# Syntactic Analyses for Parallel Grammars: Auxiliaries and Genitive NPs

**Miriam Butt** and **Christian Fortmann** and **Christian Rohrer**
Institut für Maschinelle Sprachverarbeitung
Universität Stuttgart
Azenbergstr. 12
70174 Stuttgart, Germany
{mutt|fortmann|rohrer}@ims.uni-stuttgart.de

## Abstract

This paper focuses on two disparate aspects of German syntax from the perspective of parallel grammar development. As part of a cooperative project, we present an innovative approach to auxiliaries and multiple genitive NPs in German. The LFG-based implementation presented here avoids unnessary structural complexity in the representation of auxiliaries by challenging the traditional analysis of auxiliaries as raising verbs. The approach developed for multiple genitive NPs provides a more abstract, language independent representation of genitives associated with nominalized verbs. Taken together, the two approaches represent a step towards providing uniformly applicable treatments for differing languages, thus lightening the burden for machine translation.

## 1 Introduction

Within the cooperative parallel grammar project PARGRAM (IMS-Stuttgart, Xerox-Palo Alto, Xerox-Grenoble), the analysis and representation of structures in the grammars must be viewed from a more global perspective than that of the individual languages (German, English, French). One major goal of PARGRAM is the development of broad coverage grammars which are also modular and easy to maintain. Another major goal is the construction of *parallel* analyses for sentences of the same type in German, English, and French. If this can be achieved, the problem faced by machine translation (MT) could be greatly reduced. Due to the recent development of a faster and more powerful version of the LFG (Lexical-Functional-Grammar) based *Grammar Writer's Workbench* (Kaplan and Maxwell 1993) at Xerox, the implementation of a linguistically adequate, broad coverage grammar appears viable. Given the flexible projection-based architecture of LFG (Dalrymple et al. 1995) and the MT approach presented in Kaplan et al. (1989),[1] a robust MT system is already in place.

In this paper, we concentrate on two issues within the broader perspective of PARGRAM: the treatment of auxiliaries and the transparent representation of multiple genitive NPs in German. These phenomena represent two areas for which generally accepted proposals exist, but whose implementation in the context of parallel grammar development throws up questions as to their wider, crosslinguistic, feasibility. With respect to auxiliaries, the standard *raising* approach that is usually adopted yields undesirable structural complexity and results in idiosyncratic, language particular analyses of the role of auxiliaries. With regard to genitive NPs, the standard analysis for German yields structures which are too ambiguous for a succesful application of machine translation. The following sections present a solution in that morphological wellformedness conditions are stated at a separate component, the *morphology projection*. Furthermore, a representation of argument structure is implemented that is related to, but not identical to the representation of grammatical functions. Language particular idiosyncratic requirements are thus separated out from the language universal information required for further semantic interpretation, or machine translation.

## 2 The Formalism

The architecture of LFG assumed here is the "traditional" architecture described in Bresnan (1982), as well as the newer advances within LFG (Dalrymple et al. , 1995). A grammar is viewed as a set of *correspondences* expressed in terms of *projections* from one level of representation to another. Two fundamental levels of representations within LFG are the c(onstitutent)-structure and the f(unctional)-structure. The c-structure encodes idiosyncratic phrase structural properties of a given language, while the f-structure provides

[1] See also Sadler et al. (1990), Sadler and Thompson (1991), Kaplan and Wedekind (1993), Butt (1994) for further work on MT within LFG.

a language universal representation of grammatical functions (e.g., SUBJect, OBJect), complementation, tense, binding, etc. The correspondence between c-structure and f-structure is not onto or one-to-one, but many-to-one, allowing an abstraction over idiosyncratic c-structure properties of a language (e.g., discontinuous constituents).

