

Effects of Word Order and Grammatical Function on Pronoun Resolution in German

Gerlof Bouma and Holger Hopp
Center for Language and Cognition (CLCG),
Rijksuniversiteit Groningen
gerlof [at] let.rug.nl, h.c.hopp [at] rug.nl

Abstract

This paper presents novel psycholinguistic evidence on the factors governing pronoun resolution in German. To determine the relative influence of linear order versus grammatical function of potential referents in the German *Mittelfeld* on pronoun interpretation, an interpretation-preference task was run on 23 German natives. Subjects chose the preferred referent for fully ambiguous personal pronouns. The results across different verb types show that grammatical function, yet not linear order, predicts pronoun resolution in German.

Introduction

Anaphors constitute a prime example of ambiguity in language. Many factors have been proposed that influence the interpretation of anaphora, and pronouns in particular. We can distinguish three general types of factors: (a) world knowledge considerations such as the plausibility of the resulting reading; (b) linguistic constraints like agreement or binding; and (c) salience or activation of available referents. With respect to (c), researchers have proposed NP-form, distance to the anaphor, position in the sentence and grammatical function as form aspects that influence salience in different languages. One particular issue that has received attention is the relative impact of linear order (LO) and grammatical function (GF).

For instance, Rambow (1993) claims that in the German *Mittelfeld* – the topological field between the finite verb in second position, and the sentence final verb cluster – linear order determines which NP is most likely to be the antecedent of a pronoun in subsequent discourse. In particular, Rambow claims that in (1) and (2) the predictive role of LO is borne out, outweighing any potential effects of GF:

- (1) Q Glauben sie, dass...
 Do you think that...
 eine solche Maßnahme₁ der russischen Wirtschaft₂ helfen kann?
 a such measure.F.NOM the Russian economy.F.DAT help can
 such a measure could help the Russian economy?
 A Nein, sie₁ ist viel zu primitiv.
 no it.F is much too primitive
 No, the measure is much too primitive.
- (2) Q Glauben sie, dass der russischen Wirtschaft₂ eine solche Maßnahme₁ helfen kann?
 A Nein, sie₂ ist viel zu primitiv.
 No, the economy is much too primitive.

The grammatical functions, subject and indirect (dative) object, do not change between (1) and (2). 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.

Note that the pronoun *sie* in (1) and (2) remains ambiguous: interpretation sie_2 is readily available in (1) as is sie_1 in (2). In other words, preferences for either interpretation are not categorical; rather, they reflect tendencies potentially based on factors like GF and LO. It is notoriously difficult to capture these tendencies using isolated examples. Yet, previous experimental and corpus research has found statistical evidence of the reality of such factors.

This paper adds to this body of research by presenting new empirical evidence on the relative influences of LO and GF on pronoun resolution in relation to the German Mittelfeld. We carried out an interpretation-preference task to investigate the relative impact of either factor. This paper is structured as follows: We briefly review some previous corpus and experimental research on pronoun resolution preferences across languages (Section 1). Then we present the experiment and its findings (Section 2). In Section 3, we discuss the results and offer suggestions for future research.

1 Previous results on LO vs GF

Much work on salience is related to, or based on Centering Theory, a theory of local discourse coherence (Grosz, Joshi, and Weinstein, 1995, CT). 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, BFP), who implemented CT for pronoun resolution, and operationalized GF by ranking entities according to a refined obliqueness scale.

Gordon, Grosz, and Gillom (1993) found increased reading times for repeated full NP versus pronoun subjects if the referent was the subject or the first mentioned NP of the previous utterance. They conclude that LO is a salience factor in English. Moreover, they speculate that any GF effect might be due to subjects often being topics.

Rambow (1993) claims LO determines salience in the German Mittelfeld. Also arguing for the importance of LO, Strube and Hahn (1999) propose a reformulation of Centering Theory based on Prince's familiarity categorization of entities (1981). The system does not use GF information. Empirically, their model outperforms BFP, both on a German and an English corpus.

