

# Coreference Preferences for Personal Pronouns in German

**Gerlof Bouma & Holger Hopp**

*University of Groningen, The Netherlands*

---

This paper presents psycholinguistic evidence on the factors governing the resolution of German personal pronouns. To determine the relative influence of linear order versus grammatical function of potential antecedents, two interpretation-preference tasks were designed. Their specific aim was to disentangle salience factors conflated in previous research on pronoun interpretation, such as linear order, first mention and topicalization. Experiment 1 tested pronoun resolution to non-sentence-initial position (scrambling) and Experiment 2 tested pronoun resolution to sentence-initial position (topicalization). The results across different verb types and across different syntactic contexts in Experiments 1 and 2 show that grammatical function, yet neither linear order, first mention nor topicalization predicts pronoun resolution in German.

---

## **1 Introduction**

Language comprehension involves the resolution of ambiguities, for instance, in determining co-reference relations between ambiguous pronouns and their potential antecedents. In order to elucidate the nature of these ambiguity resolution strategies, pronoun interpretation preferences have been investigated with reference to three types of factors: (a) world knowledge such as the plausibility of linking a pronoun to a particular antecedent; (b) linguistic constraints such as agreement or binding; and (c) psychological salience or activation of available referents. With respect to (c), researchers have proposed, for instance, NP-form, distance to the anaphor, position in the sentence and grammatical function as linguistic and (non)linguistic form aspects that influence salience. The relative impact of these factors has been explored cross-linguistically in corpus and computational research as well as psycholinguistic experimentation or introspection. However, especially for languages allowing for free word order, the results have proven inconclusive.

This paper aims to supply new evidence on the relative contribution of grammatical form aspects, namely, grammatical function (GF) and surface form

aspects, namely, linear order (LO) to the resolution of personal pronouns in German. Both GF and LO have been argued to underlie pronoun resolution preferences (e.g. Brennan, Friedman, and Pollard, 1987; Rambow, 1993; Strube and Hahn, 1999). We present two experiments to discriminate between GF and LO preferences in interpretation. These experiments investigate pronoun resolution in two different syntactic contexts in order to establish the scope of interpretive preferences for German personal pronouns.

In particular, we explore the idea that part of the conflicting previous findings on effects of LO and GF may result from methodological confounds of LO with first mention (FM) effects or information-structural effects of topicalization (TOP) in many of the previous studies. Since psycholinguistic studies on pronoun resolution have so far considered only word order variation in main clauses (e.g. SVO versus OVS), effects of LO, FM and TOP are conflated, as the first-mentioned entity in topicalized sentence-initial position is identical to the leftmost possible pronoun antecedent. Consider the German (1), taken from Bosch et al. (2007): The bolded sentence-initial NP in (1b) is topicalized, first-mentioned as well as leftmost.

- (1) a) Der Oberarzt untersucht den Notfallpatienten. (SVO)  
 The senior physician.nom examines the emergency patient.acc  
 The senior physician examinesthe emergency patient.
- b) **Den Notfallpatienten** untersucht der Oberarzt. (OVS)

Both FM and TOP have been advanced as factors affecting salience: First, Gernsbacher (e.g. Gernsbacher & Hargreaves, 1988) argues that the first-mentioned words in a sentence carry the highest salience due to general cognitive principles favouring initial information. Second, the information structure of topicalization has been shown to affect the relative salience of topicalized entities in sentence processing (see, e.g., Weskott, 2003).

In the two experiments in this paper, we first isolate effects of LO by testing LO against GF in non-sentence-initial position (Experiment 1) and, second, we consider potentially additive or interacting effects of LO, FM and TOP by testing LO against GF in sentence-initial topicalized position (Experiment 2).

Counter to previous research on German (e.g. Rambow, 1993; Strube and Hahn, 1999), both experiments document strong effects of GF, and no effects of LO for German. These results resonate with prior findings for other free word order languages, e.g. Finnish (Järvikivi et al., 2005; Kaiser and Trueswell, 2007). Our findings are near-identical for potential pronoun antecedents in sentence-initial and sentence-medial position, which suggests that the preference of personal pronouns for subject antecedents is robust against LO, FM as well as TOP.

This paper is structured as follows: In section 2, we summarize previous corpus and psycholinguistic research on effects of LO and GF in pronoun resolution. Section 3 illustrates word order variation in German and outlines differences between sentence-initial reordering (topicalization) and sentence-medial reordering (scrambling). In section 4, the first interpretation experiment on scrambling is presented and discussed. Section 5 reports the second interpretation experiment on topicalization. In section 6, we discuss the general findings and put them in perspective of previous research.

## 2 Previous research

Effects of grammatical function and surface order on pronoun resolution have been explored in computational and corpus research as well as psycholinguistic studies.

### 2.1 Computational and corpus research

Computational and corpus research on pronoun resolution has been mainly carried out within the framework of Centering Theory (CT), a theory of local discourse coherence (Grosz, Joshi, and Weinstein, 1995). Originally, CT proposed GF as the main determinant of salience: subjects are more salient than objects, which are more salient than other elements in an utterance. Although CT was not intended as a theory of anaphora resolution, it has inspired many such theories and systems. A well-known early example is Brennan, Friedman, and Pollard (1987), who implemented CT for pronoun resolution, and operationalized GF by ranking entities according to a refined obliqueness scale. Pitting GF against LO in a corpus-based computational evaluation, Poesio et al. (2004) report that substituting GF information by LO does not lead to substantially different results. Note, though, that the comparatively fixed word order of English leads to a close correspondence of linear order and grammatical function (subjecthood).

Thus, free word order languages are better suited to provide evidence that differentiates between LO or GF. For free(er)-word order languages like German, Strube & Hahn (1999) claim that information structure as expressed, *inter alii*, in LO of NPs underlies pronoun resolution; in their computational model that has been tested against corpus data, GF plays no role. On the other hand, corpus studies of newspaper texts by Bosch et al. (2003) and Wunsch (2006) identify GF, and specifically subjecthood, as the best predictor of referential antecedence for personal pronouns in German.

## **2.2 Psycholinguistic studies**

Psycholinguistic studies conceptualize effects of GF and LO in terms of the status of grammatical (GF) and non-grammatical (LO) factors guiding the resolution of ambiguities in comprehension.