In addition, several proposals exploring possible representations of a s(emantic)-structure have been made over the years (e.g. Halvorsen and Kaplan (1988), Dalrymple et al. (1993)). As the realization of a separate semantic component is only planned for the latter stages within PARGRAM, no further discussion of possible formalisms will take place here. It should be noted, however, that rudimentary semantic information, such as argument structure information (lexical semantics), is encoded within the f-structures in order to facilitate transfer in some cases. A case in point is presented in the section on German genitive NPs.

## 3 Auxiliaries — a flat approach

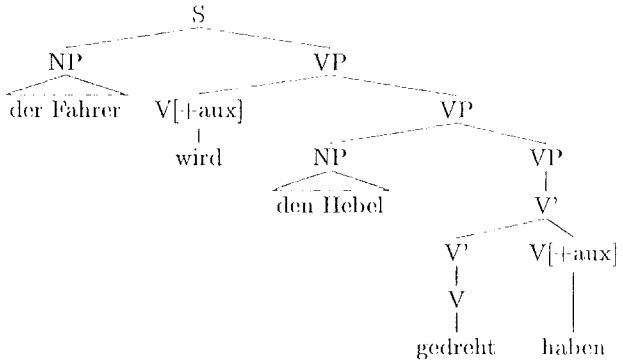
### 3.1 The Received Wisdom

Auxiliaries have given rise to lively debates concerning their exact syntactic status (e.g. Chomsky (1957), Ross (1967), Pullum and Wilson (1977), Akmajian et al. (1979), Gazdar et al. (1982)): are they simply main verbs with special properties, or should they instantiate a special category AUX? Within current lexical approaches (Lexical-Functional-Grammar (LFG), Head-driven Phrase Structure Grammar (HPSG)), auxiliaries (e.g. *have*, *be*) and modals (e.g. *must*, *should*) are treated as *raising* verbs, which are marked as special in some way: in HPSG through an [AUX: +] feature (Pollard and Sag 1994), in LFG (Bresnan 1982) by a difference in PRED value.[2] However, newer work within LFG (Bresnan 1995, T.H. King 1995) has been moving away from the raising approach towards an analysis where auxiliaries are elements which contribute to the clause only tense/aspect, agreement, or voice information, but not a subcategorization frame. This view is also in line with approaches within GB (Government-Binding), which see auxiliaries simply as possible instantiations of the functional category I (see also Halle and Marantz (1993)).

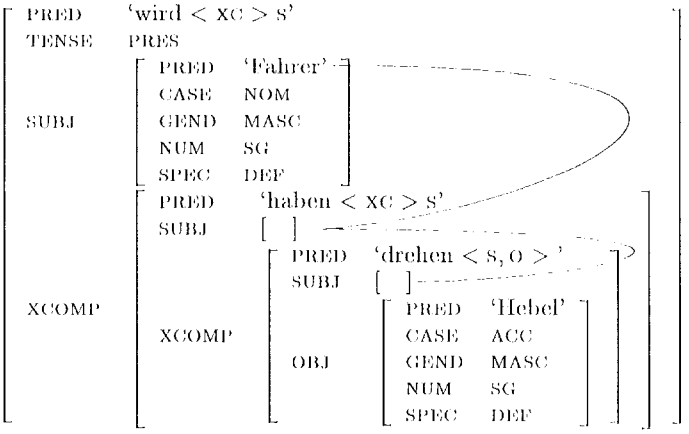
The "traditional" treatment of auxiliaries in both HPSG (Pollard and Sag 1994) and LFG has its roots in Ross's (1967) proposal to treat auxiliaries and modals on a par with main verbs.[3] In particular, auxiliaries are treated as a subclass of raising verbs (e.g. Pollard and Sag (1994), Falk (1984)). For example, a simple sentence like (1) would correspond to the c-structure and f-structure shown in (2) and (3), respectively. Note that the level of embedding in the f-structure exactly mirrors the c-structure: each verbal element takes a complement.

(1) Der Fahrer wird den Hebel gedreht haben
the driver will the lever turned gave
'The driver will have turned the lever.'

