# Word order variation in German main clauses: A corpus analysis

**Andrea Weber**
Computational Linguistics
Saarland University
PO Box 151150
66041 Saarbrücken, Germany
aweber@coli.uni-sb.de

**Karin Müller**
Informatics Institute
University of Amsterdam
Kruislaan 403
1098 SJ Amsterdam, The Netherlands
kmueller@science.uva.nl

## Abstract

In this paper, we present empirical data from a corpus study on the linear order of subjects and objects in German main clauses. The aim was to establish the validity of three well-known ordering constraints: *given* complements tend to occur before *new* complements, *definite* before *indefinite*, and *pronoun* before *full noun phrase* complements. Frequencies of occurrences were derived for subject-first and object-first sentences from the German Negra corpus. While all three constraints held on subject-first sentences, results for object-first sentences varied. Our findings suggest an influence of grammatical functions on the ordering of verb complements.

## 1 Introduction

Word order variation has been described in great detail by theoretical linguists. There is, however, also an increasing interest in the topic from both computational linguists and psycholinguists (e.g. Kaiser and Trueswell (submitted); Keller (2000); Kruiff and Duchier (2003); Pechmann et al. (1996); Röder et al. (2000)). Nevertheless, the empirical resources that researchers can draw from for their studies are still very limited, since only a few studies report on the actual amount of word order variation (see Kempen and Harbusch (2004); Kurz (2000)). This paper therefore presents a corpus study on the linear order of subjects and objects in German, and factors related to the positioning of complements before or after the verb. Our study is also new in that it looks at main clauses rather than the *Mittelfeld*, for which most ordering principles were originally intended.

German is a language with a relatively free word order in which the subject usually precedes the object, but can also follow it: In (1), the subject "Turnverein Neurönnebeck" precedes the object "Faireßpokal"; in (2) the same object precedes the subject, without changing the original meaning of the sentence.

(1) SVO
**Der Turnverein Neurönnebeck** gewann **den Faireßpokal**.
*The club Neurönnebeck (NOM) won the fairness price (ACC).*

(2) OVS
**Den Faireßpokal** gewann **der Turnverein Neurönnebeck**.
*The fairness price (ACC) won the club Neurönnebeck (NOM).*

For a German newspaper corpus, we investigate subject-verb-object (SVO) and object-verb-subject (OVS) sequences, and examine the extent to which certain ordering constraints influence the positioning of verb complements. In particular, we investigate the validity of the well-known constraints to place given before new, definite before indefinite, and pronoun before full noun phrase (NP) complements (cf. Müller (1999); Uszkoreit (1987)) using the Negra corpus (Brants et al., 1999).

## 2 Ordering Principles

In this section, we describe the three ordering principles tested in this study. Whereas the first principle attends to the contextual dependencies of a sentence, the scope of the second and the third principle is a sentence.

In scrambling languages, the position of verb complements can reflect their connection to the preceding context. In these languages, discourse new information tends to occur towards the end of a sentence, whereas discourse old information is more likely to occur at the beginning (cf. Birner and Ward (1998)). Thus, the information structure of a sentence in a scrambling language such as German, can reflect its fit within a given discourse (Selkirk (1984); Steedman (1991))[1]: When an object precedes a subject, the object is likely to be given and the

[1]Different terms and concepts such as theme/rheme, background/focus and given/new are used in the lit-

subject new; when a subject precedes an object, the subject is likely to be given and the object new, although the canonical subject-first order is also expected when both complements are either new or given. We establish for Negra how often both SVO and OVS main clauses adhere to this basic order pattern.

German is also a language with definite and indefinite articles. According to a second linear ordering principle, definite NPs should tend to precede indefinite NPs. We also presume, that definiteness is correlated to the information status of complements. As already Chafe (1976) pointed out, indefiniteness often goes together with newness, and definiteness with givenness or newness. Thus, on the NP itself information status can be partially encoded by the choice of article. If the correlation with givenness drives the positioning of definite and indefinite complements, we should find for both SVO and OVS sentences that definite complements tend to precede indefinite ones. Another possibility is, however, that definiteness is bound to grammatical functions (i.e. subjects are usually definite). In that case, we would expect to find a reversal of the ordering principle for definiteness in OVS sentences.