In the structure-building approach by Gernsbacher (Gernsbacher & Hargreaves, 1988; Gernsbacher, Hargreaves, & Beeman, 1989), LO is implemented as first mention (FM) of an entity in a sentence. According to Gernsbacher, the first-mentioned entity bears higher salience than later-mentioned entities, since it forms the foundation onto which later information in the sentence is mapped. In probe-recognition studies of sentences containing two noun phrases, Gernsbacher and Hargreaves (1988) and Carreiras, Gernsbacher and Villa (1995) report that probe recognition is faster for the first-mentioned noun phrase. Note, though, that neither study used pronoun probes. For pronouns, reading-time experiments by Frederiksen (1981) and Crawley, Stevenson and Kleinman (1990) report faster reading times for sentences beginning with a pronoun referring to the subject, rather than the object, in the preceding sentence. In consequence, these authors argue that grammatical information predominantly underlies strategies for pronoun resolution.

Since subjecthood usually coincides with first position in English, though, these results cannot discriminate between GF and LO or FM. However, Gordon, Grosz, and Gillom (1993) found increased reading times for repeated full NP versus pronoun subjects for both subjects or first mentioned NPs in the preceding sentence. Analogously to subjects, reaction times increased for first mentioned non-subject NPs, embedded in sentence-initial adverbials, such that Gordon et al. (1993) conclude that LO/FM is an (additional) salience factor in English.

Again, potentially more decisive evidence comes from free-word order languages. For Finnish, Kaiser and Trueswell (2007) tested pronoun resolution in relation to previous SVO and OVS sentences in order to establish the relative impacts of GF and surface order, i.e. LO/FM/TOP. In interpretation experiments as well as eye-tracking using the visual-world paradigm, they report that Finnish personal pronouns preferentially refer to subjects of preceding sentences, irrespective of whether these subjects are in sentence-initial pre-verbal position (SVO) or whether they follow the verb in sentences involving topicalization (OVS). Using similar materials in eye-tracking, Järvikivi et al. (2005) find that both subjecthood (GF) and FM/LO/TOP have effects on resolution preferences, although subjecthood seems to be a stronger factor than FM/LO/TOP. For German, however, judgement and eye-tracking studies by Hemforth and Konieczny (2002) find effects of LO, as realized in TOP, for personal pronouns in embedded clauses that refer to antecedents in the matrix clause. In off-line interpreta-

tion and completion tasks, Bosch et al. (2007) report a slight preference for German personal pronouns to refer to antecedent subjects, although this effect does not reach statistical significance. Unfortunately, none of these studies allows us to disentangle effects of LO from those of FM and TOP.

In contrast, Rambow (1993) claims – for the German *Mittelfeld* – that LO determines which NP is most likely to be the antecedent of a pronoun in discourse. To illustrate, Rambow argues that in the *Mittelfeld* – the topological field between the finite verb in second position, and the sentence final verb cluster – the predictive role of LO is borne out, outweighing any potential effects of GF in (2) and (3). (The indicated interpretation preferences are from Rambow, 1993.)

(2) Q: Glauben Sie,  
Do you believe  
dass [eine solche Massnahme] [der russischen Wirtschaft] helfen kann?  
that such a measure.nom the Russian economy.dat help can  
Do you believe that such a measure can help the Russian economy?

A: Nein, sie ist viel zu primitiv.  
No, it (=the measure) is much too primitive.

(3) Q: Glauben Sie,  
Do you believe  
dass [der russischen Wirtschaft] [eine solche Massnahme] helfen kann?  
that the Russian economy.dat such a measure.nom help can  
Do you believe that such a measure can help the Russian economy?

A: Nein, sie ist viel zu primitiv.  
No, it (=Russian economy) is much too primitive.

The grammatical functions, subject and indirect (dative) object, do not change between (2) and (3). However, according to Rambow, the interpretation of the pronoun in these answers is always the leftmost *Mittelfeld* NP of the preceding sentence, which does change between the examples. Importantly, the example from Rambow (1993) isolates LO by focussing on potential antecedents in non-initial positions; it thus abstracts away from effects of FM or TOP.

In sum, the cross-linguistic evidence on pronoun preferences to date remains inconclusive. For German, there is some corpus evidence pointing to GF as the main factor for disambiguating personal pronouns. However, this preference is, if at all, only weakly attested in psycholinguistic judgement and comprehension data. Rather, introspective judgement data attribute a greater role to LO than GF, at least for non-sentence-initial antecedents. Against this background, the present study presents novel data from two interpretation experi-

ments in German to establish whether effects of LO can be documented experimentally.

Across two experiments, we differentiate between GF and LO to identify their individual contributions. Starting from Rambow's (1993) observation on the role of LO in the German *Mittelfeld*, we first examine the effects of GF and LO in non-initial sentence position in Experiment 1. The second experiment probes whether the effects of GF and LO extend to sentence-initial antecedents, where LO coincides with FM and TOP. This way, we attempt to specify the contributions or interactions of these salience factors.

### 3 Two Types of Word Order Variation in German

German is an SOV language with verb-second order in main clauses, but it allows for a comparatively free word order with respect to the arguments. This property makes German suitable for an experiment contrasting GF and LO. German has two types of word order variation that may alternate the order of the arguments, which allows us to separate different aspects of LO.

The first type of word order variation in German is *topicalization*. In topicalization, a constituent is moved to the front of the clause, directly in front of the finite verb. If the fronted constituent is not the subject, the subject will appear postverbally. An example of topicalization was given in (1b). Topicalization of arguments is not an information-structurally uniform phenomenon. The fronted argument may express a topic and hence be out of focus, but in rarer cases it may also be a narrow focus or even the projecting part of a wider focus (Féry, 2007). Hence, topicalization serves multiple information-structural functions in that it allows, *inter alii*, for topic-topicalization and focus-topicalization (Gundel, 1988). Although effects of definiteness and givenness on topicalization can be observed (data in Weber and Muller, 2004), these effects are not categorical; the only near categorical constraint on topicalization is that reduced pronominal objects cannot topicalize (Travis, 1984).

Apart from argument topicalization to the front of main clauses, German allows for sentence-internal argument reordering, in the so called *Mittelfeld*. This type of word order variation is referred to as *scrambling*. Like topicalization, scrambling can lead to objects preceding the subject in terms of word order (see 2 & 3 above for indirect objects). A direct object and indirect objects may in principle also appear in either order.

- (4) a) Gern hat ein junger Geiger die Frauen geküßt.  
Gladly has a young violinist.nom the women.nom/acc kissed  
A young violinist was keen on kissing the women.  
b) Gern hat die Frauen ein junger Geiger geküßt.
- (5) a) Der Mann sandte dem Bauern den Arbeiter.  
The man sent the farmer.dat the worker.acc  
The man sent the worker to the farmer.  
b) Der Mann sandte den Arbeiter dem Bauern.

