The Non-Subject Bias of German Demonstrative Pronouns

Peter Bosch, Graham Katz, Carla Umbach
Institute of Cognitive Science, University of Osnabrück

Abstract

This paper presents evidence from both corpus studies and psycholinguistic experiments regarding the referential preferences of German personal pronouns (*er sie, es*) and demonstrative pronouns (*der, die, das*). Both types of pronouns are naturally used in German to refer to both animate and inanimate referents, but they appear to have different preferences with respect to their choice of antecedents: While personal pronouns show a slight preference for antecedents which are grammatical subjects, demonstrative pronouns show a strong bias for non-subject antecedents.

1. Introduction

Linguistic constraints on potential antecedents for anaphoric expressions have been investigated in syntactic theory under the rubric of Binding Theory (Chomsky 1981) and in more semantic settings in discourse representation theory (Kamp 1981) and centering theory (Grosz, Joshi and Weinstein 1995), among others. It is well known that syntactic, semantic, and pragmatic factors all play a role in resolving the question of which of the many potential antecedents for a particular use of a pronoun is intended by the author as the antecedent. Since Hobbs (1978), computational linguists have sought to investigate these preferences (Lappin and Leass 1994; Preuss et al. 1994; Mitkov 1998). In this paper we investigate a preference for antecedents that appears to be a lexical property of the pronominal element itself. The anaphoric expressions in question, the little-studied class of German demonstrative pronouns, appear to evidence a bias for non-subject antecedent noun phrases.

German makes use of two sets of pronominal elements, the *er (sie, es)* series, which we will term personal pronouns, and the *der (die, das)* series, which we will term demonstrative pronouns. Both varieties of pronouns can be naturally used to refer to animate antecedents, but they appear to have slightly different preferences with respect to which antecedent is preferred. In particular while the personal pronouns show a familiar slight preference for antecedents which are grammatical subjects, demonstrative pronouns show a strong bias for non-subjects. In this paper we will present an investigation of these contrasting preferences, with evidence taken from both large scale corpora and from two psycholinguistic studies. Our results seem to confirm the results from Dutch and Finnish reported by Kaiser & Trueswell (2004; in press), and the analogous results reported for Italian by Carminati (2002)[1].

German personal and demonstrative pronouns appear at first blush to be synonymous, differing, if at all, only in register or affect. In (1), for example, either pro-form will be understood as referring back to the subject of the first sentence.

(1)   Ein Mann sprach mich an. {Der/Er} wollte wissen, wo die Kolpingstraße ist.
        [A man was addressing me. {DEM[2]/He} wanted to know where Kolpingstraße is.]

---

[1] Only after we finished this paper we became aware of Comrie (1997), who investigated the opposition between personal and demonstrative pronouns for Dutch with results that are entirely compatible with our results for German and who was probably the first to carry out a systematic corpus study of this opposition.
[2] In the English glosses we represent the German demonstrative simply by "DEM".

But this synonymy does not extend to all contexts. In (2), for example, we find different interpretation preferences for the two pronouns:

(2)   Paul wollte mit Peter laufen gehen. Aber {er / der } war erkältet.
        [Paul wanted to go running with Peter. But {he /DEM} had a cold.]

Here, the pronoun *er* is most naturally interpreted as referring to *Paul*, the grammatical subject of the prior sentence, while the demonstrative *der* has a preference for *Peter*, the grammatical object. Subtle pragmatic factors can in some cases override the preferences, but absent these, the lexical preferences remain. The studies we report here support the intuition that such a preference exists.

2. Linguistic Background

The *der (die, das)* series demonstrative pronouns are for the most part homophonous with the definite determiners in German. While it might be suspected that these forms are to be analyzed as full definite noun phrases with elided head nouns, and that the difference evident in (2) is derivable from these properties, there are reasons to set this concern aside. First, the interpretation of demonstrative pronouns depends on salience information to a degree that we cannot find in the interpretation of full definite NPs (Gundel, Hedberg and Zacharski 1993). Secondly, while many forms of the demonstrative pronouns look like definite determiners, there are forms, such as the genitive form and the dative plural, in which a difference becomes apparent. For example, the dative plural demonstrative pronoun is *denen* (3a) but the definite article is *den* (3b).

(3) a.   Denen ist es egal.
    b.   Den Frauen ist es egal.
         [It doesn't matter to them / the women.]

We have in much of the work limited our attention to the masculine singular form *der* (accusative *den*) because the other forms present problems which make direct investigation difficult. The feminine singular form *die* is syncretic with the plural prounoun and the neuter form *das* is used to refer to abstract objects such as facts and events (cf. also Consten et al. (this volume)). For a language like German with grammatical gender, this is less of a limitation than it would be for a natural gender language such as English. We have also ignored the morphologically more complex (and significantly rarer) demonstrative paradigms *dieser, jener, derselbe,* ('this', 'that', 'the same') and their likes. Here it is not clear that their uses are similar to the simple demonstrative pronouns, however.

Traditional grammatical description of the difference between personal pronouns and demonstrative pronouns is surprisingly cursory. Duden (1984) suggests that personal and demonstrative pronouns are in free variation, with only a difference in tone to distinguish them. This appears to have to do with a difference in evaluative attitude toward the intended referent. In the case of (1), for example, the use of the demonstrative *der* appears to distance the speaker in some way from the man being referred to and is somehow pejorative. We have nothing to say about this potential pragmatic difference here. Our investigation is more semantically focused. The IDS Grammar (Zifonun et al. 1997) suggests, in this regard, that the semantic difference is that personal pronouns are used if the antecedent relates to a topic which has been established before whereas the demonstrative pronoun is used if the antecedent is part of the rheme or is in contrast with other topics.

It is also a common observation that the use of demonstrative pronouns is limited to or primarily to be found in spoken language. This observation is one which we were able to test directly. We compared in two different corpora[3], one of spoken German and one of written German texts, the number of demonstrative uses of *der* with the number of uses of *er* (the corresponding personal pronoun). While in our newspaper corpus fewer than 7% of the masculine pronoun tokens were demonstratives, in our spoken language dialog corpus approximately 80% were. This is a dramatic difference, but one which is likely as much attributable to text genre as to the spoken/written contrast. While this is an interesting finding in its own right, it tells us little about the factors that contribute to the choice of one pronominal form over the other, given that in both spoken and written language, both forms are used. In the next sections we will investigate one of the factors that appears to be crucial in determining how demonstrative pronouns are used, namely the grammatical function of the intended antecedent.