Of course, the comparatively fixed word order of English leads to a close correspondence of linear order and grammatical function. In a corpus based comparison of CT variants, Poesio et al. (2004) find that substituting GF information by LO has hardly any effect. Thus, free word order languages are better suited to provide evidence that differentiates between LO or GF. Contrastive research on pronoun resolution in such languages seems to point at GF as a determinant of pronoun resolution, or at best to a mixed GF & LO model.

Järvikivi et al. (2005) present eye tracking data for Finnish. Both subjecthood (GF) and first mention (LO) influence the order and number of fixations on a picture of the stories being read. However, the data suggest that subjecthood is the stronger factor. In an interpretation study, Kaiser (2003) concludes that Finnish personal pronouns refer to the subjects, initial or not, unless the object was both initial and pronominal. For Turkish, Hoffman (1998) and Turan (1998) claim subjecthood rather than LO determines salience, based on the introspective

interpretation and acceptability of null pronouns. For Hindi, Prasad and Strube (2000) argue for GF and against LO on the basis of individual examples.

Against the background of these inconclusive and conflicting findings, this study presents psycholinguistic evidence on pronoun resolution in relation to the German Mittelfeld. By using scrambling in the Mittelfeld rather than movement to sentence initial position, this study avoids potential confounds of first mention effects and/or the particular information-structural effects of topicalization on pronoun resolution.

2 The Present Study

German, an SOV language with verb-second order in main clauses, allows comparatively free NP order. Apart from argument *topicalization* to the front of the main clause, German allows for subject-object and object-subject orders in the Mittelfeld (e.g. (1) and (2) above), so-called *scrambling*. Scrambling is constrained by various factors, such as definiteness, animacy, and information structure (Müller, 1999). Constituents scramble felicitously only when they are defocussed, i.e. when they constitute given information (Haider and Rosengren, 1998; Lenerz, 1977). In other contexts, scrambling of objects across subjects is distinctly marked.

2.1 Materials

In order to test whether GF or LO in the German Mittelfeld determine pronoun resolution, we designed an interpretation-preference task. In the task, subjects indicated the preferred referent of a pronoun in relation to a preceding sentence. For this sentence, the task manipulated the linear order and the grammatical function of NPs. The factor LO (left-right) was crossed with GF in three conditions: subject-direct object (3), subject-indirect object (4), and direct object-indirect object (5). Note that none of the relevant NPs is first mentioned.

- (3) subject-direct object (su-do):
- Die Hoffnung war, dass [der Beschluss]_{su} [den Plan]_{do} beeinflussen würde.
The hope was that the decision.NOM the plan.ACC affect would
 - Die Hoffnung war, dass [den Plan]_{do} [der Beschluss]_{su} beeinflussen würde.
- (4) subject-indirect object (su-io):
- Alle dachten, dass [der Sohn]_{su} [dem Vater]_{io} ähnelte.
All thought that the son.NOM the father.DAT resembled
 - Alle dachten, dass [dem Vater]_{io} [der Sohn]_{su} ähnelte.
- (5) direct object-indirect object (do-io):
- Die Professorin stellte [dem Kollegen]_{io} [den Studenten]_{do} vor.
The professor introduced the colleague.DAT the student.ACC part.
 - Die Professorin stellte [den Studenten]_{do} [dem Kollegen]_{io} vor.