Scrambling is constrained by various factors, such as definiteness, animacy, and information structure (Müller, 1999). Constituents scramble felicitously only when they constitute given information and hence are defocussed, and scrambled word orders are considerably improved when the rightmost constituent is focussed (Haider and Rosengren, 1998; Lenerz, 1977). In other contexts, scrambling of objects across subjects is distinctly marked. In sum, as opposed to topicalization, scrambling shows uniform behaviour in terms of information structure.

Although both topicalization and scrambling implicate syntactic reordering of subject and object, scrambling neither involves the first position in a sentence, nor does it serve multiple information-structural functions.

In the next two sections, we present two experiments on pronoun resolution in German. The first experiment on scrambling investigates effects of GF and LO without the interference of FM or TOP, and the second experiment on topicalization investigates LO as realized in FM and TOP.

#### **4 Experiment 1: Manipulating Word Order by Mittelfeld Scrambling**

In order to investigate the relative importance of GF and LO in pronoun resolution in German, we designed an interpretation preference task. Subjects had to indicate their preferred co-reference interpretation for a pronoun in a small discourse. In this first experiment, we manipulated word order by varying the order of subject and object, or indirect object and direct object in the Mittelfeld, i.e., using scrambling. Since for scrambling, the constituents in question never appear in sentence-initial position, Experiment 1 tests the role of LO avoiding possible confounds of FM or TOP.

#### 4.1 *Materials, Procedure and Participants*

The task manipulated the linear order and the grammatical function of NPs. The factor LO (left-right) was crossed with GF, defined as obliqueness of NPs (subject > indirect object > direct object). In a non-fully-factorial design, three conditions were constructed according to GF: *sudo* (subject-direct object (6)), *suio* (subject-indirect object (7)), and *iodo* (indirect object-direct object (8)). Example sentences for each condition are given below, with the (a) examples showing ordering by GF and the (b) examples showing syntactic reordering.

(6) *sudo* (a: *su* before *do*. b: *do* before *su*)

- a) Die Hoffnung war, dass [der Beschluss]<sub>SU</sub> [den Plan]<sub>DO</sub> beeinflussen würde.  
 the hope was, that the decision.nom the plan.acc influence would  
*It was hoped that the decision would influence the plan.*
- b) Die Hoffnung war, dass [den Plan]<sub>DO</sub> [der Beschluss]<sub>SU</sub> beeinflussen würde.

(7) *suio* (a: *su* before *io*. b: *io* before *su*)

- a) Alle dachten, dass [der Sohn]<sub>SU</sub> [dem Vater]<sub>IO</sub> ähnelte.  
 Everybody thought, that the son.nom the father.dat resembled  
*Everybody thought that the son resembled the father.*
- b) Alle dachten, dass [dem Vater]<sub>IO</sub> [der Sohn]<sub>SU</sub> ähnelte.

(8) *iodo* (a: *io* before *do*. b: *do* before *io*)

- a) Die Professorin stellte [dem Kollegen]<sub>IO</sub> [den Studenten]<sub>DO</sub> vor.  
 the professor.f introduced the colleague.dat the student.acc vpart  
*The professor introduced the colleague to the student.*
- b) Die Professorin stellte [den Studenten]<sub>DO</sub> [dem Kollegen]<sub>IO</sub> vor.

For each of the 3 conditions, 6 items were devised, yielding a total of 18 items. In the subject conditions (*sudo* and *suio*), the two potential antecedent NPs appeared in a subordinate clause (like in (6) and (7) above) to ensure both arguments' appearances in the Mittelfeld are unmarked. In the *iodo*-condition, the subject – ruled out as antecedent by means of gender agreement – was in sentence-initial position. All potential antecedent NPs were controlled and matched for animacy, definiteness, number and gender. In terms of information structure, both NPs were given since their referents were previously mentioned in an introduction sentence.

The ambiguous pronoun was introduced as the subject of a short third and final sentence after the introduction sentence and the context sentence. Gender and number agreement ensured that only two of the NPs appearing in the mini-discourse were available as antecedents. A complete stimulus, consisting of a

three-sentence discourse and a question with three answer options, is given in (9). The translations in italics were not present in the actual stimuli.

- (9) Das Gremium fasste einen Beschluss, der den Plan zur Umsatzsteigerung ändern sollte.  
*The board made a decision that should change the plan for increasing turnover (introduction sentence)*  
Die Hoffnung war, dass der Beschluss den Plan beeinflussen würde.  
*It was hoped that the decision.masc would affect the plan.masc (canonical word order, context sentence).*  
Aber er war zu unstrukturiert.  
*But it.masc was too unstructured (ambiguous pronoun).*

Was war zu unstrukturiert?

*What was too unstructured?*

- a) Der Beschluss  
*The decision*
- b) Der Plan  
*The plan*
- c) Etwas anderes  
*Something else*

Participants indicated their interpretation preference by answering a question about the discourse, using three fixed answers corresponding to either of the potential antecedent NPs in the context sentence, or something else. To control for plausibility of interpretation and to test for lexical biases in pronoun preferences, all items were tested in a plausibility-rating study with a separate group of 12 German natives. In this plausibility experiment, the discourses from the main experiment were taken, but with the ambiguous pronoun in the third sentence replaced by the two potential antecedent NPs. For the discourse in (9), example continuations are given in (10).

- (10) a) Aber der Plan war zu unstrukturiert.  
*But the plan was too unstructured.*  
b) Aber der Beschluss war zu unstrukturiert.  
*But the decision was too unstructured.*

Participants were asked to indicate on a five point scale whether they considered continuation (a) or continuation (b) to be more plausible. On the basis of average plausibility ratings, 4 of the 18 items were excluded from analysis in the main experiment; 3 from the *sudo* condition and 1 from the *suio* condition. For the

remaining 14 items, order and syntactic function of the NPs constituted the only differences.