We sought to investigate the contrast using two complementary methodologies. First we examined the distribution of demonstrative and personal pronouns in text corpora, looking both at the sentence internal distribution and the pronoun antecedent relation. The corpus studies are reported in Section Three. We also carried out two psycholinguistic studies, these are reported on in Section Four. Our goal here is not to account for the observed differences but to establish what the differences are that need to be accounted for. In the concluding section we will offer some speculations about how to explain the observed differences.

3. Corpus Investigation

We made use of a hand-parsed corpus of newspaper texts[4] to investigate the distributional patterns associated with personal and demonstrative pronouns. In this study we investigated two different aspects of pronoun use, their distribution of the pronominal element within its own sentence and the distributions of the antecedents for these elements in prior sentences. In both cases there were striking differences between the two classes of pronouns.

3.1 Sentence-internal distribution

The first question we investigated was whether there were obvious differences in the sentence-internal distribution of demonstrative and personal pronouns in simple German sentences. As is well known, German is a verb-second language, with a distinguished topic position preceding the tensed verb of the main clause. This pre-verbal position can be filled by any of the grammatical arguments of the verb, or by adverbial material. This means that in simple transitive clauses there are four different potential positions that a pronoun can occupy: pre-verbal subject, post-verbal subject, post-verbal object and pre-verbal object. These are illustrated in (4).

(4) a. {Er/Der} hat Hans gesehen.
         [{He/DEM} has Hans seen]
     b. Hans hat {er/der} gesehen.
         [Hans has {he/DEM} seen]

"He has seen Hans"
c.   Hans hat {ihn/den} gesehen.
     [Hans has {him/DEM} seen]
d.   {Ihn/den} hat Hans gesehen.
     [{Him/DEM} has Hans seen]
     "Hans has seen him"

In this study we compared the number of token uses of personal pronouns in each of the four positions to those of demonstrative pronouns. The results are indicated in Fig. 1. The first thing to note is the clear dominance of subject uses over object uses for both types of pronouns (in both cases about four to one). In contrast to personal pronouns, demonstrative pronouns have a clear preference for pre-verbal positions. Approximately 93% of all demonstrative tokens occur in this position, either as subjects or objects. Personal pronouns used as subjects are evenly divided between pre- and post-verbal uses, while personal pronouns used as objects are overwhelmingly post-verbal (not even 0.5% are used pre-verbally). This contrasts strikingly with demonstrative pronouns, where pre-verbal object uses far dominate post-verbal object uses.
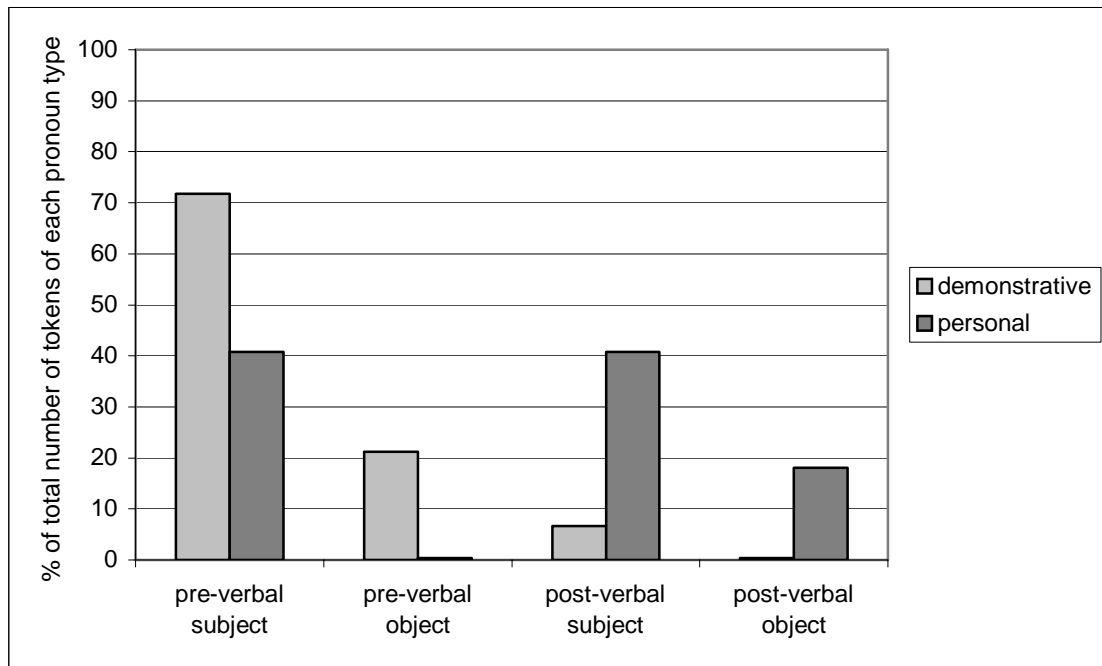

Fig.1: The distribution of demonstratives and personal pronouns over different
        syntactic positions

We might summarize by saying that the pre-verbal subject position is a natural position for both demonstrative and personal pronouns, while the post-verbal subject position is an unnatural position for demonstrative pronouns and the pre-verbal object position is an unnatural position for personal pronouns. The preferred syntactic positions of a pronoun can, of course, give only a very indirect hint at the discourse parameters governing its use. Even if we believe that the choice of syntactic position is to some extent a function of information structure, we expect the pronoun-antecedent relation to be of central importance to this contrast. A natural next step was to compare the antecedents of demonstrative pronouns with those of personal pronouns.

3.2 Preference for antecedent types

For this study, in order to investigate the potential differences in antecedent-pronoun relations between demonstrative and personal pronouns we hand-annotated a corpus of German newspaper texts with pronoun antecedent relations.[5] For practical reasons, we limited ourselves to cases in which the antecedent for a pronoun appeared in the immediately preceding sentence. Even in these cases there are a number of aspects of the pronoun-antecedent relation that one could choose to investigate. The grammatical function of the antecedent, the distance of the antecedent measured in syntactic nodes, the depth of embedding of the antecedent, the semantic class of the antecedent, syntactic parallelism between anaphora and antecedent are all factors that have been suggested to play a role in anaphora resolution (cf. Lappin and Leass 1994).