(2)




(3)




The main reasons to treat auxiliaries as complement taking verbs in English are: 1) an account of VP-ellipsis, VP-topicalization, etc. follows immediately; 2) restrictions on the nature of the verbal complement (progressive, past participle, etc.) following the auxiliary can be stated straightforwardly (Pullum and Wilson (1977), Akmajian et al. (1979), Gazdar et al. (1982)). The latter point holds for German as well, and in fact, without

[2] See Falk (1984) for an early LFG treatment of 'do' in line with that proposed here, and Abeillé and Godard (1994) for a similar treatment in French.

[3] The term *auxiliary* has often been taken to subsume both modals and elements such as *have* and *be*. However, the distinction between the two is necessary not only semantically, but also syntactically. In German and (some dialects of) English modals can be stacked, while the distribution of auxiliaries is more restricted. Also, assuming that semantic interpretation is driven primarily off of the f-structure, the relative embedding of modals must be preserved at that level in order to allow an interpretation of their scope and semantic force.

some sort of a hierarchical structure, stating well-formedness conditions on a string of multiple auxiliaries becomes wellnigh impossible in light of the greater ordering possibilities granted by the flexible German word order. There are also major reasons, however, for not adopting this analysis: 1) linguistic adequacy; 2) unmotivated structural complexity; 3) non-parallel analyses for predicationally equivalent sentences. Consider the French equivalent of (1) in (4).

(4) Le conducteur aura tourné le levier
the driver will have turned the lever
'The driver will have turned the lever.'

As argued by Akmajian et al. (1979), crosslinguistic evidence indicates that elements bearing only tense, mood, or voice should belong to a distinct syntactic category. In many languages, like French or Japanese, the information carried by *will* (future), or *have* (perfect) is realized morphologically rather than periphrastically. The analysis in (4) thus effectively claims that there exists a deep difference in the predicational structure of auxiliaries like *will* and *have* and the French *aura*.[4] This is not desirable from a crosslinguistic point of view, nor is it helpful for MT.

### 3.2 Alternative Implementation

The approach adopted here is a *flat* analysis of auxiliaries at f-structure ((5)).

(5)

$$\begin{bmatrix} \text{PRED} & \text{'drehen} < \text{SUBJ, OBJ} > \text{'} \\ \text{TENSE} & \text{FUTPERF} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'Fahrer'} \\ \text{CASE} & \text{NOM} \\ \text{GEND} & \text{MASC} \\ \text{NUM} & \text{SG} \\ \text{SPEC} & \text{DEF} \end{bmatrix} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'Hebel'} \\ \text{CASE} & \text{ACC} \\ \text{GEND} & \text{MASC} \\ \text{NUM} & \text{SG} \\ \text{SPEC} & \text{DEF} \end{bmatrix} \end{bmatrix}$$

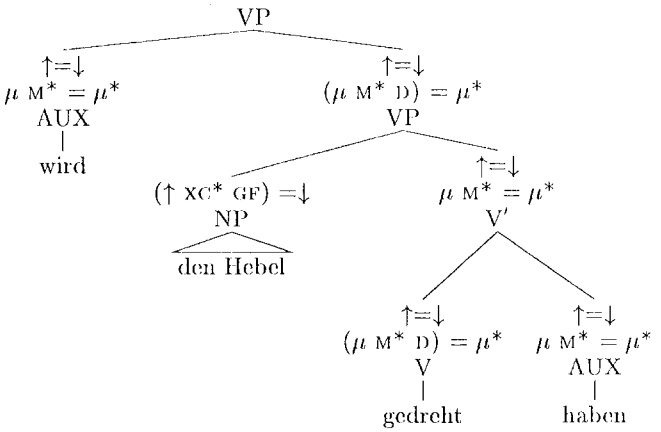
The auxiliaries *wird* 'will' and *haben* 'have' now only contribute information as to the overall tense, but do not subcategorize for complements. Structural phenomena like VP-ellipsis, coordination, or topicalization can, however, still be accounted for in terms of an appropriate embedding at c-structure (cf. (2)). The role of auxiliaries in natural language is now adequately modeled, in particular with respect to a more realistic treatment of tense (compare (3) and (5)), as the French (4) has essentially the same f-structure as (5).[5]

[4]Note that *wird* 'will' is often analyzed as a modal in accordance with Vater (1975). However, the arguments presented there are not conclusive.