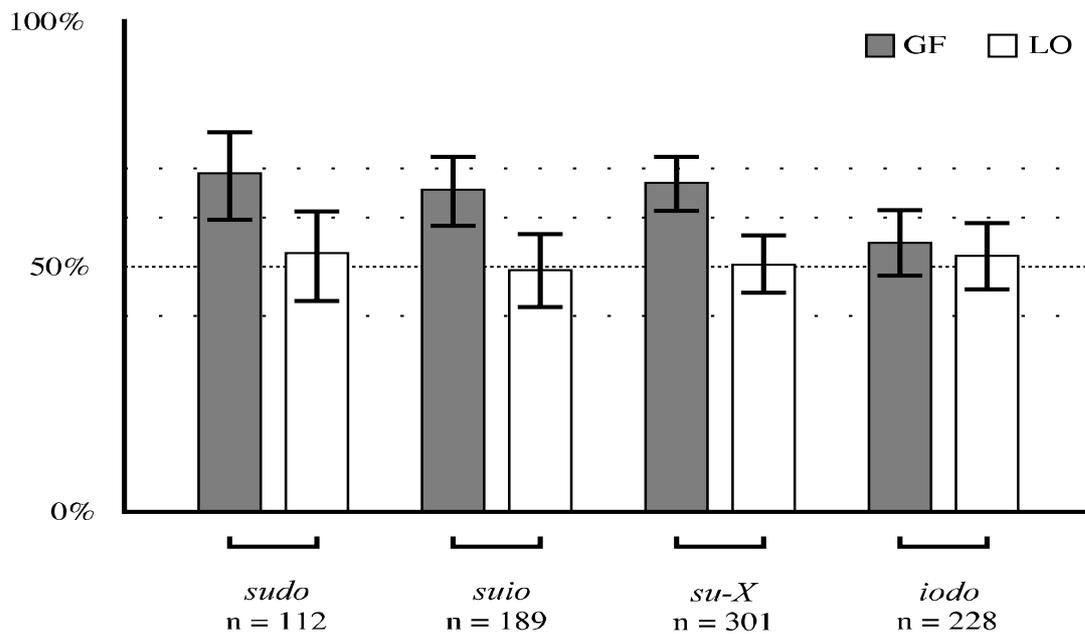
The untimed experiment was conducted online, using a simple web-based interface. Participants were not informed about linguistic concepts like pronoun interpretation, antecedents, etc.; however, they were given examples of how sentences and discourses can be ambiguous, and how, even in cases of ambiguity, it is possible that one reading is preferred over others. Participants were instructed to indicate their preferred interpretation by answering questions about the discourses, as outlined above. They were not able to change their answers once they had moved on to a new item. The 18 items were interspersed with 18 fillers, and distributed over 2 lists that balanced the order of the answers, the order of the items, as well as the order of the arguments in the context sentence. Each participant only saw each item once.

In Experiment 1, 38 adult, native speakers of German, who were recruited individually via email, participated.

## 4.2 Results

The results of Experiment 1 are summarized in Figure 1. The four groups of bars refer, from left to right, to the subject conditions individually (*sudo*, *suio*), the subject conditions collapsed (*su-X*), and the double object condition (*iodo*). The dark bars refer to the proportion of cases in which a subject was picked over an object (*sudo*, *suio*, *su-X*), or the indirect object over the direct object (*iodo*). The light bars refer to the proportion of cases in which the leftmost NP was preferred over the rightmost NP as an antecedent in each condition. The error bars indicate the 95% confidence intervals for the estimated proportions. There were three cases of participants answering 'something else', two in the *sudo*, one in the *suio* condition. These cases were removed from analysis.

In the conditions involving subjects (*sudo* and *suio*), roughly two-thirds of the time people prefer the subject over the object as the antecedent. In each subject condition, and in the collapsed conditions (*su-X*), the subject is chosen significantly more often than chance (*sudo*: 78/112=70%,  $p<.001$ ; *suio*: 124/189=66%,  $p<.001$ ; *su-X*: 202/301=67%,  $p<.001$ , all 2-tailed exact binomial tests, see also the confidence intervals in the chart). In the double object condition, the preference for the indirect objects over direct objects is not statistically significant (125/228=55%,  $p=.164$ , 2-t exact binomial). The results of Experiment 1 indicate that there is a preference for subject antecedents.



**Figure 1:** Results of Experiment 1, manipulating word order by scrambling. Error bars indicate 95% conf. int. for the probability of predictor being correct.

By contrast, there is no evidence in any of the conditions that participants select the leftmost potential antecedent more often than the rightmost (*sudo*: 59/112=53%,  $p=.636$ ; *suio*: 93/189=49%,  $p=.884$ ; *su-X*: 152/301=50%,  $p=.908$ ; *iodo*: 119/228=52%,  $p=.551$ , all 2-t exact binomial tests).

Comparing the preferences for less oblique antecedents and leftmost antecedents directly, we find that participants select the least oblique NP significantly more often than they select the leftmost NP in the conditions *sudo* and *suio* (*sudo*:  $p=.015$ ; *suio*:  $p=.002$ ; *su-X*:  $p<.001$ , all 2-t exact sign tests); however, there is no significant difference between least oblique and leftmost antecedent in the *iodo* condition ( $p=.152$ ). In the *sudo* and *suio* conditions, therefore, obliqueness is a better predictor of pronoun interpretation than linear order.

In summary, the data strongly suggests that LO is not a factor in pronoun interpretation, since there is no preference for the leftmost NP in any condition. By contrast, GF partially is a factor, since there is a strong subject preference.

**Table 1:** Breakdown of preferences in Experiment 1, per word order, per condition. None of the 2x2 sub-tables show significant associations.

Condition	Word order	GF	
		correct	incorrect
sudo	su before do	41	16
	do before su	37	18
suio	su before io	61	33
	io before su	63	32
iodo	io before do	63	47
	do before io	62	56

However, it might be the case that the general pattern we see in Figure 1 is due to a complex interaction between GF and LO. For instance, subjects might only be selected as antecedents when they appear as the leftmost NP. To investigate this possibility, we break down the results by word order in Table 1.

Table 1 shows that the subject preference does not interact with reordering, i.e. even when the more oblique noun phrase shifts leftwards, anaphoric preferences do not shift leftwards. Instead, anaphor resolution preferences remain oriented to the subject in the *sudo* and *suio* conditions; in the do-io condition, the preferences remain indeterminate. This further demonstrates that linear order does not underlie pronoun resolution in relation to NPs in the German Mittelfeld. None of the three 2x2 sub-tables show signs of association between word order and GF performance (all  $ps > 0.5$ , 2-t Fisher's Tests).

### 4.3 Discussion

The results of Experiment 1 show a clear effect of subjecthood on antecedent choice and, as such, point to the role of GF for personal pronoun resolution in German. At the same time, the results suggest that LO plays no role in pronoun resolution. With respect to subjecthood, our results are in line with Järvi-kivi et al. (2005), Kaiser (2003) for Finnish. However, since there were no effects of GF for the *iodo* condition, the data in this experiment do not support distinguishing further obliqueness levels, that is, distinguishing between indirect object and object (contra Brennan et al. 1987, for English).<sup>1</sup>

By testing the influence of LO of arguments in the German Mittelfeld, we avoided the possible confound of a first mention (FM) effect, and a possibly related (but not well understood) information-structural effect of topicalization

<sup>1</sup> See the discussion of the results of Experiment 2 for a more detailed evaluation of the lack of an obliqueness effect in the double object data.