As a first step, we chose to investigate the grammatical function of the antecedent NP. We compared the number of uses of demonstrative pronouns with antecedent NPs that were grammatical subjects of their sentences with the number of uses of in which the antecedents were objects. The same comparison was made for personal pronouns. The four types of discourses illustrated in (5) and (6) were compared, where we have underlined the pronouns and their antecedents.

(5) a.   Hans reparierte <u>den Rolladen</u>. <u>Der</u> hatte sich verklemmt.
           [Hans was repairing <u>the roller blind</u>. <u>DEM</u> had got stuck.]
     b.   <u>Hans</u> reparierte den Rolladen. <u>Er</u> fluchte ununterbrochen.
           [<u>Hans</u> was repairing the roller blind. <u>He</u> kept swearing all of the time]

(6) a.   <u>Hans</u> reparierte den Rolladen. <u>Der</u> fluchte ununterbrochen.
           [<u>Hans</u> was repairing the roller blind. <u>DEM</u> kept swearing all of the time]
     b.   Hans reparierte <u>den Rolladen</u>. <u>Er</u> hatte sich verklemmt.
           [Hans was repairing <u>the roller blind</u>. <u>It</u> had got stuck.]

The results, displayed in Fig. 2, show a strong preference of demonstratives for object NP antecedents; Personal pronouns, on the other hand, show a complementary behaviour and prefer subject NPs as their antecedents. In other words, discourses such as (5a) and (5b) are common, while those such as (6a) and (6b) are rare. The frequencies displayed in Fig. 2 reflect quite well native speakers' intuitions about the relevant uses. The discourses in (5) are judged natural, while those in (6), in particular the one in (6a), appear less natural.
Our corpus studies showed us that there was indeed a striking difference in use between demonstrative and personal pronouns. The typical use of a demonstrative is as a pre-verbal subject referring back to an NP object of the prior sentences, while the typical use of a pronoun is as a pre-or post-verbal subject referring back to the NP subject of the prior sentence. We sought to confirm that this difference in use is correlated with a processing difference by carrying out two psycholinguistic studies. These are described in the next section.

4. Experiments

The two psycholinguistic experiments we carried out were designed to complement each of our corpus studies. The first experiment investigated the naturalness of demonstrative and

---

[5] We used again the NEGRA corpus. As one might expect, in very few cases was the intended antecedent unclear. These cases were discarded. The counts include singular masculine and feminine and plural forms. Neuter forms were excluded.

personal pronouns in the various sentence-internal syntactic positions, while the second investigated the naturalness of the pronoun-antecedent relation for various positions of the antecedent.
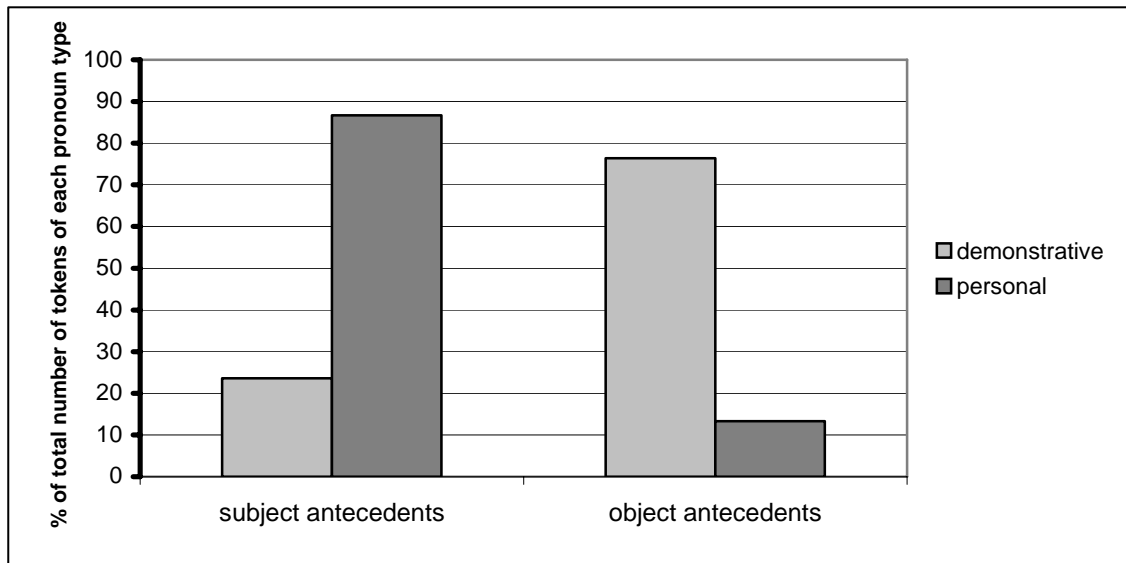

Fig.2: The distribution of demonstratives and personal pronouns over two types
of antecedents

4.1 Experiment One: Sentence-internal parameters

Thirty students of the University of Osnabrück participated in the first experiment (9 female and 21 male). Experimental materials consisted of a total of 64 sentences. Thirty-two of these, 16 containing a personal pronoun and 16 containing a demonstrative pronoun, were extracted form the *Frankfurter Rundschau* corpus[6]. The sentences were chosen so that the position of the pronoun varied along the parameters pre-verbal/post-verbal position and subject/object, as illustrated in (7):

(7) a. pre-verbal subject: Der stellte der Hausverwaltung ein Ultimatum.
[DEM put an ultimatum to the property administration.]
b. pre-verbal object: Doch den könnte die Gemeinde verteilen.
[But the council could distribute DEM.]
c. post-verbal subject: Ein Buch wollte er nicht geschenkt haben.
[A book he didn't want for a present.]
d. post-verbal object: Man konnte ihn nicht mehr erneuern.
[One couldn't renew it any more.]

