yesterday, the professor has tried them'); this analysis fails when the infinitival complement cannot be attached.

On the other hand, the pronoun *sie* is regarded as the argument of a following infinitival clause, *i.e.* marked as 'uninterpreted'. When the infinitival complement *zu küssen* is attached, this uninterpreted argument is treated as a new argument of the infinitival verb and interpreted as its direct object, resulting in the structure (11).

(11) [$_{CP}$ [$_{AdvP}$ Gestern] [$_{\bar{C}}$ hat$_i$ [$_{IP}$ [$_{DP}$ sie]$_j$ [$_{DP}$ der Professor]$_k$ [$_{\bar{I}}$ [$_{VP}$ [$_{VP}$ [$_{CP}$ t]$_l$ versucht] t_i] [$_{CP}$ [$_{IP}$ PRO_k [$_{VP}$ [$_{DP}$ t]$_j$ zu küssen]]]$_l$] t'_i]]]

Example (12a), discussed by Rambow (1994: 17-23), is a sentence containing multiple coherent verbs, which illustrates the recursive application of argument transfer. The structure of the sentence is given in (12b).

(12)a. ?...weil das Fahrrad niemand zu reparieren verspricht zu versuchen.
'because the bicycle (acc) no-one (nom) to repair promises to try'
...because no one promises to try to repair the bicycle

b. [$_{CP1}$ weil [$_{DP}$ das Fahrrad]$_i$ niemand [$_{CP3}$ t_i zu reparieren]$_j$ t_k verspricht [$_{CP2}$ t_j zu versuchen]$_k$]

The direct object *das Fahrrad* of the most deeply embedded infinitive *zu reparieren* is attached to the main clause CP1 (long-distance scrambling). In addition, the CP3 is scrambled out of CP2, which is extraposed after the finite main verb.

The parser proceeds as follows: The three arguments preceding the main verb *verspricht* are attached and inserted into the provisional argument table. When the parser reads *verspricht*, the matching procedure applies. *Das Fahrrad* can be nominative or accusative. Therefore, three readings are temporarily possible: subject, direct object and uninterpreted (direct object of a following infinitival complement). Since *niemand* is non-ambiguously interpreted as the subject of *versucht*, the subject reading for *das Fahrrad* fails. On the one hand, the CP3 *zu reparieren* can be interpreted as sentential complement of the main verb *versucht*, which produces an interpretation of *das Fahrrad* as the long-distance scrambled argument of *zu reparieren*, resulting in the grammatical sentence (13).

(13) ...weil [$_{DP}$ das Fahrrad]$_i$ niemand [$_{CP}$ t_i zu reparieren] verspricht.
because no one promises to repair the bicycle

This interpretation will fail, since the CP2 *zu versuchen* cannot be attached. On the other hand, CP3 can be left uninterpreted; when the CP2 *zu versuchen* is attached, and interpreted as the sentential complement of *versprichi*, the two uninterpreted arguments *das Fahrrad* and *zu reparieren* are transferred to the CP2 for interpretation. The CP3 *zu reparieren* is interpreted as sentential object of *versuchen*, while *das Fahrrad* is regarded as uninterpreted again, and thus is transferred to the CP3, where it is interpreted as the direct object of *reparieren*.

The same strategy of argument transfer holds for ECM constructions in which a subject is scrambled to the upper clause (*cf.* example (6)). In almost the same way, this strategy applies to the arguments of the infinitival clause in raising constructions. While the complements of the infinitival clause are treated in the same way as in the restructuring case, the subject is inserted into the argument table of the raising verb with the grammatical function 'subject', but without thematic role; thus, it is inserted a second time into the argument table as 'uninterpreted'; therefore, it can be treated like other restructuring arguments, *i.e.* it is transferred to the infinitival clause and interpreted as the logical subject of the embedded clause.

(14) ...daß ihn die Frau zu schlagen scheint.
'that him the woman to beat seems'
that the woman seems to beat him.

In (14), the direct object *ihn* of the infinitival verb *schlagen*, being in a scrambled position, is inserted into the argument table of the raising verb and marked as 'uninterpreted'. The subject *die Frau* is inserted into this argument table as the surface subject of *scheint* without a thematic

role and, in addition, as an uninterpreted argument of a following infinitival complement. When the infinitival clause *zu schlagen* is interpreted as the sentential complement of *scheint*, the uninterpreted arguments *ihn* and *die Frau* are transferred to this clause in order to be interpreted with respect to the verb *schlagen*; *die Frau* is taken as the subject of *schlagen* with the thematic role 'agent' and *ihn* as the direct object of the same verb with the thematic role 'patient'.