(TOP). Conversely, this also means that the lack of an LO effect in Experiment 1 does not rule out the existence of an FM effect, nor does it rule out a TOP effect., since scrambling and topicalization might have different influences on the salience of discourse referents. To cast light on these issues, Experiment 2 probes word order variation targeting the sentence-initial position.

## 5 Experiment 2: Manipulating Word Order by Topicalization

### 5.1 Materials and Procedure

In order to ensure the greatest degree of similarity between the two experiments for direct comparisons, we used the materials of Experiment 1, adapting them for topicalization. Apart from the syntax of the context sentence, all materials were identical to those for Experiment 1.

In Experiment 2, all potential pronoun antecedents were direct constituents of the main clause, so that topicalization was possible. In some cases auxiliary verbs and adverbial material were added to make the sentence more natural and to substitute for the meaning contribution of the deleted matrix clause material (e.g. ‘He thinks ...’) in the *sudo* and *suio* conditions of Experiment 1. In the double object conditions in Experiment 2, the subject (not a potential antecedent) always directly follows the verb, so that either of the objects appears in topicalized position. The counterparts to (6)–(8) in Experiment 2 are given in (11)–(13). Note that there is always one potential antecedent in sentence-initial position.

(11) *sudo* (a: *su* before *do*. b: *do* before *su*)

- a) [Der Beschluss]<sub>SU</sub> sollte [den Plan]<sub>DO</sub> beeinflussen.  
 the decision.nom should the plan.acc influence  
 The decision was intended to influence the plan.
- b) [Den Plan]<sub>DO</sub> sollte [der Beschluss]<sub>SU</sub> beeinflussen.

(12) *suio* (a: *su* before *io*. b: *io* before *su*)

- a) [Der Sohn]<sub>SU</sub> ähnelte [dem Vater]<sub>IO</sub> ein wenig.  
 The son resembled the father a little.  
 the son.nom resembled the father.dat a little
- b) [Dem Vater]<sub>IO</sub> ähnelte [der Sohn]<sub>SU</sub> ein wenig.

(13) *iodo* (a: *io* before *do*. b: *do* before *io*)

- a) [Dem Kollegen]<sub>IO</sub> stellte die Professorin [den Studenten]<sub>DO</sub> vor.  
 the colleague.dat introduced the professor.f the student.acc vpart  
 The professor introduced the student to the colleague.
- b) [Den Studenten]<sub>DO</sub> stellte die Professorin [dem Kollegen]<sub>IO</sub> vor.

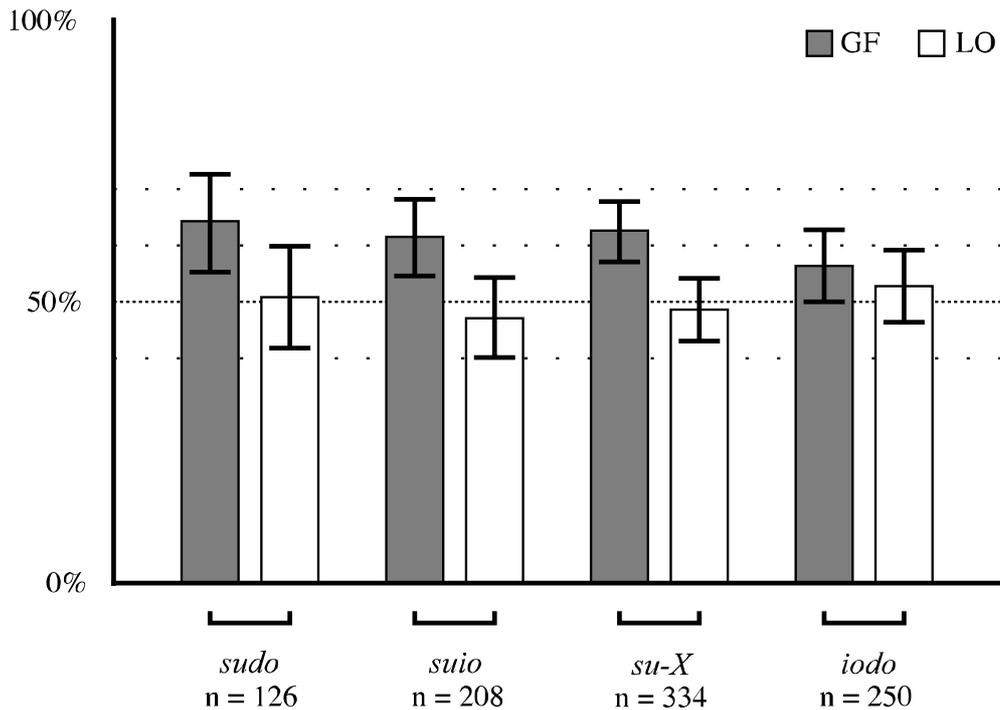
The 18 items were again distributed over 2 balanced lists with fillers; recruiting and procedure were identical to Experiment 1. Forty-two adult native speakers of German participated in Experiment 2.

## 5.2 Results

The results of the second interpretation experiment are summarized in Figure 2. Four datapoints were excluded from analysis, because 'something else' had been answered (two in *suio*, two in *iodo*).

In all conditions, participants showed a significant preference for the least oblique possible NP over the more oblique NP, although this effect only just reaches significance in the double object condition (*sudo*: 81/126=64%,  $p=.002$ ; *suio*: 128/208=62%,  $p=.001$ ; *su-X*: 209/334=63%,  $p<.001$ ; *iodo*: 141/250=56%,  $p=.049$ , all 2-t exact binomial tests). There is no evidence for a preference for the leftmost NP over the rightmost NP in any condition (*sudo*: 64/124=51%,  $p=.788$ ; *suio*: 98/208=47%,  $p=.446$ ; *su-X*: 162/334=49%,  $p=.623$ ; *iodo*: 132/250=53%,  $p=.411$ , all 2-t exact binomial tests). Finally, the preference for the least oblique possible NP is significantly stronger than the preference for the leftmost possible NP in all conditions (*sudo*:  $p=.043$ ; *suio*:  $p=.004$ ; *su-X*:  $p<.001$ ; *iodo*:  $p=.049$ , all 2-t exact sign tests). The results of Experiment 2 suggest that GF is a predictor of pronoun interpretation throughout, although the effect of subjecthood is the most robust. LO does not predict pronoun resolution.

As before, we can inspect the preferences for GF in comparison to word order, in order to see whether there is an interaction between preferences and order (Table 2). Again, there is no indication that the preference for GF is modulated by word order (Fisher's Tests for each 2x2, for all conditions: all  $p>.4$ ). In other words, the null effect of LO remains even when the more oblique NP is leftmost.



