Demonstrative Pronouns and Personal Pronouns. German *der* vs. *er*.

Peter Bosch  
Inst. of Cognitive Science  
University of Osnabrück  
49069 Osnabrück, Germany  
pbosch@uos.de

Tom Rozario  
Inst. of Cognitive Science  
University of Osnabrück  
49069 Osnabrück, Germany  
trozario@uos.de

Yufan Zhao  
Inst. of Cognitive Science  
University of Osnabrück  
49069 Osnabrück, Germany  
yzhao@uos.de

Abstract

This paper deals with the problem of determining both the distribution and reference of the German demonstrative pronoun *der* and its case and gender variants. We first show how occurrences of this pronoun can be distinguished from formally largely identical relative pronouns and definite articles. Then we develop the hypothesis that demonstrative and personal pronouns have fundamentally complementary functions: The latter prefer discourse topics as referents, while the former prefer non-topics. This hypothesis is tested against a 350,000 word corpus of written German. The result shows that our hypothesis would correctly describe a dominant tendency in the usage of these two pronoun classes.

1 Introduction

Unlike English, German has demonstrative pronouns that are inflected for gender and can be used also in reference to persons. Here are a couple of examples for a first impression:

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

(2) Paul sah eine Frau hereinkommen. *{Die/Sie}* trug einen schwarzen Mantel.  
Paul saw a woman enter. *She* was wearing a black coat.

In (1) the demonstrative *der* can only refer to Peter, while the personal pronoun *er* can only refer to Paul. In (2) the demonstrative *die* refers to the woman, but it could be substituted for, without noticeable difference, by a personal pronoun *sie*. The particular difficulty with these demonstratives is of several kinds: (a) Their forms are identical with the definite article (see Section 2.1); (b) They are also formally identical with the relative pronoun (see Section 2.2); (c) They overlap in their distribution as well as in their referential options with the personal pronouns (see Section 3), even though, as we shall argue, they function on fundamentally complementary principles.

There are as yet, to the best of our knowledge, no computational accounts of German demonstratives. The good reason for this, apart from the difficulties mentioned under (a)-(c), is certainly the fact that the demonstrative pronouns of the *der* paradigm – quite in contrast with other German demonstrative pronouns like *dieser, jener, derjenige, derselbe*, etc. – are comparatively rare, at least in written German. In a large German newspaper corpus 94.8% of all occurrences of *der* were found to be occurrences of the definite article, a further 4.8% were occurrences of the relative pronoun, and a mere 0.4% were occurrences of the demonstrative pronoun.

So, why would one want a computational identification of demonstratives, and why would one want a better understanding of their reference? On the one hand, there is the motivation from applications: A treatment of demonstratives is clearly needed for machine translation. Nearly no German demonstrative can be adequately translated into English, or most other languages, without reference resolution. Further, even though personal pronouns are much more frequent than demonstratives in written language, this relation is nearly reversed in spoken German. Hence any speech application that requires reference resolution is even more in need of a treatment of demonstratives than of a treatment of personal pronouns.

---

1 These figures are based on a 41 Mio word corpus of *Frankfurter Rundschau*, a German national daily newspaper.

2 The Verbmobil corpus of spoken German (http://verbmobil.dfki.de/) has a ratio of personal pronouns to demonstratives of 1:4, while the Negra corpus of written German yields about 8:1.
On the other hand, independent of applications, we believe that an improved understanding of demonstratives will contribute both to a better understanding of processes of reference in general and to theoretical accounts of information structure and the dynamics of discourse. This paper takes a few modest steps towards an account of the distribution and reference of demonstratives, and what we have to offer is still work in progress. We have not yet looked at all forms of the demonstrative. What we have worked out so far is an hypothesis about the distinction between different uses of the demonstrative and personal pronouns, and we have put this hypothesis to the test with a small 350,000 word corpus of written German.

2 Identifying demonstrative uses of der

The first problem we have to attend to is the identification of those occurrences of the relevant forms that are actually used as demonstratives. The forms of the demonstrative der in German are largely identical with those of the relative pronoun and the definite article. So, how do we identify demonstrative uses?