[5]The construction of the value for the composed tenses results from a complex interaction between the lexical entries. Note that this treatment does not as yet include a fine-grained representation of tense and aspect. This is the subject of ongoing work. The treatment presented here provides the basis needed for a thorough crosslinguistic analysis of temporal and aspectual phenomena.

However, the flat f-structure in (5) provides no room for a statement of selectional requirements, allowing massive overgeneration (e.g. nothing blocks the presence of two *haben* in (1)). Neither can the particular order of auxiliaries be regulated. Our solution takes advantage of LFG's flexible projection-based architecture by implementing a projection which models the hierarchical selectional requirements of auxiliaries, yet does not interfere with the subcategorizational properties of verbs, as would be the case under a raising analysis.

(6)




In LFG, the flexible word order of German is handled via *functional uncertainty*, which characterizes long-distance dependencies without resorting to movement analyses (Netter (1988), Zaenen and Kaplan (1995)). As in (6), which illustrates our alternative solution, functional uncertainty is represented by the Kleene Star (*).[6] The annotation on the NPs indicates that they could fulfill the role of any possible grammatical function (GF), e.g. SUBJ or OBJ, and that the level of embedding ranges from zero to infinite. With every auxiliary subcategorizing for an XCOMP, the two NPs could conceivably be arguments of three different verbs: *wird*, *haben*, or *gedreht*. Thus, the greater structural complexity unnecessarily increases the search space for the determination of a verb's arguments. In (6), however, the m-structure is projected from the c-structure parallel to the f-structure through annotations similar to the usual f-structure annotations.[7] Statements about "mor-

[6]For space reasons, the XC indicates XCOMP, the D a DEP.

[7]The annotation $\mu$ M* in (6) refers to the m-structure associated with the parent c-structure node, and $\mu$* refers to the m-structure associated with the daughter node. The more familiar ↑ and ↓ of LFG are simply shorthand notations of the same idea,

phological" dependents (DEP) are thus decoupled from functional uncertainty: the relation of NP arguments to their predicator now does not extend through various layers of artificial structural complexity (XCOMPs). For VP-topicalization or extraposition an unbounded long-distance dependency must still be assumed. However, as the functional uncertainty path for auxiliaries is distributed only over the m-structure of the verb complex ($(\mu \uparrow$ DEP$^*) = \mu \downarrow$), and does not involve the resolution of the role of NP arguments, there are in fact differing paths of functional uncertainty involved. The dependencies between predicators and their arguments and auxiliaries and their dependents are thus neatly factored out. The m-structure corresponding to the matrix VP in (6) is (7). The desired flat f-structure resulting from the usual $\uparrow$ and $\downarrow$ annotations is as in (5).

(7)

$$\begin{bmatrix} \text{AUX} & + \\ \text{FIN} & + \\ \text{DEP} & \begin{bmatrix} \text{AUX} & + \\ \text{FIN} & - \\ \text{VFORM} & \text{BASE} \\ \text{DEP} & \begin{bmatrix} \text{FIN} & - \\ \text{VFORM} & \text{PERFP} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Like the f-structure, the m-structure is an attribute-value matrix. It encodes language-specific information about idiosyncratic constraints on morphological forms. The m-structure is not derived from the f-structure. Rather, both representations are in simultaneous correspondence with the c-structure. The following (abbreviated) lexical entry exemplifies the pieces of information needed. The disjunctive lexical entry for *wird* 'will' in (8) takes the various combinatory possibilities of auxiliaries and main verbs into account, and provides the appropriate tense feature. For example, it requires that the embedded VFORM be BASE, and that there be no passive involved for a simple future like *wird drehen*.