A third common linear ordering principle states that pronouns tend to precede full NPs. Similarly to definiteness, the use of pronouns can potentially encode information structure. Almost by definition, pronouns refer to an antecedent in the discourse and represent therefore given information (with the possible exception of indefinite pronouns). Whereas pronoun complements usually represent given information, full NP complements are not necessarily new. Again, if the correlation with givenness drives the positioning of pronominalized complements, we should find that pronouns tend to precede full NPs in both SVO and OVS sentences. If, however, for example grammatical functions determine which complements are pronominalized, we might not find this tendency.

## 3 Corpus Analysis

The Negra corpus (Skut et al., 1998) is an annotated collection of 20,602 sentences (355,096 tokens) extracted from the German newspaper Frankfurter Rundschau. The syntactic structure of sentences is represented in dependency trees for which the nodes describe constituents and the edges between the nodes are labeled with grammatical functions expressing syntactic relations. For our study, we choose the 'Penn format', a transformation of the original Negra treebank, in which crossing edges and traces are omitted.

### 3.1 Data extraction

Using a Perl program and the tree-search program Tgrep2 (Rhode, 2002) all OVS sentences in Negra are extracted by looking for object-verb-subject sequences with the same depth of embedding. Object and subject themselves as well as the sentences in which the OVS structure occur could be complex (see 3).

(3) OVS.
**Den Satz** von der Vergangenheit, die noch nicht einmal vergangen sei, **zitiert** auch **Peter Rühl** in seinem wie stets gescheiten Begleittext zur jüngsten CD des Trompeters Frank Koglmann, einem der wichtigsten Musiker des europäischen Jazz.
*The sentence (ACC) about the past that has not even passed yet, cites also Peter Rühl (NOM) in his as always smart accompanying text to the latest CD of the trumpet player Frank Koglmann, one of the most important musicians in European jazz.*

Clausal objects with verbal constructions in addition to direct and indirect questions are manually omitted from the list after extraction. A total of 625 OVS sentences are kept for analysis (3% of all sentences in Negra). Next, comparable 625 (out of 2773) SVO sentences are chosen.[2] Since the total number of SVO sentences in Negra exceeds that of OVS sentences notably, for practical reasons a subset of 625 SVO sentences is selected. Since the selection is random, we assume that findings within the subset are generally valid for SVO sentences. In addition, for each selected OVS and SVO sentence, the two immediately preceding sentences in Negra are extracted and serve as context to determine the information status (given or new) of complements.

erature to express information structure (for a recent overview see Kruiff-Korbayová and Steedman (2003)). Since annotations in the present study are based on single referents rather than parts of sentences, we distinguish between given and new.

[2] We do not allow SVO sentences in which the object is a reflexive pronoun (e.g. "er fürchtet sich", *he is afraid*), since reflexive pronouns are most unlikely to be fronted. In fact, none of our OVS sentences contained a fronted reflexive pronoun. Furthermore, reflexive pronouns express coreference within a clause, whereas we are interested in references across sentence boundaries.

### 3.2 Data Coding

For each extracted OVS and SVO sentence, the authors annotate information status, definiteness, and pronominalization of its verb complements. For complex subjects and objects, annotations are based on the semantic head of the complement (i.e. noun or pronoun if the semantic head coincides with the syntactic head). Definiteness of NPs is assigned by the determiner and information status of NPs by the noun. Whereas it seems obvious to base annotations on the head of complex complements, the decision is less clear in the case of more than one equivalent NP within a complement (i.e. coordinated NPs). Rather than annotating all NPs, for comparability reasons, we base annotations on the first NP only, whenever a complement consists of a listing of more than one NP. Furthermore, some SVO sentences contain both direct and indirect objects, though none of our OVS sentences do. For these SVO sentences, only the direct object are considered for the annotations. An exception are SVO sentences with reflexive pronouns as indirect object, for which annotations are based on the direct object (1.6% of all SVO sentences).

**Givenness**. Two preceding context sentences are used to determine whether verb complements present new or given information.[3] We code complements as given if they present accessible information (Lambrecht, 1994). Accessability can either be textually or inferentially provided. Textual accessability requires an explicit coreferential antecedent (i.e. the occurrence of the same lemma in the context). Inferentially accessible complements do not require an explicit antecedent. Such *inferables* (Prince, 1981) are assumed to be activated via *bridging inferences* (Clark, 1977) that logical relations such as synonymy, hyponymy, and meronomy can provide. In such cases, shared general knowledge of the relations between objects and their components or attributes is assumed. Whenever more specific knowledge is required to establish such a relation, however, complements are considered to be new. The distinction between general and specific knowledge is particularly hard to maintain, since the distinction is often clearly not binary. For instance, geographic familiarity with the catchment area of the Frankfurter Rundschau is considered specific knowledge: In (4), "Waldstadion" is one of the local soccer stadiums in Frankfurt. Even though many local readers of the Frankfurter Rundschau will know this it can not be assumed to be known by all readers of the newspaper, and is therefore coded as new information.