There were six items per condition, 18 items in total. All verbs were of the canonical su-obj order; the ditransitives were verbs of the su-io-do class (Haider, 1993). In the design of the items care was taken to avoid potential interfering effects on pronoun interpretation: All items were controlled and matched for animacy, definiteness, number and gender, and they all displayed unambiguous case-marking. Further, the items were preceded by a context sentence that introduced both NPs. This context sentence neutralized information-structural differences between the NPs and, by making the NPs given information, established the contextual requirements for felicitous scrambling. The items and their preceding contexts were followed

by a sentence with a referentially fully ambiguous pronoun and a question to be answered by the subjects. For (3) this would be:

- (6) Aber er war zu unstrukturiert.
But it was too unstructured.
- (7) Was war zu unstrukturiert? (a) Der Plan (b) Der Beschluss (c) Etwas anderes
What was too unstructured? (a) The plan (b) The decision (c) 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 the plausibility test, the pronoun in sentence (6) was replaced by the full noun phrases. Subjects were presented with both options (as in (8) for the sentence in (3)) and they had to indicate which sentence, if any, seemed more plausible.

- (8) 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.

Due to a bias towards one of the readings, four items were excluded from further analysis. For the remaining 14 items, order and syntactic function of the NPs constituted the only differences for pronoun resolution.

2.2 Subjects and Procedure

Twenty-three German adult native speakers took part in the experiment. The subjects were recruited individually by e-mail. The task was untimed and conducted via the Internet. Subjects logged on to the experimental web site and filled in some biographical information. Subjects were told that they would read ambiguous sentences and were asked to give their preferred interpretation by choosing an answering to a question about the items. A complete experimental example as seen by the subjects for item (3), looked like the following (but without the italicized gloss):

- (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.
- Die Hoffnung war, dass der Beschluss den Plan beeinflussen würde.
It was hoped that the decision would affect the plan (canonical word order).
- Aber er war zu unstrukturiert.
But it was too unstructured
- Was war zu unstrukturiert?"
What was too unstructured?
- a. Der Beschluss
The decision
- b. Der Plan
The plan
- c. Etwas anderes
Something else

The experimental items were interspersed with 18 filler items of various other structures involving ambiguous and unambiguous pronouns, and pseudo-randomized. Two lists were used which counterbalanced the order of items, the order of answering options and scrambling.

2.3 Results

The results are given in Tables 1 and 2. In these tables, we refer to grammatical function in terms of obliqueness, with the subject being least oblique and the indirect object most oblique. Table 1 shows that for the su-do condition, the subject was chosen as the preferred antecedent of the fully ambiguous pronoun roughly two-thirds of the time, whereas the leftmost NP was chosen at levels indistinguishable from chance. In the su-io condition, the subject was also preferred as the pronoun antecedent; again, the leftmost NP was chosen at chance levels. The same finding obtains if we collapse the su-do and the su-io conditions to see the effects of subjecthood vs objecthood (see column *su-X*). For ditransitive verbs, the rightmost column in Table 1 demonstrates that neither obliqueness nor leftmost NP accounts for more than 50% of pronoun resolution. In subsequent statistical analyses, three pairwise Wilcoxon Signed-Rank tests were run for each condition: Subject vs Leftmost NP, Subject vs Object (for do-io: Direct Object vs Leftmost NP and Direct Object vs Indirect Object, respectively) and Leftmost NP vs Rightmost NP. For the su-do condition, only the Subject vs Object comparison showed a significant difference ($z = -2.425$, $p = 0.015$). For the su-io condition, the Subject vs Leftmost NP ($z = -2.517$, $p = 0.012$) and the Subject vs Object comparison ($z = -3.823$, $p < 0.001$) reached significance. For the combined subject-object conditions (*su-X*), both Subject vs Leftmost NP ($z = 2.862$, $p = 0.004$) and the Subject vs Object ($z = -4.509$, $p < 0.001$) comparisons showed significant differences. Across conditions, none of the Leftmost vs Rightmost NP comparisons reached significance. In the do-io condition, no comparison yielded significant differences. In sum, subjecthood turns out to be a better predictor of anaphoric reference than linear order, irrespective of the obliqueness of the object. For object antecedents, neither grammatical function nor linear order seems to govern pronoun resolution.

To show the lack of relation between anaphoric reference and word order in more detail, Table 2 presents a breakdown of the results by condition and word order. Table 2 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, anaphoric resolution preferences remain oriented to the subject in the su-do and su-io 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.