This was the naturally occurring set. The other 32 were generated by substituting a demonstrative for each personal pronoun and conversely a personal pronouns for each demonstrative. The intent was to determine whether the naturally occurring set was judged more normal sounding by the participants than the artificially generated set, the assumption

---

[6] The *Frankfurter Rundschau* Corpus is a large (41 Mio words) collection of German national daily newspaper text.

being that the sentence-internal factors that influenced the original author to use the form that he did use would make the other form in some way unnatural. This we investigated via a self-paced reading study and a follow-up questionnaire study.

In the self paced reading study, participants were asked to read 16 of the naturally occurring sentences and 16 artificial sentences. The stimuli were evenly divided among the different sentence types. Participants were asked to read the sentences as they appeared on a computer monitor carefully and with a normal speed. The sentences were displayed in a word by word fashion, with only one word displayed at a time. The reading times for each word were recorded. After each sentence a comprehension question was asked.

An analysis of the results showed no overall significant difference between the time course for reading the naturally occurring sentences and the artificially generated sentences. Also an analysis of variance with z-transformed reading times (so that the reading times of different subjects become comparable) showed no significant differences for the reading times of the target pronoun or the following word. While we had hoped to find that the reading times of the substituted pronouns was elevated, indicating recognition of a disfluency, the lack of significant elevation may only indicate that the effect is too small to be detected by this method.

After the completion of the self-paced reading study, subjects were presented with a list of all 64 sentences in pairs of original and substitute and were told that one sentence in each pair had actually appeared in a newspaper while the other one had not and they were asked to guess which was which. If a subject marked a sentence as "appeared in print" this was interpreted as an intuitive preference for the relevant choice of pronoun. The analysis of the responses, shown in Fig. 3, revealed a clear and significant preference for the personal pronoun in the post-verbal position of the sentence. The evident preference of the demonstrative for the pre-verbal positions was suggestive, but not significant.
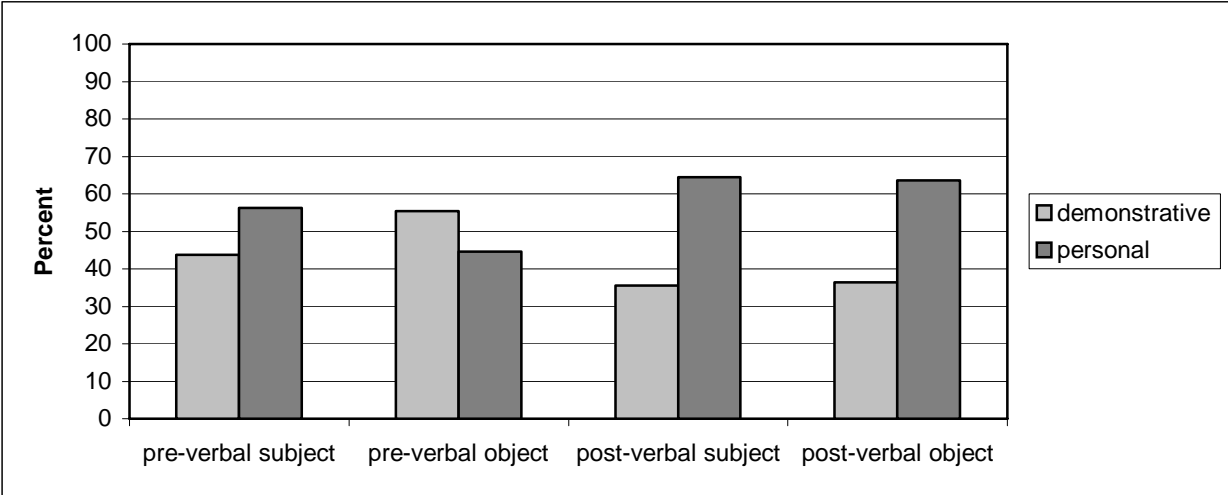

Fig. 3: Questionnaire results. Preference for personal and demonstrative pronouns in different syntactic positions

While the result of the questionnaire study is congruent with the quantitative corpus study reported above, there is very little that can be concluded.

4.2 Experiment Two: Discourse Parameters

While the first experiment looked at sentences in isolation, the second experiment takes both discourse context and background knowledge into account. We saw already in example (2) that the interpretation preferences for personal and demonstrative pronouns are sensitive to the way the potential referents are introduced in preceding sentences. And, as always, interpretation preferences may be influenced by what a listener or reader expects on the basis of general world knowledge. Consider (8)

(8)   Paul spielte Tennis mit Peter. {Er / Der} hat ihn gewinnen lassen.
        [Paul was playing tennis with Peter. {He /DEM} let him win.]

It is plausible to assume that any knowledge of the reader of (8) about how good Peter and Paul are at tennis would influence the interpretation of the pronoun. If the reader knows that Paul is an international champion while Peter plays tennis only once in a while, the likely interpretation is going to be that Paul let Peter win – never mind the type of pronoun. So we may expect a competition between intrinsic interpretation preferences of the pronoun type, as they are suggested by our corpus data, and world-knowledge-based preferences. The basic idea in experiment two is to make use of this competition. If an antecedent which is preferred due to factors relating to grammar is unlikely due to world knowledge, interpretation load should be increased.

Accordingly, we designed the stimulus materials taking both factors into account. Each stimulus set consisted of four lines of text. The first line (headline) introduces a situation and the second (context sentence) brings in two protagonists. The third line (target sentence) refers to one of the protagonists by either a demonstrative or a personal pronoun, and the final line (completion sentence) contains a gap, which subjects are asked to fill in with a noun so that this line would paraphrase their understanding of the target sentence. For example:

1. Im Krankenhaus
   [At the hospital]
2. Der Oberarzt untersucht den Notfallpatienten.
   [The senior doctor is examining the emergency patient.]
3. Er ist gerade erst gekommen.
   [He has only just arrived.]
4. Der ＿＿＿ ist gerade erst gekommen.
   [The ＿＿ has only just arrived.]

In this example world knowledge implies no preference for the doctor or the patient as a referent for the target pronoun. The intrinsic interpretation preferences for the personal pronoun in the target sentence, however, may still lead to an interpretation that takes the doctor as the referent.