(4) Frankfurt - Waldstadion

Moreover, when two entities X and Y of a potentially larger group Z are considered equally specific, Y is coded as new information after X is mentioned in the context (see (5)): Here, "Klassik" and "Jazz" are two examples of music styles.

(5) Klassik - Jazz
classical music - jazz

A special case form constructions with "es". They are almost exclusively used impersonally in sentences such as "Karten gibt es", *tickets are available*. We then annotate "es" as new information.

**Definiteness**. For all complements, we annotate whether they are definite or indefinite. We largely follow the classification suggested by Prince (1992). Markers of definite complements are definite articles, demonstrative articles, possessive articles, personal pronouns, and unmodified proper names. Markers of indefinite complements are indefinite articles, zero articles, quantifiers, and numerals. Note, that all quantifiers are marked as indefinite even though certain quantifiers like *all* and *every* have been suggested in the literature to mark definite descriptions. Furthermore, certain syntactically indefinite DPs have been argued to be semantically definite and syntactically definite DPs to be semantically indefinite. In our study, however, only formal syntactic properties are critical for the assignation of definiteness.

**Pronominalization**. For the annotation of pronominalization, we check whether complements are realized as pronouns or full NPs.

## 4 Results

### 4.1 Givenness

In 74 cases, the antecedent of an anaphoric complement occurs prior to our context window of two sentences. Sentences containing such complements are excluded from the analysis of

[3] A recent study by Morton (2000) showed for singular pronouns, that 98.7% of the times the antecedents were available within two preceding sentences. His findings are similar with those reported by Hobbs (1976). Since the context in our study regularly introduces complements, we believe that it gives an adequate picture of the interaction between information status and word order.

givenness. In addition, 106 sentences with "es"-complements are excluded for the analyses we present in this paper. Table 1 shows the observed orderings of given and new complements for our set of SVO and OVS sentences.

(SVO)

| first NP \ second NP | given O | new O |
|---|---|---|
| given S | 113 | **187** |
| new S | **88** | 175 |

(OVS)

| first NP \ second NP | given S | new S |
|---|---|---|
| given O | 96 | **144** |
| new O | **134** | 170 |

Table 1: Frequency of subject (S) and object (O) pairs ordered by givenness.

In SVO sentences, given subjects precede new objects more often (187 times) than new subjects precede given objects (88 times). This tendency is in compliance with the linear ordering principle for information structure, even though the principle is not strictly obeyed as the 88 cases of new-before-given complements show. In OVS sentences, both orders occur about equally often. Given objects precede new subjects 144 times, and new objects precede given subjects 134 times. A chi-square test confirms a significant interaction between sentence type (SVO, OVS) and ordering (given-before-new, new-before-given; $\chi^2(1) = 14.44$, $p < .001$). Thus, in contrast with for example Finnish (Kaiser and Trueswell, submitted), information structure seems not to be encoded in German OVS sentences, in the sense that fronted given objects do not cue upcoming new subjects for language perceivers. Obviously, factors other than givenness must have influenced the fronting of objects as is also apparent by the frequent occurrence of OVS sentences with given-before-given (96 times) and new-before-new (170 times) ordering of complements. If not, the canonical SVO order would be expected.

We want to point out a second way of looking at the results, one that involves a more language producer-oriented view. Discourse context defines the information status of complements. Supposing now that a subject has been introduced in a context, but not an object, we can check which sentence structure occurs more often. We find more SVO (187 times) than OVS sentences (134 times). On the other hand, when the object of a sentence is given, but the subject new, we find more OVS (144) than SVO (88) sentences ($\chi^2(1) = 21.45$, $p < .001$). Information status of complements seems to have influenced the choice of word order. However, this interpretation must be taken with caution. First, we only look at a subset of all SVO sentences of Negra. Second, at least to a certain degree language producers can not only choose word order but also the grammatical function (subject or object) of discourse referents. The assignment of grammatical functions to constituents is assumed to happen during the functional stage of grammatical encoding in sentence production; only at a later positional stage the linear order is determined (e.g. Bock and Levelt (1994)).