(8)

wird AUX
($\mu$ M$^*$ AUX) $=+$
{ ($\mu$ M$^*$ DEP VFORM) $=_c$ BASE
($\mu$ M$^*$ DEP DEP VFORM) $\neq$ PERFP
($\uparrow$ PASSIVE) $\neq +$
"simple future: wird drehen"
($\uparrow$ TENSE) $=$ FUT
|
($\mu$ M$^*$ DEP VFORM) $=_c$ BASE
($\mu$ M$^*$ DEP DEP VFORM) $=_c$ PERFP
($\uparrow$ PASSIVE) $\neq +$
"future perfect: wird gedreht haben"
($\uparrow$ TENSE) $=$ FUTPERF }

but restricted to the projection from c-structure to f-structure: $\uparrow = \phi$ M$^*$, $\downarrow = \phi$ $^*$.

Features needed only to ensure language particular wellformedness are no longer unified into the f-structure, cluttering a representation that is meant to be language independent. In our analysis, only features needed for further semantic interpretation, MT, or for the expression of language universal syntactic generalizations are represented at f-structure. For example, morphologically encoded information like case, gender, or agreement is needed for statements as to binding, predicate-argument relations, or the determination of complex clause structures (given that agreement is generally clause-bounded), and is therefore represented at f-structure. Wellformedness conditions on adjective inflection or relative pronoun agreement, however, can now be stated on the m-structure as idiosyncratic, language particular information which can be ignored for purposes of MT or semantic interpretation.

## 4 Multiple Genitive NPs

The differing surface realization of genitives within NPs in English (preverbal NPs, postverbal PPs), French (postverbal PPs), and German (preverbal NPs, postverbal PPs or NPs), poses a particular challenge for a parallel grammar development project like PARGRAM. In this paper, we suggest a treatment of multiple genitive NPs which not only accounts for some restrictions on their distribution within German, but also allows a language independent (universal) representation, thus facilitating MT.

In general, the distribution of multiple NPs within NPs is an area of German syntax which has not received a satisfactory account to date (e.g., Pollard and Sag (1994), Bhatt (1990), Haider (1988)). In German, nouns generally have at most one genitive which may occur in a prenominal or postnominal position adjacent to the noun. Both kinds of genitives have the same morphological shape. However, nominalizations that are derived from a transitive verb allow for two genitives, one in the prenominal, the other in the postnominal position.

The function of a genitive is generally expressed as indicating a possessor: POSS within LFG. However, in the case of two genitives, the assignment of two POSS values violates the uniqueness-condition on f-structures and is furthermore insufficient to distinguish the two differing kinds of genitives. We therefore propose the utilization of two functions named GEN1 and GEN2 in order to avoid association with any specific semantic role. Furthermore, as genitives in the NP are generally optional, they are taken to express no governed functions, i.e., they are not subcategorized for by the noun. So GEN1 and GEN2 are semantic functions in LFG on a par with, say, adjuncts. The NP rule for German

then is (9).[8]

(9) NP → ({DET: ↑=↓
| NP: (↑ GEN1) =↓ })
N: ↑=↓
(NP: (↑ GEN2) =↓)

If the head-noun is not derived from, say, a verb, the single genitive in either position is interpreted as a possessor. In case of a derived nominal, however, a genitive is interpreted according to the thematic roles assigned to the arguments of the verbal base. That means the functions GEN1 and sc gen2 have to be linked to the appropriate roles. Neither of the two functions is in principle restricted to any specific role. But if both genitives are present they must be interpreted according to a thematic role hierarchy.