2.1 Demonstratives and definite articles

Demonstrative uses of der and its morphological variants are formally identical with definite article uses – except for the genitive and the dative plural forms (cf. tables 1 and 2). Their gender and number is determined either by the semantic classification (animacy, sex) of the intended referent or, occasionally, by the gender and number of an antecedent or an appropriate noun that is associated as a default description with the intended referent (cf. Bosch 1987:72-73).

<table>
<thead>
<tr>
<th></th>
<th>sing</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>masc</td>
<td>fem</td>
<td>neut</td>
<td>m/f/n</td>
</tr>
<tr>
<td>nom</td>
<td>der</td>
<td>die</td>
<td>das</td>
<td>die</td>
</tr>
<tr>
<td>gen</td>
<td>dessen</td>
<td>deren</td>
<td>dessen</td>
<td>deren/derer 3</td>
</tr>
<tr>
<td>dat</td>
<td>dem</td>
<td>der</td>
<td>dem</td>
<td>denen</td>
</tr>
<tr>
<td>acc</td>
<td>den</td>
<td>die</td>
<td>das</td>
<td>die</td>
</tr>
</tbody>
</table>

Table 1: Forms of the demonstrative pronoun

Definite determiner occurrences always start an NP, i.e. they are followed by an (attributive) adjective, adverb, numeral, or noun, etc., while demonstrative pronouns are themselves of the category NP. Some unclarity may arise though in cases that could look like cases of elided nouns, as in (3).

(3) Die Gelder sollen nicht aus dem Etat des Umweltministeriums, sondern aus dem des Entwicklungsministeriums fließen.

[The money should not come out of the budget of the Ministry of Environment, but from the budget of the Ministry of Development.]

One might want to argue that dem is really elliptical for dem Etat; and indeed the insertion of Etat would lead to a fully acceptable paraphrase. Still, dem can here only be analysed as a demonstrative pronoun. The evidence comes from a plural variant:

(4) Die Gelder sollen nicht aus den Mitteln des Umweltministeriums, sondern aus {denen/ *den} des Entwicklungsministeriums fließen.

[The money should not come out of the means of the Ministry of Environment, but out of those of the Ministry of Development.]

Here the article form den is ungrammatical and only the demonstrative denen is acceptable. Our claim that dem in the entirely parallel (3) is also a demonstrative would thus be born out.

2.2 Demonstrative and relative pronouns

The distinction between occurrences of the demonstrative pronoun and occurrences of the relative pronoun is slightly trickier, partly because the inventory of forms is exactly the same. In principle there should be a difference in their distribution though, because German relative clauses have sub-clause, i.e. verb-final, word order, while demonstrative pronouns may occur in any NP position. But this won’t help with the distinction, of course, when the entire clause consists of a pronoun plus a finite verb.

3 derer only when immediately followed by a relative clause, or as a relative pronoun when preceded by a preposition. But cf. Baarentzen (1995) for details.
Eventually we believe that demonstrative pronouns must unexceptionally occur referentially and (restrictive) relative pronouns are never referential, but this is not the subject of the present paper. Although, in theory, there is clearly a problem in distinguishing relative and demonstrative pronouns (cf. Gärtner 2001) and there are areas of apparent overlap, we have here decided to take the slightly risky position that all and only those pronouns from the paradigm are counted as relatives that are clearly distinguished as relatives by V-final word order plus those from syntactically ambiguous structures with a pronoun plus a finite verb. All other occurrences are regarded as referential and as demonstratives.

Having clarified which occurrences of the relevant forms actually function as demonstratives, we will now turn to the question of how they relate to personal pronouns.

3 Demonstrative and personal pronouns

There is a considerable overlap in the distribution of demonstrative and personal pronouns, but here at least we have clearly distinct forms. In sentences like the following, as already in example (2), either form seems acceptable and semantic, pragmatic, or even stylistic differences are hard to pinpoint, also for the native speaker.

(5) Am Dienstag, 16. Juni, können dann Falkensteinern ihren Sondermüll von 9 bis 10 Uhr zu dem Wagen bringen. {Er/Der} parkt auf dem Parkplatz beim Bürgerhaus. [On Tuesday, 16 June, the Falkensteiners can take their hazardous waste to the van between 9 and 10 a.m. It is parked in the parking lot at the Civic Centre.]