### 4.2 Definiteness

One hundred and six sentences containing "es"-complements as well as four sentences in which the object or subject is a citation are excluded from the analysis. Table 2 shows the observed orderings of definite and indefinite complements.

(SVO)

| first NP \ second NP | def O | indef O |
|---|---|---|
| def S | 237 | **242** |
| indef S | **50** | 62 |

(OVS)

| first NP \ second NP | def S | indef S |
|---|---|---|
| def O | 286 | **48** |
| indef O | **190** | 25 |

Table 2: Frequency of subject (S) and object (O) pairs ordered by definiteness.

In SVO sentences, definite NPs precede indefinite NPs 242 times but the reverse ordering occurs only 50 times. Thus, the basic order of definite before indefinite NPs is largely met for SVO sentences. For OVS sentences, however, the preference to place definite NPs before indefinite NPs is reversed. Only 48 times precede definite NPs indefinite NPs, but 190 times precede indefinite NPs definite NPs ($\chi^2(1) = 205.58$, $p < .001$). Thus, the ordering principle for definiteness is violated in OVS sentences. Rather, the results suggest a strong correlation between

grammatical function and definiteness [4]: Subjects are more often definite and objects indefinite, regardless of sentence type. Considering all four ordering possibilities (see Table 2), however, this tendency is much stronger for subjects than objects.

**Definiteness and givenness.** Not unexpectedly, definiteness is significantly correlated with givenness for all complements in both sentence types (all p-values in Pearsons tests $< .01$). At a closer look, indefinite NPs represent more often new information (72%), whereas definite NPs present given information (52%) as often as new information. This result matches corpus studies in other languages, which found that indefiniteness entails newness whereas definiteness can entail both givenness and newness (see e.g. Fraurud (1990)).

**Definiteness, givenness, and word order.** We are also interested in whether the positioning of a complement before or after the verb is influenced by its information status. Is a definite complement, for example, more likely to occur before the verb if it also is given? Table 3 shows the number of occurrences of both definite-indefinite and indefinite-definite orders, split by the information status of the complements. For both sentence types (SVO and OVS), neither the positioning of the definite complement nor the positioning of the indefinite complement is affected by information status (in chi-square tests all p-values $> .3$). Thus, in SVO sentences, definite subjects precede indefinite objects more often than indefinite subjects precede definite objects, regardless of whether subjects and objects present given or new information. Similarly, in OVS sentences, indefinite objects precede definite subjects more often than the reverse, regardless of the information status of the complements.

### 4.3 Pronominalization

As with definiteness, we exclude 106 sentences containing "es"-complements. Table 4 shows the observed orderings of pronoun and full NP complements. In SVO sentences, 95 times a pronoun precedes a full NP, whereas a full NP precedes a pronoun only 33 times. Thus, as with givenness and definiteness before, the basic order of pronoun complements before full

[4]Since we only look at sentences in active and not in passive voice, subjects in our sentences are always agents and objects patients. We can therefore not exclude the possibility that thematic roles rather than grammatical functions drive determiner choice.

(SVO)

| | def_given | def_new | indef_given | indef_new |
|---|---|---|---|---|
| def<indef | 147 | 95 | 71 | 171 |
| indef<def | 26 | 24 | 16 | 34 |

(OVS)

| | def_given | def_new | indef_given | indef_new |
|---|---|---|---|---|
| def<indef | 26 | 22 | 15 | 33 |
| indef<def | 85 | 105 | 57 | 133 |

Table 3: Linear order frequency of definite-indefinite pairs for given and new complements.

NP complements is largely met for SVO sentences. In OVS sentences pronouns precede full NPs 76 times, but the reverse order also occurs 91 times ($\chi^2(1) = 23.35$, $p < .001$). Interestingly, our results differ from what Kempen and Harbusch (2004) found for subordinate clauses in Negra. In adverbial and complement OVS clauses, they found that full NP objects never precede pronominalized subjects and translated this findings into a rigid rule schema. They argued that only strong conceptual influences such as topic/focus relations could override the ordering pattern. Such influences would then be more likely to play a role in main clauses, as we test them, than in subordinate clauses, since we do observe full NP objects preceding pronominalized subjects. However, this conclusion is based on a relatively small set of sentences and needs to be verified in a corpus larger than Negra.