**Figure 2:** Results of Experiment 2, manipulating word order by topicalization

**Table 2:** Breakdown of preferences in Experiment 2, per word order, per condition. None of the 2x2 sub-tables show significant associations.

Condition	Word order	GF	
		correct	incorrect
<i>sudo</i>	su before do	41	22
	do before su	40	23
<i>suio</i>	su before io	61	43
	io before su	67	37
<i>iodo</i>	io before do	74	51
	do before io	67	58

### 5.3 Discussion

The results of Experiment 2 are virtually identical to the results of Experiment 1: There is a robust GF effect in the conditions that involve a possible subject antecedent, and there is no evidence for an LO effect in any of the conditions. Since word order variation in Experiment 2 targeted the sentence-initial position, we can conclude that there is no first mention effect (FM) on pronoun resolution in German. In the same vein, we conclude that the particular information-structural functions of topicalization (TOP) do not affect pronoun resolution preferences, at least when compared to scrambling.

At this stage we only have enough evidence to claim subjects are clearly more likely to be selected as antecedent, partially confirming GF. Our experiments therefore do not support discerning further levels of obliqueness, as proposed in Brennan et al (1987), for English. However, the fact that GF was a marginally significant predictor in the double object condition in Experiment 2, warrants further investigation of this data group. Note that the direction and size of the preferences in both experiments are very similar: in 55% respectively 56% of the cases, participants preferred the indirect object over the direct objects. This is in agreement with our formulation of GF: Subjects are preferred over indirect objects, which, in turn, are preferred over direct objects.

However, using such a scale of obliqueness presupposes that all verbs behave similarly with respect to this obliqueness hierarchy. In particular, it means assuming that the verbs used in the double object stimuli are a homogenous group. This need not be the case. For German, Haider (1993) points out that some verbs go with an unmarked word order of indirect object before direct object (in accordance with our obliqueness hierarchy), whereas with other verbs the preferred unmarked order seems to be direct object before indirect object. Of the six verbs in the double object condition, three display a preference for a direct object initial word order (*verschweigen* ‘not tell about’, *verbergen* ‘hide’ and *überweisen* ‘refer’). There is no consensus as to whether these word order preferences correspond to differences in the underlying syntactic structure (see, for instance, Müller, 1999, for an answer in the negative), that is, whether the objects of these verbs have different obliqueness hierarchies. Irrespective of the question of underlying order, it may still be the case that the preferences of the three verbs imply higher salience of the direct object. In this case, the lack of a clear GF effect in the double object condition may be attributed to the fact that there are two different verb classes showing dichotomous preferences.

Inspection of the data of Experiments 1 and 2 does not reveal a dichotomy between the three verbs mentioned and the remaining three verbs in terms of the preference for the indirect object NP as antecedent. However, the data are far from homogenous: three verbs show a consistent preference for the indirect object as antecedent (*empfehlen* ‘recommend’, *verschweigen* ‘not tell about’, *verbergen* ‘hide’), one verb does not show a clear trend (*überweisen* ‘refer’), and two verbs show a clear preference for the direct object (*vorstellen* ‘introduce’, *präsentieren* ‘present’). The double object data excluding the last two verbs show, in both experiments, an overall preference for the indirect object of a size similar to the subject preference in the other conditions (Experiment 1: 66%, Experiment 2: 70%). The two deviant verbs prefer the direct object in 67% and 73% of the cases respectively. Note that this pattern does not align with differences in markedness that potentially index different underlying argument orders; rather, the pattern cross-cuts these differences. At this point, we have no expla-

nation for this pattern in the data, although we would like to point out that it is possible that verb semantics play a role, given the fact that the two deviant verbs are near-synonyms. Because our plausibility study controlled for pragmatic or lexical plausibility of pronoun preferences, we deem it unlikely that plausibility alone could be responsible for the observed difference between verbs. It will be interesting to consider different verb types and different verb classes more systematically in future research in order to investigate effects of verb semantics and argument structure on pronoun preferences.

To summarize, we find a robust subject effect in both experiments, such that GF is partly confirmed. In conjunction, Experiments 1 and 2 document that LO, irrespective of whether conceptualized as LO, FM or TOP, does not have an effect on pronoun interpretation in German. Since our study involved off-line interpretation preferences, our results directly contradict the claims by Rambow (1993) and Strube and Hahn (1999) suggesting LO should be one of the factors (possibly in conjunction with other factors) when assigning an interpretation to a pronoun.

## **6 Overall Discussion and Directions for Future Work**

The results from Experiments 1 and 2 across different types of NP arguments in the German sentence-initial and sentence-medial position indicate that GF, or, more specifically, subjecthood is a predictor of pronoun resolution. Although referentially fully ambiguous, personal pronouns refer at above-chance levels to the subject of a preceding sentence that offered multiple potential antecedents. This result generalizes for subjects over sentence type, i.e. main (Experiment 2) and embedded clauses (Experiment 1) and over syntactic order, i.e. SO and OS (Experiments 1 & 2). The finding that subjects are chosen as antecedents roughly two-thirds of the time shows that there is a robust resolution preference that nevertheless is not categorical. However, GF turned out not to have a general effect on pronoun resolution for object-object ambiguities in either Experiment 1 or Experiment 2.

Counter to the claims by Strube and Hahn (1999) and Rambow (1993), LO was not found to determine pronoun resolution in either Experiment 1 or 2. Moreover, counter to the claims of the structure-building approach by Gernsbacher (e.g. Gernsbacher and Hargreaves, 1988), FM does not predict pronoun resolution preferences in German (Experiment 2). In sum, LO, whether conceptualized as surface order, first mention or topicalization, was not found to play any role in pronoun interpretation preferences.

Of course, we need to exclude alternative explanations of the null effect of LO and the positive effect of GF in Experiments 1 & 2 that do not attribute the pattern of results to the involvement of grammatical-function-based strategies in

pronoun resolution. To this end, we consider account in terms of parallelism and a frequency explanation, in turn.

One alternative explanation of the robust preference for subject NPs, and the lack of such a clear GF preference in the double object data, might be given in terms of *parallelism*. According to the parallelism account (e.g. Smyth 1994), a pronoun is preferentially interpreted as co-referential with a noun phrase that has the same grammatical role. The ambiguous pronoun in our experiments was invariably a subject. If readers show a preference for an antecedent that has the same grammatical function as the pronoun that is resolved, one would expect a preference for subject antecedents, and one would not expect to see any preference in those cases where the subject is not available. Thus, parallelism would explain both the positive finding for GF and the null finding with respect to LO.