Stimulus sets were varied with respect to the parameters world knowledge bias and pronoun type. In each stimulus set world knowledge would support either the first or the second protagonist, that is, either the subject or the object of the context sentence as the referent of the target pronoun, or be neutral in this respect. The pronoun would either be a demonstrative or a personal pronoun. So we had six versions of the stimulus set with identical headline and context sentence, differing with respect to the target sentence and the completion sentence:

1. Im Krankenhaus [At the hospital.]
2. Der Oberarzt untersucht den Notfallpatienten.
   [The senior doctor is examining the emergency patient.]

Version one (world knowledge supporting the subject; personal pronoun)
    3. Er ist nämlich Herzspezialist. [He is a heart specialist.]
    4. Der ___ ist nämlich Herzspezialist. [The ____ is a heart specialist.]

Version two (world knowledge supporting the subject; demonstrative pronoun)
    3. Der ist nämlich Herzspezialist. [DEM is a heart specialist.]
    4. Der ___ ist nämlich Herzspezialist.

Version three (world knowledge supporting the object; personal pronoun)
    3. Er muß sofort operiert werden. [He must be operated on at once.]
    4. Der ___ muß sofort operiert werden. [The ___ must be operated on at once.]

Version four (world knowledge supporting the object; demonstrative pronoun)
    3. Der muß sofort operiert werden. [DEM must be operated on at once.]
    4. Der ___ muß sofort operiert werden.

Version five (world knowledge support neutral; personal pronoun)
    3. Er ist gerade erst gekommen. [He has only just arrived.]
    4. Der ___ ist gerade erst gekommen. [The ___ has only just arrived.]

Version six (world knowledge support neutral; demonstrative pronoun)
    3. Der ist gerade erst gekommen. [DEM has only just arrived.]
    4. Der ___ ist gerade erst gekommen

Seven such sets were constructed: Five in which the context sentence is in SVO (subject-verb-object) order (e.g. *Der Oberarzt untersucht den Notfallpatienten*) and two with context sentences in OVS (object-verb-subject) order (e.g. *Den Notfallpatienten untersucht der Oberarzt*). This yields a total of 42 stimulus sets. To prevent participants from getting used to the routine of choosing antecedents we interspersed additional stimulus sets as distractors, in which there was a different completion task in line four. 69 students of the University of Osnabrück participated in the experiment (23 female and 46 male). Participants were randomly assigned to one of six groups so that each participant only saw one version of each block of stimulus sets.

      Participants read the stimulus sets line by line as they were displayed in the upper left corner of a computer monitor. Each set was started by a star displayed at the position of the first character of the first line to focus the participant's attention before the stimulus sets were displayed. Only one line of the stimulus set was displayed at a time. When the line with the gap was shown, participants had to type in their completion. By pressing the enter key the stimulus set was finished and the next one was started. The reading times for the target sentence were measured and the answers that participants typed into the gap were recorded.

4.2.1 Reading time results

The reading times for the target sentence are as presented in Figs. 4 and 5. The columns for 'subject support' represent the averaged reading times in milliseconds for those targets in which world knowledge would support an interpretation of the pronoun with reference to the

subject of the context sentence and the columns for 'object support' give the reading times when world knowledge suggests the object of the context sentence as the appropriate referent. The columns given as 'unsupported' are for those cases where world knowledge does not induce any preference.

As Fig. 4 shows for context sentences with SVO structure, reading times are longer for the demonstrative pronoun when the pronoun is interpreted in reference to a preceding subject protagonist. Thus in the context of (9) continuation (a) takes more time to read than (b):

(9) Der Oberarzt untersucht den Notfallpatienten.
    a.   <u>Der</u> ist nämlich Herzspezialist.
    b.   <u>Er</u> ist nämlich Herzspezialist.
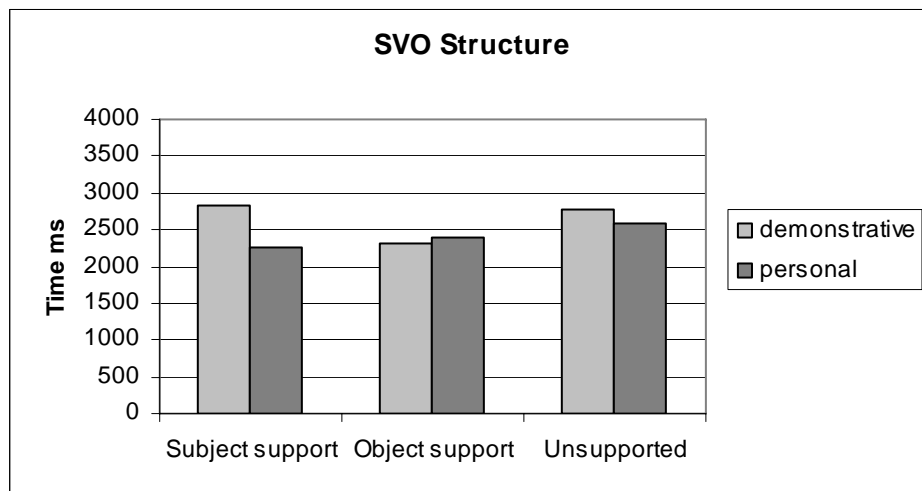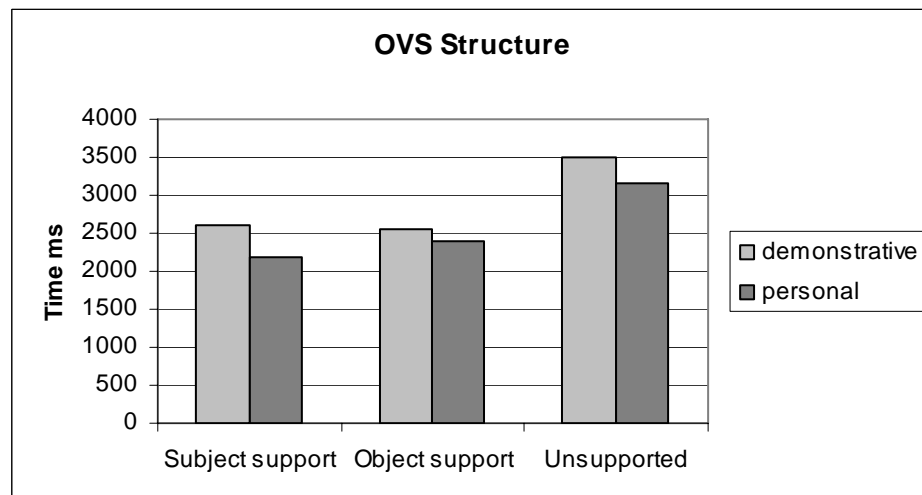
Fig. 4 Reading times with SVO context sentences

Fig. 5 Reading times with OVS context sentences

And the same holds for context sentences with OVS structure, as shown in Fig 5. Also here (a) takes significantly longer than (b)[7]. Cf. (10).