(SVO)

| first NP \ second NP | pro O | full O |
|---|---|---|
| pro S | 12 | **95** |
| full S | **33** | 450 |

(OVS)

| first NP \ second NP | pro S | full S |
|---|---|---|
| pro O | 20 | **76** |
| full O | **91** | 366 |

Table 4: Frequency of subject (S) and object (O) pairs ordered by pronominalization.

**Pronominalization and givenness.** As expected, pronominalization is highly correlated with givenness for all complements in both sentence types (all p-values in Pearsons tests $< .001$). Almost by definition, pronouns are given, except for a few cases in which the

referent of a pronoun follows rather than precedes it within the same sentence. On the other hand, clearly not all given complements are pronouns. In fact, only 33% of all given subjects and 26% of all given objects are pronouns.

**Pronominalization, givenness, and word order.** Table 5 shows the number of occurrences of both pronoun-full NP and full NP-pronoun orders, split by the information status of the complements. For both SVO and OVS sentences, the positioning of the pronoun is affected by its information status (for SVO: $\chi^2(1) = 16.92$, $p < .001$; for OVS: $\chi^2(1) = 4.76$, $p < .03$). Thus, the givenness of the pronoun significantly increases the likelihood for this complement to precede the other complement. No such effect is found for full NPs.

(SVO)

| | pro_given | pro_new | full_given | full_new |
|---|---|---|---|---|
| pro<full | 92 | 3 | 28 | 67 |
| full<pro | 23 | 10 | 14 | 19 |

(OVS)

| | pro_given | pro_new | full_given | full_new |
|---|---|---|---|---|
| pro<full | 69 | 7 | 30 | 46 |
| full<pro | 70 | 21 | 37 | 54 |

Table 5: Linear order frequency of pronoun-full NP pairs for given and new complements.

## 5 Discussion and Conclusion

We present in this paper a corpus-based study on the linearization of subjects and objects in German main clauses. We examined the extent to which the parameters of givenness, definiteness, and pronominalization influence the ordering of verb complements in German SVO and OVS sentences.

In general, our corpus data only support the validity of the ordering principles for SVO sentences: given subjects indeed tend to precede new objects, definite subjects indefinite objects, and pronominalized subjects full NP objects. However, clearly none of the ordering constraints is absolute since the reversed orders were also observed, just not as often. For OVS sentences, our results differed. None of the three basic order patterns for givenness, definiteness, and pronominalization was confirmed: For givenness and pronominalization both orders (given-new and new-given; pronoun-full NP and full NP-pronoun) occurred about equally often. For definiteness the basic order preference was reversed (more indefinite-definite than definite-indefinite orders). The fact that in OVS sentences indefinite objects preceded definite subjects more often than the converse, suggests that grammatical functions rather than the linear structure of sentences influence the choice of word order (i.e. subjects are more likely to be definite, regardless of word order). Furthermore, even though both definiteness and pronominalization were correlated with givenness, only the positioning of pronouns before or after the verb was additionally influenced by its information status. Definiteness did not interact with information structure as we would have expected.

In sum, our data indicate that linearization principles are soft constraints, and that a combination of principles rather than one primary constraint impact the choice of word order. The fact, that for canonical SVO sentences cases of reversed basic order patterns (e.g. new subject preceded given object) were observed for all three tested linearization principles, suggests that ordering constraints other than the ones under investigation here influence the linearization of complements. Indeed, in the literature a range of such parameters has been suggested, including animacy and length of complements (Dietrich and Nice (in press); Hawkins (1994)). A recent Negra corpus study by Kempen and Harbusch (2004), confirmed the direct influence of animacy on linearization in German subordinate clauses. On the other hand, a Negra corpus study by Kurz (2000) found no influence of length on the ordering of subject and object.

## 6 Acknowledgements

We thank Stefan Baumann, Matthew Crocker, and Amit Dubey for helpful comments on an earlier version of this paper. This research was supported by SFB 378 "ALPHA", awarded by the German Research Council and by the Pioneer Project "Computing with Meaning", funded by the Netherlands Organization for Scientific Research.

## References

Betty Birner and Geoffrey Ward. 1998. *Information status and noncanonical word order in English.* John Benjamins, Amsterdam.