Although stronger conclusions about the plausibility of this alternative hypothesis ultimately have to await a study in which the grammatical function of the pronoun is varied, the results of the two present experiments suggest that parallelism of pronoun and antecedent does not play a significant role in the data. Smyth (1994) argues that there are several aspects to parallelism of which grammatical function is only one. One of the other aspects is structural position of the constituent. The effect of parallelism on pronoun resolution is strongest when more of these different aspects apply. This means that parallelism would predict a linear order effect in Experiment 2. To see how, consider the configuration of the sentence containing the pronoun compared to different word orders in the context sentence. The ambiguous pronoun was immediately preverbal (SV) in all but one of the stimuli. In the preceding context sentence in Experiment 2, the subject is either in preverbal (SVO) or postverbal (OVS) position. This means that there is a higher degree of parallelism between the pronoun sentence and the context sentence when the subject in the latter is preverbal (SVO – SV) compared to when it is postverbal (OVS – SV). According to parallelism, then, preverbal subjects should be selected as antecedents more often than postverbal subjects. As Table 2 shows, this prediction is not borne out: Subjects are preferred, no matter whether they occur preverbally or postverbally. By the same reasoning, parallelism would predict a difference between the SVO cases in Experiment 2, and both of VSO and VOS in Experiment 1. Comparing Tables 1 and 2 shows that this is not borne out either. We therefore conclude that parallelism is not a likely explanation of the results.

A second alternative explanation of the findings might be that *frequency* differences between the word order variants interfere with pronoun interpretation. Non-canonical syntactic orders, such as scrambling and topicalization, are marked and infrequent in German (e.g. Hoberg 1981). Moreover, they are restricted to specific discourse contexts. In particular scrambling of objects across subjects is a marked and infrequent reordering option in German. Since unambi-

guous effects of LO in the present study can surface only if pronouns are resolved to initial objects in scrambled or topicalized position, it could be that a frequency bias against non-canonical word orders, such as scrambling and topicalization, mitigates against or even masks order effects.

However, there is some evidence both across Experiments 1 and 2 and within each experiment that frequency does not affect the pattern of linear order preferences in pronoun resolution. First, topicalization is far more frequent a non-canonical word order than scrambling in German. If relative frequency affected order effects in pronoun resolution, some difference between Experiment 1 and 2 would be expected, contrary to fact.

Second, frequency differences within the experimental conditions in Experiment 1 and 2 would lead one to expect differences in interpretive patterns. Consider scrambling in this respect: In a corpus study on NP order using the NEGRA II corpus that consists of about 20.000 written sentences, Kempen and Harbusch (2003) report that there are large frequency differences between accusative-initial and dative-initial orders compared to nominative-initial orders of full NPs in the German Mittelfeld. Compared to 513 nominative-accusative (i.e. *su* before *do*) orders, there is only one case of an accusative-nominative (*do* before *su*) order, for dative-marked indirect objects, there are 20 cases of dative-nominative (*io* before *su*) orders compared to 43 nominative-dative (*su* before *io*) orders. This corresponds to a ratio of roughly 1 to 500 for *do*-before-*su* orders and roughly 1 to 2 for *io*-before-*su* orders. If frequency differences of scrambling affected linear order effects in pronoun resolution, we would thus expect to see some difference in anaphoric preferences between the *sud* and the *suio* conditions reflecting the frequency divide. Yet, the figures in Figures 1 & 2 demonstrate that there is no such difference between conditions. Moreover, Kempen and Harbusch (2003) note that, for ditransitive verbs, *io*-before-*do* orders by far outnumber *do*-before-*io* orders in the corpus (14 to 3). However, anaphoric preferences do not shift depending on NP ordering in the *do*-*io* condition (see Tables 1 & 2). It thus seems unlikely that frequency effects can account for the null effect of LO in Experiments 1 and 2. Having discussed and dismissed non-grammatical-function accounts of the null effect of LO and the robust subjecthood preference, we conclude that grammatical function indeed underlies pronoun resolution in German. With respect to the role of subjecthood, these results are in line with the findings for personal pronouns in Finnish by Järvikivi et al. (2005) and Kaiser and Trueswell (2007). For German, the grammatical *subject* preference for *personal* pronouns found here complements the grammatical *object* preference for *demonstrative* pronouns elicited by Bosch et al. (2007). In conjunction, these findings document that grammatical function underlies German pronoun resolution; crucially, though, grammatical function interacts with grammatical form (personal versus demonstrative pronouns) in

determining pronoun interpretation preferences (see also Kaiser 2003; Kaiser and Trueswell 2007).

Given the distributional overlap of FM and TOP in German, this study could not dissociate effects of FM and TOP on pronoun resolution. It thus cannot be excluded that the (diverse) information-structural effects associated with topicalization somehow cancel FM effects in Experiment 2. An across-study comparison with the Finnish study by Kaiser and Trueswell (2007), however, suggests that this is unlikely. In Finnish, OVS orders in main clauses are canonically associated with old/given information and are thus obligatorily defocussed (e.g. Kaiser and Trueswell 2004). In other words, OVS in Finnish main clauses has uniform information structure. Yet, similarly to topicalized OVS orders in German in Experiment 2, a sentence-completion and an interpretation preference study for Finnish reported in Kaiser and Trueswell (2007) find that personal pronouns preferentially resolve to the subject in Finnish OVS sentences. To the extent the null findings across experiments in different studies allow for comparisons, we speculate, *pace* Hemforth and Konieczny (2002), that whatever the (information-structural) contribution of topicalization to reordering is, it does not seem to affect pronoun resolution preferences.

Notwithstanding this, it would be rewarding to further investigate the effects of Information Structure on pronoun resolution preferences. For instance, scrambling in German is felicitous only in particular discourse contexts, namely those in which the scrambled constituent denotes given information. This requirement on scrambling was met in the present study in that all relevant NPs were given in preceding discourse contexts as in previous studies (Scheepers, Hemforth, and Konieczny, 2000). This way, information-structural differences between these NPs were neutralized, so that the effects of word order could be isolated. The prototypical case of Mittelfeld scrambling, however, is arguably when a given object NP fronts across an information-structurally new (and focussed) constituent (Lenerz, 1977; Müller, 1999). In future research, it would be interesting to vary the information-structural contexts for ambiguous pronouns systematically to test for potential interactions between word order and information structure in anaphor resolution. A complicating aspect of such a setup would be, however, that one has to make sure that one measures the result of word order variation on pronoun interpretation, rather than a direct effect of the particular context that is needed to facilitate, e.g., scrambling.