Resolved to	Condition			
	su-do	su-io	<i>su-X</i>	do-io
least oblique	44	78	122	68
%	64.7	67.8	66.7	49.3
leftmost	36	59	95	69
%	52.9	51.3	51.9	50.0
#Items	68	115	183	138

Table 1: Comparison of obliqueness & leftmost NP preferences

Condition	Word order	Least Oblique	
		#yes	#no
su-do	su left	22	13
	do left	22	11
su-io	su left	41	17
	io left	37	20
do-io	do left	33	36
	io left	35	34

Table 2: Interaction between LO & GF

3 Discussion and Conclusion

The results across different types of NP arguments in the German Mittelfeld indicate that GF, or, more specifically, subjecthood is a predictor of pronoun resolution. Even though referentially fully ambiguous, pronouns were found to refer at levels above chance to the subject of a preceding embedded clause that offered multiple potential antecedents. 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. GF turned out not to have an effect on pronoun resolution for object-object ambiguities. Counter to the claims by Rambow (1993), LO was not found to determine pronoun resolution in relation to the German Mittelfeld. In fact, in the present study, there is no evidence to suggest that LO affects resolution at all.

With respect to the role of subjecthood, these results are in line with the findings by Järvikivi et al. (2005), Kaiser (2003) for Finnish and Hoffman (1998) and Turan (1998) for Turkish. However, in their eye-tracking experiments, Järvikivi et al. (2005) tested for differences between subjecthood and first-mention effects, and not linear order. In a similar vein, the effects of LO, first-mention and topicalization potentially overlap and interact in the other studies. The present study avoids the interference of first-mention or topicality effects with LO, and the lack of LO effects found in the present study might suggest that LO effects previously reported in the literature could be due to these interfering factors. Future research on e.g. topicalization in German etc., will show whether this tentative suggestion is on the right track and holds cross-linguistically.

However, testing anaphoric preferences with scrambling in the German Mittelfeld also faces several potential problems. First, scrambling in German is marked and infrequent. Secondly, scrambling is restricted to specific discourse contexts. In particular scrambling of objects across subjects is a marked and infrequent reordering option in German. Hence, it could be that linear order effects in this study are affected or even masked by the frequency bias against scrambling. While it is difficult to isolate the effects of frequency, there is some evidence that frequency does not modulate linear order preferences in pronoun resolution with respect to the German Mittelfeld. 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 huge 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-do) orders, there is only one case of an accusative-nominative (do-su) order, for dative-marked indirect objects, there are 20 cases of dative-nominative (io-

su) orders compared to 43 nominative-dative (su-io) orders. This corresponds to a ratio of roughly 1 to 500 for do-su orders and roughly 1 to 2 for io-su orders. If frequency differences of scrambling affected linear order effects in pronoun resolution, we would thus expect to see a concomitant difference in anaphoric preferences between the su-do and the su-io conditions. Yet, the figures in Tables 1 & 2 demonstrate that there is no such difference between conditions. Moreover, Kempen and Harbusch (2003) note that, for ditransitive verbs, io-do orders by far outnumber do-io orders in the corpus (14 to 3). However, anaphoric preferences do not shift depending on NP ordering in the do-io condition (see Table 2). Consequently, we conclude that frequency differences of NP order in the German Mittelfeld do not modulate pronoun resolution preferences and thus cannot account for the lack of LO effects attested in this study.

As for the discourse requirements, 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 (Müller, 1999). In future research, it would be interesting to vary the information-structural contexts of ambiguous pronouns systematically to test for potential interactions between word order and information structure in anaphor resolution.

Ideally, experimental psycholinguistic studies on pronoun resolution as this one should be supplemented by corpus research to determine whether the same factors mandate anaphoric preferences in comprehension and naturalistic production. However, the multitude of potential relevant variables that need to be controlled for in conjunction with the rarity of scrambling in German will most likely lead to problems of data sparseness. Worse still, identifying the exact properties of contexts in a natural language corpus is difficult and subject to inter-annotator variation. In future research, we hope that these problems can be overcome so that we gain a more comprehensive understanding of pronoun interpretation.