4 Conclusion

The task of our DIPS parser consists of not only building one or more trees for an input sentence, but also of determining the grammatical function and the thematic interpretation of arguments. We have discussed the parsing strategy in detail and shown that it is adequate for the treatment not only of finite clauses, but also of non-finite clauses. This strategy relies on the following steps: immediate attachment, provisional and definitive interpretation, the testing of constraints, creation of chains, and restructuring. An argument interpretation strategy has been developed, which analyses arguments in a uniform fashion, regardless of whether they precede or follow the verb. This strategy has been extended to handle long-distance scrambling, so that arguments are transferred from the clause in which they are attached to an embedded clause in which they receive an interpretation.

DIPS is a practical system under development, which uses a large-sized lexicon (over 150,000 entries) and which, at present, covers a large range of grammatical constructions such as simple and complex sentences, finite and non-finite clauses, active and passive voice, *wh*-constructions, topicalization, extraposition, scrambling, long-distance dependencies, and verb raising.

References

Abney, S. 1987. *The English Noun Phrase in its Sentential Aspect*, doctoral dissertation, MIT, Cambridge, Mass.

Bader, M & I. Lasser. 1993. "German verb-final clauses and sentence processing: evidence for immediate attachment", ms., in C. Clifton, L. Frazier & K. Rayner (eds.), *Perspectives in Sentence Processing*, Lawrence Erlbaum, Hillsdale N.J.

Berwick, R., S. Abney & C. Tenny (eds.). 1991. *Principle-Based Parsing: Computation and Psycholinguistics*, Kluwer, Dordrecht.

Chomsky, N. 1981. *Lectures on Government and Binding*, Foris Publications, Dordrecht.

Chomsky, N. & H. Lasnik. 1992. "Principles and Parameters Theory", in J. Jacobs, A. van Stechow, W. Sternefeld & T. Vennemann (eds.), *Syntax: An International Handbook of Contemporary Research*, Walter de Gruyter, Berlin.

Haegeman, L. 1994. *Introduction to Government and Binding Theory* Blackwell, Oxford.

Gazdar, G, E. Klein, G. Pullum & I. Sag. 1985. *Generalized Phrase Structure Grammar*, Harvard University Press, Cambridge, Mass.

Grimshaw, J. 1991. *Extended Projections*, ms. Brandeis University.

Kaplan, R. 1980/1986 "Algorithm Schemata and Data Structures in Syntactic Processing", technical report CSL-80-12, Xerox PARC, 1980. reprinted in B. Grosz, K. Sparck Jones & B. Lynn Webber (eds.), *Readings in Natural Language Processing*, Morgan Kaufmann Publishers, Los Altos, 1986, 35-70.

Karttunen L. & M. Kay. 1985. "Parsing in a free word order language", in D. Dowty, L. Karttunen et A. Zwicky (eds.) *Natural Language Parsing*, Cambridge University Press, Cambridge, 279-306.

Kashket, M. 1991. "Parsing Warlpiri, a free word order language", in Berwick et al. (eds.), 123-151.

Kay, M. 1967. "Experiments with a powerful parser", in *COLING-67*.

Laenzlinger, C. & E. Wehrli. 1991. "FIPS: Un analyseur interactif pour le français", *TA informations*, 32.2, 35-49.

Netter, K. 1986. "Getting things out of order (An LFG-proposal for the treatment of German word order)", in *COLING-86*, 494-496.

Rambow, O. 1994. "Formal and computational aspects of natural language syntax", IRCS Report 94-08, University of Pennsylvania.

Uszkoreit, H. 1987. *Word Order and Constituent Structure in German*, CSLI, Lecture notes 8, Stanford University.

Wehrli, E. 1988. "Parsing with a GB grammar", in U. Reyle and C. Rohrer (eds.), *Natural Language Parsing and Linguistic Theories*, Reidel, Dordrecht, 177-201.

Wehrli, E. 1992. "The IPS System", in *COLING-92*, 870-874.