In conclusion, we look to future research that considers how potential effects of information structure and argument structure interact with the strong effect of grammatical function for pronoun resolution attested in the present experiments.

## 7 Acknowledgements

The authors wish to thank the audience at the Conference on Intersentential Pronominal Reference in Child and Adult Language in Berlin for helpful comments. The results of a pilot version of Experiment 1 have been reported in Bouma and Hopp (2006). The first author gratefully acknowledges financial support from the Netherlands Organization for Scientific Research (NWO), grant number 051.02.071, as part of the NWO Cognition Programme. The second author now works in the Department of English Linguistics at the University of Mannheim, Germany.

## 8 References

- Bosch, P., G. Katz and C. Umbach. (2007). The non-subject bias of German pronouns. To appear in M. Schwarz-Friesel, M. Consten and M. Knees (eds.): *Anaphors in Texts*. Amsterdam: Benjamins.
- Bouma, G. and H. Hopp. (2006). Effects of word order and grammatical function on pronoun resolution in German. In *Ambiguity in Anaphora*. Workshop Proceedings, ESLLI 2006, Málaga, 5–13.
- Brennan, S.E., M.W. Friedman, and C. J. Pollard. (1987). A centering approach to pronouns. In *Proceedings of the 25th Annual Meeting of the ACL*, 155–162.
- Carreiras, M., M.A. Gernsbacher, and V. Villa. (1995). The advantage of first mention in Spanish. *Psychonomic Bulletin & Review* 1, 124–129.
- Crawley R., R. Stevenson and D. Kleinman. (1990). The use of heuristic strategies in the interpretation of pronouns. *Journal of Psycholinguistic Research* 4, 245–264.
- Frederiksen, J. (1981). Understanding anaphora: Rules used by readers in assigning pronominal referents. *Discourse Processes* 4, 323–347.
- Féry, C. (2007). Wide focus object fronting. To appear in S. Ishihara, M. Schmitz, and A. Schwarz (eds.): *Interdisciplinary Studies on Information Structure* 8, Potsdam.
- Gernsbacher, M.A., and D.J. Hargreaves. (1988). Accessing sentence participants: The advantage of first mention. *Journal of Memory and Language* 27, 699–717.
- Gernsbacher, M.A., D.J. Hargreaves, and M. Beeman. (1989). Building and accessing clausal representations: The advantage of first mention versus the advantage of clause recency. *Journal of Memory and Language* 28, 735–755.
- Gordon, P., B. Grosz, and L. Gillom. (1993). Pronouns, names, and the centering of attention in discourse. *Cognitive Science* 17(3), 311–347.
- Grosz, B., A. Joshi, and S. Weinstein. (1995). Centering: A framework for modelling the local coherence of discourse. *Computational Linguistics* 21(2), 203–225.
- Gundel, J. (1988). The Role of Topic and Comment in Linguistic Theory. In G. Hankamer (ed.): *Outstanding Dissertations in Linguistics*. Garland Publishing Co., New York.
- Haider, H. (1993). *Deutsche Syntax, generativ*. Günter Narr, Tübingen.
- Haider, H. and I. Rosengren. (1998). Scrambling. *Sprache und Grammatik*. Volume 49. Lund University.
- Hemforth, B. and L. Konieczny. (2002). Where pronouns and relative clauses differ: Information Structure and binding preferences. Poster presented at CUNY.

- Hoberg, U. (1981). *Die Wortstellung in der geschriebenen deutschen Gegenwartssprache*. München: Hueber.
- Järvikivi, J., R. van Gompel, J. Hyönä, and R. Bertram. (2005). Ambiguous pronoun resolution: Contrasting the first-mention and subject-preference accounts. *Psychological Science* 16(4).
- Kaiser, E. (2003). Word order, grammatical function, and referential form: On the patterns of anaphoric reference in Finnish. *Nordlyd* 31(1), 245–260.
- Kaiser, E. and J.C. Trueswell. (2004). The role of discourse context in the processing of a flexible word-order language. *Cognition* 94, 113–147.
- Kaiser, E. and J.C. Trueswell. (2007). Investigating the interpretation of pronouns and demonstratives in Finnish: Going beyond salience. To appear in E. Gibson and N. Pearlmuter (eds.): *The Processing and Acquisition of Reference*. Cambridge, Mass.: MIT Press.
- Kempen, G. and K. Harbusch. (2003). An artificial opposition between grammaticality and frequency: comment on Bornkessel, Schlesewsky & Friederici (2002). *Cognition* 90, 205–210.
- Lenerz, J. (1977). *Zur Abfolge nominaler Satzglieder im Deutschen*. Günter Narr, Tübingen.
- Müller, G. (1999). Optimality, markedness and word order in German. *Linguistics* 37, 777–818.
- Poesio, M., R. Stevenson, B. di Eugenio, and J. Hitzeman. (2004). Centering: A parametric theory and its instantiations. *Computational Linguistics* 30(3).
- Prince, E. (1981). Toward a taxonomy of given-new information. In P. Cole, (ed.): *Radical Pragmatics*. Academic Press, New York, 223–256.
- Rambow, O. (1993). Pragmatic aspects of scrambling and topicalization in German: A Centering approach. MS University of Pennsylvania.
- Scheepers, C., B. Hemfort and L. Konieczny. (2000). Linking syntactic functions with thematic roles: psych-verbs and the resolution of subject-object ambiguity. In B. Hemfort and L. Konieczny (eds.): *German Sentence Processing*. Kluwer, Dordrecht, 95–135.
- Smyth, R. (1994). Grammatical determinants of ambiguous pronoun resolution. *Journal of Psycholinguistic Research* 23, 197–229.
- Strube, M. and U. Hahn. (1999). Functional Centering. Grounding referential coherence in information structure. *Computational Linguistics* 25(3), 309–344.
- Travis, L. (1984). *Parameters and Effects of Word Order Variation*. Ph.D. thesis, MIT.
- Turan, Ü.D. (1998). Ranking forward-looking center in Turkish: Universal and language-specific properties. In A. Joshi, M. Walker and E. Prince (eds.): *Centering Theory in Discourse*. Clarendon University Press, Oxford, pages 139–161.
- Weber, A. and K. Müller. (2004). Word order variation in German main clauses: A corpus analysis. In *Proceedings of the 20th International Conference on Computational Linguistics*, Geneva, 71–77.
- Weskott, T. (2003). *Information Structure as a Processing Guide: The Left Periphery of German Verb-Second Sentences and its Interpretation in Context*. PhD dissertation. University of Leipzig.
- Wunsch, H. (2006). Anaphora resolution – What helps in German. In *Pre-proceedings of the International Conference on Linguistic Evidence 2006*, Tübingen.