Acknowledgments

As with the antecedents of pronouns, linear order of the authors' names does not correlate with their respective contributions to the paper. We would like to thank three anonymous reviewers for helpful comments on an earlier version of this paper. The first author gratefully acknowledges financial support by the Netherlands Organization for Scientific Research (NWO), grant number 051.02.071, as part of the NWO Cognition Programme.

References

- Brennan, Susan E., M.W. Friedman, and Carl J. Pollard. 1987. A centering approach to pronouns. In *Proceedings of the 25th Annual Meeting of the ACL*, pages 155–162.
- Gordon, Peter, Barbara Grosz, and Laura Gillom. 1993. Pronouns, names, and the centering of attention in discourse. *Cognitive Science*, 17(3):311–347.
- Grosz, Barbara, Aravind Joshi, and Scott Weinstein. 1995. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, 21(2):203–225.

- Haider, Hubert. 1993. *Deutsche Syntax, generativ*. Günter Narr, Tübingen.
- Haider, Hubert and Inger Rosengren. 1998. *Scrambling*, volume 49 of *Sprache und Grammatik*. Lund University.
- Hoffman, Beryl. 1998. Word order, information structure, and centering in Turkish. In M. Walker, A. Joshi, and E. Prince, editors, *Centering Theory in Discourse*. Clarendon University Press, Oxford, pages 251–273.
- Järvikivi, Juhani, Roger van Gompel, Jukka Hyönä, and Raymond Bertram. 2005. Ambiguous pronoun resolution. contrasting the first-mention and subject-preference accounts. *Psychological Science*, 16(4).
- Kaiser, Elsi. 2003. Word order, grammatical function, and referential form: On the patterns of anaphoric reference in Finnish. *Nordlyd*, 31(1):245–260.
- Kempen, Gerard and Karin Harbusch. 2003. An artificial opposition between grammaticality and frequency: comment on Bornkessel, Schlesewsky & Friederici (2002). *Cognition*, 90:205–210.
- Lenerz, Jürgen. 1977. *Zur Abfolge Nominaler Satzglieder im Deutschen*. Günter Narr, Tübingen.
- Müller, Gereon. 1999. Optimality, markedness and word order in German. *Linguistics*, 37:777–818.
- Poesio, Massimo, Rosemary Stevenson, Barbara di Eugenio, and Janet Hitzeman. 2004. Centering: A parametric theory and its instantiations. *Computational Linguistics*, 30(3):309–363.
- Prasad, Rashmi and Michael Strube. 2000. Discourse salience and pronoun resolution in Hindi. In *Penn Working Papers in Linguistics 6.3*. University of Pennsylvania, pages 189–208.
- Prince, Ellen. 1981. Toward a taxonomy of given-new information. In P. Cole, editor, *Radical Pragmatics*. Academic Press, New York, pages 223–256.
- Rambow, Owen. 1993. Pragmatic aspects of scrambling and topicalization in German: A Centering approach. MS, University of Pennsylvania.
- Scheepers, Christoph, Barbara Hemforth, and Lars Konieczny. 2000. Linking syntactic functions with thematic roles: psych-verbs and the resolution of subject-object ambiguity. In B. Hemforth and L. Konieczny, editors, *German sentence processing*. Kluwer, Dordrecht, pages 95–135.
- Strube, Michael and Udo Hahn. 1999. Functional Centering. grounding referential coherence in information structure. *Computational Linguistics*, 25(3):309–344.
- Turan, Ümit Deniz. 1998. Ranking forward-looking centers in Turkish: Universal and language-specific properties. In A Joshi M. Walker and E. Prince, editors, *Centering Theory in Discourse*. Clarendon University Press, Oxford, pages 139–161.