---

[7] The reading times displayed in Figs. 4 and 5 are not directly comparable because the number of stimulus sets was different. They were measured for SVO structure on 30 different stimulus sets, while the OVS structure only had 12 different stimulus sets. So the results for the subject verb object structure are more reliable.

(10) Den Notfallpatienten untersucht der Oberarzt.
     a. <u>Der</u> ist nämlich Herzspezialist.
     b. <u>Er</u> ist nämlich Herzspezialist

The remaining reading time results for object support and the unsupported cases unfortunately show too much variation among participants to count as significant.

4.2.2 Completion results

The results of the completion study, i.e., the answers that were filled into the gap in the fourth line of the stimulus sets are as in Figs. 6 and 7 (demonstrative pronouns are displayed as 'DEM' and personal pronouns as 'PERS'):
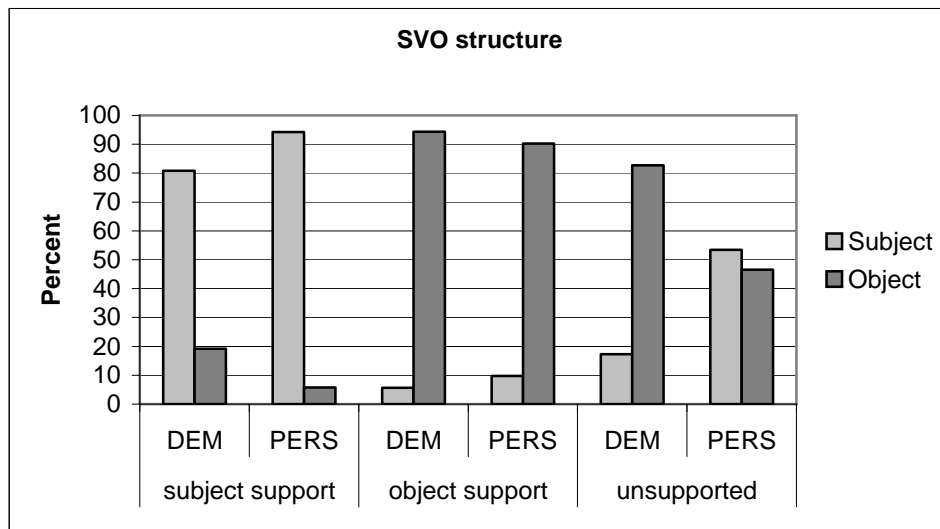

Fig 6   Choices for the subject or the object of the SVO context
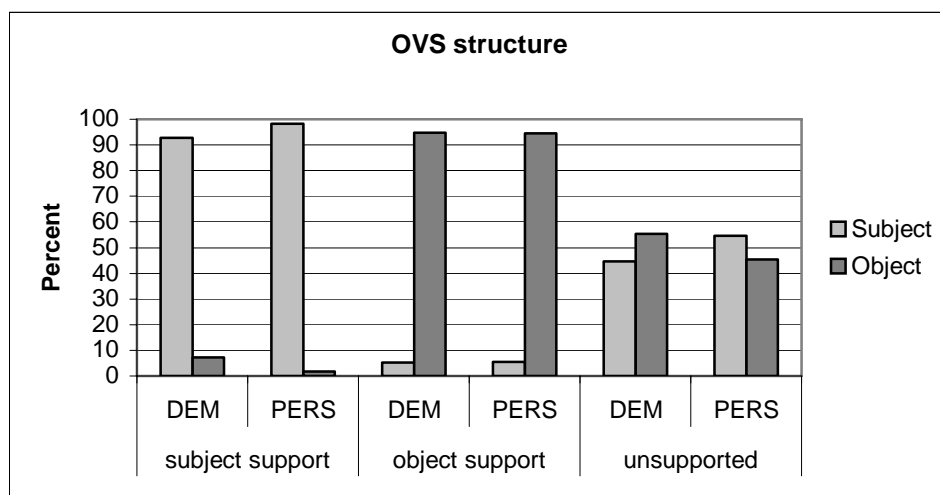   sentence as referent for the target pronoun.

Fig 7   Choices for the subject or the object of the OVS context
          sentence as referent for the target pronoun.

The central observation is that wherever there is a world knowledge bias for the interpretation of the pronoun, world knowledge is clearly the dominating factor in pronoun comprehension. In the unbiased cases we find a clear preference for one interpretation only where the demonstrative pronoun refers to the context sentence object in SVO structure (cf. Fig. 6, unsupported, DEM), while in the remaining cases (the other three column pairs of the unbiased cases) we can only detect tendencies: Preference for the object of the context sentence as antecedent for the demonstrative also in OVS structure and preference for the subject of the context sentence as antecedent for the personal pronoun in both SVO and OVS structure.

We can further observe that in the condition where world knowledge supports the subject of the context sentence, more participants disregard world knowledge and choose the object when the pronoun is a demonstrative than in those cases where the pronoun is a personal pronoun. Analogously, although less clearly, when world knowledge has a bias for the object of the context sentence more participants disregard world knowledge and choose the subject when the pronoun is a personal pronoun than in those cases where the pronoun is a demonstrative.

In sum, the completion study strongly supports the results from the reading time study: In both SVO and OVS conditions, the natural preference for a demonstrative would seem to be a reference to the object of the context sentence and for a personal pronoun the preference would be the subject of the context sentence. When these preferences agreed with the world-knowledge bias, reading times were shortest.