Kay Bock and Wilhelm Levelt. 1994. Language production: Grammatical encoding. In

M. Gernsbacher, editor, *Handbook of Psycholinguistics*. Academic Press, San Diego.

Thorsten Brants, Wojciech Skut, and Hans Uszkoreit. 1999. Syntactic annotation of a German newspaper corpus. In *Proceedings of the ATALA Treebank Workshop*, pages 69–76, Paris, France.

Wallace Chafe. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In C. Li, editor, *Subject and Topic*, pages 25–55. Academic Press, New York, NY.

Herbert Clark. 1977. Bridging. In P.N. Johnson-Laird and P.C. Wason, editors, *Thinking: Readings in Cognitive Science*, pages 411–420. Cambridge University Press, London, New York.

Rainer Dietrich and Kathy Nice. in press. Belebtheit, Agentivität und inkrementelle Satzproduktion. In C. Habel and T. Pechmann, editors, *Sprachproduktion*. Westdeutscher Verlag, Wiesbaden.

Kari Fraurud. 1990. Definiteness and the processing of NPs in natural discourse. *Journal of Semantics*, 7:395–433.

John Hawkins. 1994. *A Performance Theory of Order and Constituency*. CUP, Cambridge.

Jerry R. Hobbs. 1976. Pronoun Resolution. Research report 76-1, Department of Computer Sciences, City College, City University of New York, August.

Elsie Kaiser and John Trueswell. submitted. Role of discourse context in the processing of a flexible word-order language.

Frank Keller. 2000. *Gradience in grammar: Experimental and computational aspects of degrees of grammaticality*. Ph.D. thesis, University of Edinburgh.

Gerard Kempen and Karin Harbusch. 2004. How flexible is constituent order in the midfield of German subordinate clauses? A corpus study revealing unexpected rigidity. In *Proceedings of the International Conference on Linguistic Evidence*, pages 81–85, Tübingen, Germany.

Geert-Jan Kruiff and Denys Duchier. 2003. Information structure in Topological Dependency Grammar. In *Proceedings EACL 2003*, Budapest, Hungary.

Ivana Kruiff-Korbayová and Mark Steedman. 2003. Discourse and information structure. *Journal of Logic, Language and Information*, 12:249–259.

Daniela Kurz. 2000. A statistical account on word order variation in German. In *Linguistically Annotated Corpora LINC-2000, Workshop at COLING*, Luxembourg.

Knud Lambrecht. 1994. *Information structure and sentence form*. CUP, Cambridge.

Thomas S. Morton. 2000. Coreference for NLP Applications. In *Proceedings of ACL 2000*.

Gereon Müller. 1999. Optimality, markedness, and word order in German. *Linguistics*, 37:777–818.

Thomas Pechmann, Hans Uszkoreit, Johannes Engelkamp, and Dieter Zerbst. 1996. Wortstellung im deutschen Mittelfeld. Linguistische Theorie und psycholinguistische Evidenz. In C. Habel, S. Kanngiesser, and G. Rickheit, editors, *Perspektiven der Kognitiven Linguistik. Modelle und Methoden*, pages 257–299. Westdeutscher Verlag, Opladen.

Ellen Prince. 1981. Toward a taxonomy of given-new information. In P. Cole, editor, *Radical Pragmatics*, pages 223–255. Academic Press, New York, NY.

Ellen Prince. 1992. The ZPG letter: subjects, definiteness, and information status. In S. Thompson and W. Mann, editors, *Discourse description: Diverse analyses of a fund-raising text*, pages 295–325. John Benjamins.

Douglas Rhode, 2002. *Tgrep2 User Manual. Version 1.06*.

Brigitte Röder, Tobias Schicke, Oliver Stock, Gwen Heberer, Helen Neville, and Frank Rösler. 2000. Word order effects in German sentences and German pseudo-word sentences. *Sprache und Kognition*, 19(1/2):31–37.

Elisabeth Selkirk. 1984. *Phonology and Syntax*. MIT Press, Cambridge, MA.

Wojciech Skut, Thorsten Brants, Brigitte Krenn, and Hans Uszkoreit. 1998. A linguistically interpreted corpus of German newspaper text. In *Proceedings of the ESSLLI Workshop on Recent Advances in Corpus Annotation*, pages 18–24, Saarbrücken, Germany.

Mark Steedman. 1991. Structure and intonation. *Language*, 67:262–296.

Hans Uszkoreit. 1987. *Word Order and Constituent Structure in German*. Lecture Notes. CSLI, Stanford University.