As (10) shows, if only one genitive is present, its prenominal interpretation may be as agent or as patient. A postnominal (single) genitive is interpreted as agent if the head noun is derived from an intransitive, and as a patient/theme if derived from a transitive.

(10) a.
Elisabeths Lachen
Elisabeth-Gen laughing
'Elisabeth's laughter'

b. Roms Belagerung
Rome-Gen siege
'Rome's siege'

However, if two genitives occur, as in (11), the prenominal genitive is restricted to an agent, and the postnominal one to patient. This restriction must be encoded at some level, but does not follow from the distiction between GEN1 and GEN2, which are functions that do not bear any semantic content on their own.

(11)
Karls Behandlung Peters
Karl-Gen treatment Peter-Gen
'Karl's treatment of Peter'

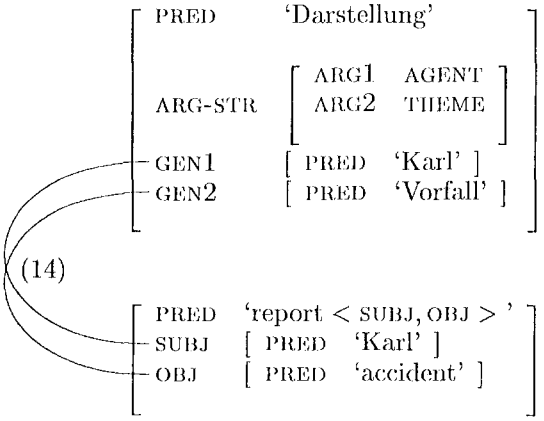
To obtain the correct linking, the argument structure of the verbal base must be available. Since MT is based on f-structures within PARGRAM, the argument structure has to be present at this level of representation.[9] Nominalization is therefore implemented as a morphologically driven process (lexical rule) which eliminates SUBJ and OBJ from the verb's subcategorization frame and enters the verb's argument structure into the lexical entry of the noun. This yields the optionality of genitives while preserving the underlying semantics, as shown in (12). The association of GEN1 and GEN2 then is determined according to a hierarchical order of arguments (Bresnan, 1995).

This approach also provides a means of handling certain cases of categorial shift. For instance, in German temporal and conditional adjuncts may be realized as PPs dominating an NP headed by a deverbal noun. English does not have this option, but employs an adjunct-clause instead. Here, the GEN1 and GEN2 functions of the German f-structure have to be related correctly to the SUBJ and OBJ functions of the English f-structure.

(12)
bei Karls Darstellung des Vorfalls
at Karl-Gen report the accident-Gen
mussten alle lachen
must-Past all laugh
'when Karl reported the accident
everyone had to laugh'

Here the linking of the GEN1 and GEN2 functions to the appropriate thematic role in the German f-structure drives the transfer of these functions to the SUBJ and OBJ functions of the English f-structure.

(13)




(14)




Under this approach, languages now only differ with respct to the categorial realisation of the function by case-marked NP or PP. Thus, an application of this treatment not only provides an adequate grammatical analysis of the NP in German, but also facilitates MT.

## 5 Conclusion

This paper has presented innovative approaches for two particular syntactic phenomena: auxiliaries and multiple genitive NPs. The analyses proposed allow the factorization of language particular, idiosyncratic information. This results in a cleaner treatment of auxiliaries by factoring out morphological wellformedness conditions, and allows for the preservation of argument structure

[8] Abstracting away from bar-level considerations and further optional constituents, this rule captures the restrictions that determine the dislocation of a genitive in the matrix NP.

[9] If a semantic or argument projection is assumed (e.g., Halvorsen and Kaplan, 1988), this information may be represented at another independent projection.

information in cases like that of the German multiple genitive NP construction, where syntactically dissimilar constructions express essentially the same predicate-argument relations. As such, the work presented here can be seen as a small but necessary step towards the realization of a broad coverage grammar. In particular, the feasibility of developing parallel grammars for differing languages is greatly increased through the formulation of uniformly applicable, transparent analyses.