4.2.3 Questionnaire study

After the experiment participants were given a questionnaire with a list of all sentences they had seen and they were asked to recall which type of pronoun they had seen in the target sentence during the experiment. If we assume that participants would not normally be able to remember which pronoun they had seen, their answers may be taken to reflect their preference for a particular pronoun type, given a particular interpretation and context. Since in the 'unsupported' cases the choice of pronoun is not constrained by any interpretation, the answers were expected to be arbitrary. The results are given in Figures 8 and 9.
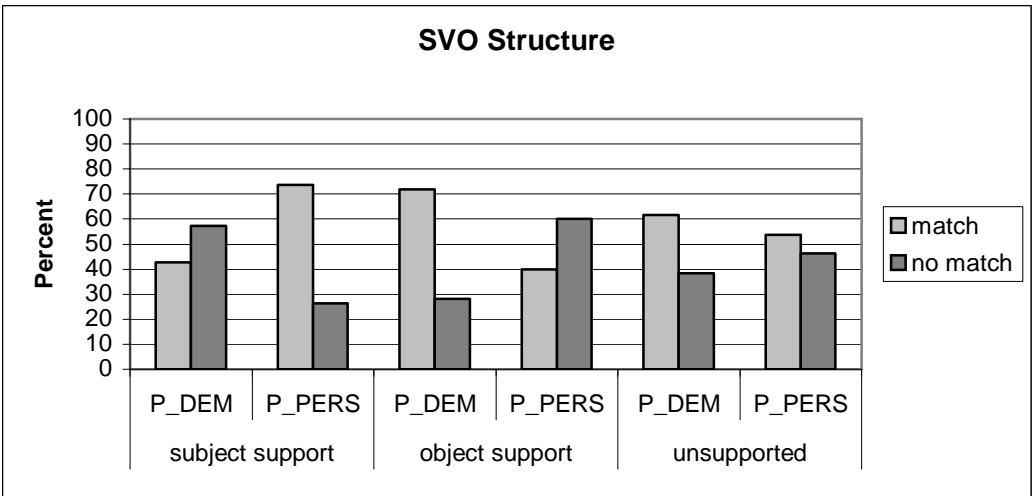

Fig 8   Recall of pronoun use in after-experiment questionnaire
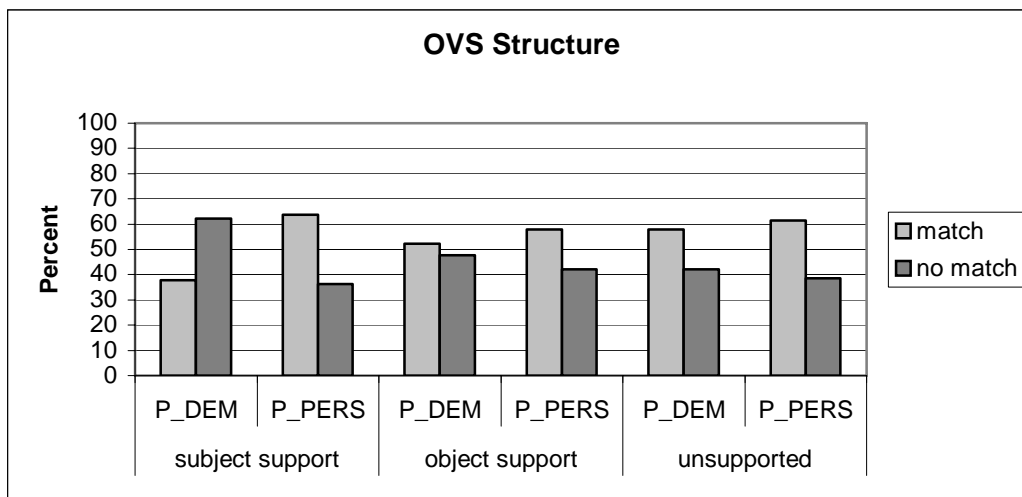
Fig 9   Recall of pronoun use in after-experiment questionnaire

'P_DEM' means that the sentence was presented with the demonstrative pronoun *der* during the reading time study. 'P_PERS' means that the sentence was presented with the personal pronoun *er*. The columns show matches between what the participants saw during the experiment and what they 'recalled' when filling in the questionnaire.

In the results for the SVO structure in Fig. 8 we find a preference for the personal pronoun to refer to the subject and the demonstrative to refer to the object of the context sentence, as we should have expected from the reading time and completion study. The OVS structure displayed in Fig. 9, however, is less clear. The tendency is indeed the same as in the SVO cases, with the exception of those targets that used the personal pronoun for reference to the context sentence object, which is remembered largely correctly and against the hypothetic preference for a demonstrative under these conditions. This may be due to a number of factors, but we have to be very careful at this point, because, as already mentioned, the OVS results rest on a fairly small number of stimulus sets (12 vs. 30 sets for SVO).

5. Conclusion

The corpus study as well as the experimental study show a clear preference of demonstrative pronouns to refer to object antecedents. These findings, however, raise the question of whether the grammatical function of the potential antecedent is the only factor biasing the choice of the antecedent for demonstrative pronouns, or whether this is correlated with another factor such as recency, salience, non-topicality, rheme, or accenting. Recency is ruled out by our findings because in the OVS condition the subject is mentioned only after the object and still the object is the preferred antecedent of the demonstrative pronoun.

Our findings are consistent with the Complementarity Hypothesis put forward in Bosch et al. (2003), which claims that demonstrative pronouns prefer non-topical referents while personal pronouns prefer referents which are established as discourse topics, assuming that topics are established by nominative NPs. Our findings are also consistent with the results of Winter (2003) who investigated demonstrative pronouns in the framework of Centering Theory (Grosz et al. 1995) showing that demonstratives choose their antecedents by inducing the least coherent transition relation (i.e. rough shift > smooth shift > retain > continue). Again, the underlying assumption is that subject antecedents are more salient than object antecedents.