## 6 Acknowledgments

We would like to acknowledge Judith Berman, Mark Johnson, Ron Kaplan, María-Eugenia Niño and Annie Zaenen for the many valuable discussions that served as input to this paper.

## References

Akmajian, A., S. Steele, and T. Wasow. 1979. The Category AUX in Universal Grammar. *Linguistic Inquiry* 10(1):1–64.

Bhatt, C. 1990. *Die syntaktische Struktur der Nominalphrase im Deutschen.* Tübingen: Narr.

Berman, J., and A. Frank. To appear. *Deutsche und Französische Syntax im Formalismus der LFG.* Tübingen: Niemeyer.

Bresnan, J. 1995. *Lexical-Functional Syntax.* Lecture Notes, ESSLI 7, Barcelona.

Bresnan, J. (Ed.). 1982. *The Mental Representation of Grammatical Relations.* Cambridge, MA: MIT Press.

Butt, M. 1994. Machine Translation and Complex Predicates In *Proceedings of KONVENS 94*, 62–72. Vienna, September.

Dalrymple, M., R. Kaplan, J. Maxwell, and A. Zaenen. (Eds.). 1995. *Formal Issues in Lexical-Functional Grammar.* Stanford, CA: CSLI Publications.

Dalrymple, M., J. Lamping, and V. Saraswat. 1993. LFG Semantics via Constraints. In *EACL 6.*

Falk, Y. 1984. The English Auxiliary System: a Lexical-Functional Analysis *Language* 60(3):483–509.

Haider, H. 1988. Die Struktur der deutschen NP. *Zeitschrift für Sprachwissenschaft* 7(1):32–79.

Halle, M., and A. Marantz. 1993. Distributed Morphology and the Pieces of Inflection. In K. Hale and S.J. Keyser (Eds.), *The View from Building 20*, 111–176. Cambridge, Massachusetts: The MIT Press.

Halvorsen, P.-K., and R. Kaplan. 1988. Projections and semantic description in Lexical-Functional Grammar. In *Proceedings of the International Conference on Fifth Generation Computer Systems*, 1116–1122.

Kaplan, R., K. Netter, J. Wedekind, and A. Zaenen. 1989. Translation by structural correspondences. In *EACL 4*, 272–281. University of Manchester.

Kaplan, R., and J. Wedekind. 1993. Restriction and correspondence-based translation. In *EACL 6.* University of Utrecht, April.

Kaplan, R., and J. Maxwell. 1993. Grammar Writer's Workbench Ms., Xerox Corporation.

King, T. H. 1995. *Configuring Topic and Focus in Russian.* Stanford, California: CSLI Publications.

Netter, K. 1988. Non-local dependencies and infinitival constructions in German. In U. Reyle, and C. Rohrer(Eds.), *Natural language parsing and linguistics theories.* Dordrecht: Reidel.

Pollard, C., and I. Sag. 1994. *Head-Driven Phrase Structure Grammar.* Chicago, IL: The University of Chicago Press.

Pullum, G., and D. Wilson. 1977. Autonomous Syntax and the Analysis of Auxiliaries. *Language* 53(4):741–788.

Ross, J. 1967. *Constraints on Variables in Syntax.* PhD thesis, MIT.

Sadler, L., I. Crookston, D. Arnold, and A. Way. 1990. LFG and translation. In *Proceedings of the Third International Conference on Theoretical and Methodological Issues in Machine Translation*, 11–13. Linguistic Research Center, University of Texas at Austin.

Sadler, L., and H. Thompson. 1991. Structural non-correspondence in translation. In *EACL 5*, 293–298. Berlin.

Vater, H. 1975. *Werden* als Modalverb. In Calbert, and H. Vater (Eds.), *Aspekte der Modalität.* Tübingen: Narr.