In general, our findings are consistent with the widely accepted idea that lower-ranked referring expressions refer to less salient antecedents (cf., e.g., Gundel et al. 1993), assuming

13

that demonstrative pronouns rank lower than personal pronouns, and that objects are less salient than subjects. For Dutch personal and demonstrative pronouns Kaiser and Trueswell (2004) show that in the case of personal pronouns the subject is preferred over the object while in the case of demonstrative pronouns the object is even more preferred over the subject.[8]

Where English is concerned, examples like (11) seem to suggest that German demonstrative pronouns pair with stressed personal pronouns in English. This assumption is disproved, however, by cases like (12) and (13). In (12a) the stressed pronoun refers to *John* whereas the demonstrative in the corresponding German sentence in (12b) does not and is actually hard to interpret. Similarly, in (13) translating the demonstrative pronoun by a stressed personal pronoun makes the sentence infelicitous because it is unclear what the antecedent of the stressed pronoun should be. The examples in (11a) and (12a) stem from Kameyama (1999) who claims that English stressed pronouns take the complementary preference of the unstressed counterpart where preferences include constraints given by world knowledge. This predicts that a stressed pronoun is interpreted as referring to an antecedent unexpected from the point of view of world knowledge; in (12a), for example, the preference for the subject is due to the fact that the agent of a hitting event is less likely to be injured. In contrast, in the case of German demonstrative pronouns preference is clearly guided by grammatical factors. World knowledge may of course interfere, strengthening or weakening an interpretation preference, but it is not an intrinsic factor as in the case of English stressed pronouns.

(11) a. Paul called Jim a Republican. Then HE insulted HIM.   HE = Jim
      b. Paul nannte Jim einen Betonkopf. Dann beleidigte der ihn.   der = Jim

(12) a. John hit Bill. Then HE was injured.   HE = John
      b. John hat Bill geschlagen. ?? Dann wurde der verletzt.   der = ??
(13) a. Paul wollte mit Peter laufen gehen. Aber der war erkältet.   der = Peter
      b. Paul wanted to go running with Peter. But HE had a cold.   HE = ??

The question of whether the preference we found for German demonstrative pronouns to refer to object antecedents is a genuine effect of the grammatical function or a side effect triggered by some factor correlating with grammatical function cannot be settled by our data. Still, there is one observation casting doubt on the idea that the use of demonstrative pronouns is exclusively determined by the grammatical role of the antecedent, namely the clear preference of demonstrative pronouns to occur in the pre-verbal position of the sentence (cf. Section 3.1). There is no reason in the first place why a pronoun that prefers non-subject antecedents should preferably occur in a pre-verbal position. Word order in German is largely guided by information-structural requirements. Preference of demonstratives for a pre-verbal position suggests that information structure plays a crucial role in the use of demonstrative pronouns. This brings us back to (Zifonun et al. 1997) who propose that demonstrative pronouns are used if the antecedent is part of the rheme or is in contrast with other topics, and we might speculate that demonstrative pronouns are used for a topic change which would be supported by the findings in Winter (2003). Still, any explanation of the contrast between personal and demonstrative pronouns in terms of topicality or the theme-rheme division has to relate these notions to grammatical function to account for the non-subject bias of demonstrative pronouns.

---

[8] Actually, in Dutch two forms of personal pronouns have to be distinguished, a reduced form (unstressed) and a full form (mostly stressed). Different from what would be predicted by the correspondence of salience of antecedents and rank of expressions, Kaiser and Trueswell found that both reduced and full pronouns show a clear preference for the subject as the antecedent.

Acknowledgements

References

Bosch, P., Rozario, T., & Zhao, Y. (2003) Demonstrative Pronouns and Personal Pronouns. German *der* vs. *er*. Proceedings of the EACL2003. Budapest. *Workshop on The Computational Treatment of Anaphora.*

Comrie, B. (1997) Pragmatic Binding: Demonstratives as anaphors in Dutch. Proceedings of the Annual Meeting of the Berkeley Linguistic Society Vol. 23 (1997) :50-61

Consten, M., Knees, M. & Schwarz-Friesel, M. (this volume)

Carminati, M. N. (2002) The processing of Italian subject pronouns. Doctoral Dissertation, University of Massachusetts, Amherst, Amherst, MA.

Chomsky, N. (1981) Lectures on Government and Binding. Dordrecht, Foris.

Duden (1984) Grammatik der deutschen Gegenwartssprache. Edited by G. Drosdowski. Bibliographisches Institut, Mannheim.

Grosz, B. J., Joshi, A. K., & Weinstein, S. (1995). Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, 21(2), 203–225.

Gundel, J., Hedberg, N., & Zacharski, R. (1993) Cognitive Status and the Form of Referring expressions in Discourse. *Language* 69:274-307.

Hobbs, J. R. (1978) Resolving Pronoun References. *Lingua*, Vol. 44, pp. 311-338.

Kaiser, E., Trueswell, J. (2004) The referential properties of Dutch pronouns and demonstratives: Is salience enough? Proceedings of the *Sinn und Bedeutung* 8, Arbeitspapier Nr. 1777, FB Sprachwissenschaft, Universität Konstanz.

Kaiser, E. & Trueswell, J. (2005) Investigating the interpretation of pronouns and demonstratives in Finnish: Going beyond salience. In E. Gibson & N. Pearlmutter (eds), *The processing and acquisition of reference.* Cambridge, Mass., MIT Press

Kameyama, M. (1999) Stressed and Unstressed Pronouns. In P. Bosch & R. van der Sandt (eds) *Focus.* Cambridge University Press.

Kamp, H. (1981) A Theory of Truth and Semantic Representation. In J. Groenendijk, T.M.V. Janssen, & M. Stokhof (eds.) *Formal Methods in the Study of Language*, Dordrecht, Volume 136, 277-322.

Kehler, A. (2002) Coherence, reference, and the theory of grammar. Stanford, CA, CSLI Publications.

Lappin, S., & Leass, H. (1994) An algorithm for pronominal anaphora resolution. *Computational Linguistics*, 20(4), 535–561.

Mitkov, R. (1998) Robust pronoun resolution with limited knowledge. *Proceedings of COLINGACL' 98*. Montreal, Canada.

Preuss, S., Schmitz, B., Hauenschild, Ch., & Umbach, C. (1994) Anaphora Resolution in Machine Translation. In W. Ramm (ed.) *Text and Context in Machine Translation*. Studies in Machine Translation and Natural Language Processing, Vol.6, 29-52.

Winter, Y. (2003) Parameter zur Bestimmung der Referenz von Demonstrativpronomina in Texten . Bachelor thesis, Institute of Cognitive Science, University of Osnabrück, www.cogsci.uni-osnabrueck.de/~CL

Zifonun, G. Hoffmann, L., & Strecker, B. (1997) Grammatik der deutschen Sprache, Berlin, de Gruyter.