(6) Zunächst waren die Mietforderungen des Investors (38 Mark pro Quadratmeter) zu hoch. {Er/Der} reduzierte sie dann auf unter 30 Mark. [At first the rent demands of the investor (38 marks per square meter) were too high. Then he reduced them to under 30 Marks.]

The impression that there is an area of overlap where the function of the demonstrative pronoun is nearly indistinguishable from the anaphoric personal pronoun and where the two forms could be substituted for each other salva grammaticalitate and without a clear semantic difference was confirmed in two experiments carried out by our project group (see Cummins e.a. (in preparation)). Self-paced reading as well as native speakers' acceptability preferences yielded no significant differences. The area of functional overlap of personal and demonstrative pronouns may well be an artefact of underspecified data though. If, as we shall argue, the difference between the two types of pronouns is a matter of information structure, then the distinction may not become visible where not enough information about information structure is available – most clearly when we consider only isolated sentences, but often also in sentence pairs.

Still, there are occurrences of the two pronoun types for which the difference is clear. First of all there is a difference with respect to referentiality, already mentioned above: demonstrative pronouns are referential in all their occurrences, but personal pronouns are not; personal pronouns, next to their referential occurrences, also have "bound" or "syntactic" occurrences (cf. Bosch 1983; Reinhart 1983) as in the following sentences

(7) Nobody thought he would make it.
(8) Everyone said he felt sick.

or their German equivalents
(7') Niemand dachte, er würde es schaffen.
(8') Jeder sagte, er fühle sich krank.

The personal pronouns in these sentences are not independently referential items and hence their job obviously cannot be done by (referential) demonstrative pronouns:
(7'') Niemand dachte, der würde es schaffen.
(8'') Jeder sagte, der fühle sich krank.

The occurrences of der here cannot be interpreted the same way as the personal pronouns before, but must refer to some referent introduced elsewhere in discourse. We find the same effect in English with stressed personal pronouns:
(7''') Nobody thought HE would make it.
(8''') Everyone said HE felt sick.

4 obviously so, if the hypothesis is right that relative pronouns have their historic origin in demonstrative pronouns (cf. Lehmann 1984:373-75, Behaghel 1923-32 vol iii:766)

5 This has nothing to do with the fact that the personal pronouns in the above examples are "bound" by "antecedent quantifiers", but is a matter of syntactic-semantic structure; cf. Bosch (1983)
Syntactic occurrences of personal pronouns are in a relation of congruence or agreement with an NP in the same sentence that c-commands them and not in a relation of co-reference. Not all personal pronouns with an explicit antecedent in the same sentence are syntactic pronouns, even though a large proportion is. It is not only syntactic personal pronouns though that differ in their interpretation options from demonstratives. Recall example (1), repeated here:

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

The \textit{er} can only be interpreted with reference to Paul, while \textit{der} must refer to Peter – a difference that is hard to emulate in an English translation. We are reminded though, of differences found in English between stressed and unstressed personal pronouns, first documented by Lakoff (1971) and worked out more recently in the Centering framework by Kameyama (1999). Cf. (9) John hit Bill. Then \textit{he/HE} was injured.

where the unstressed anaphoric pronoun clearly refers to John and the stressed – demonstrative – pronoun \textit{HE} refers to Bill.

If we combine these observations with Bosch’s (1983:203f) Aboutness Principle of Anaphora – which states (roughly) that an anaphor (and that includes all unstressed referential personal pronouns) refers to the object which the clause or sentence is about in which the anaphor occurs, i.e. the current topic, – and Givón's (1983) notion of Topic Continuity we arrive at the following Complementarity Hypothesis

**Complementarity Hypothesis**

Anaphoric personal pronouns prefer referents that are discourse topics, while demonstratives prefer to look for their referents in the complementary domain: among currently non-topical referents.

Referents of demonstratives thus should not only not be current topics, but they should not be topics of the preceding discourse either\(^6\). This gives us two types of referents that are distinguished by their different roles in discourse: On the one hand referents that occurred in previous sentences, but were not topics and hence must be retrieved by the more marked demonstrative form, while the personal pronoun would pick out a topical referent instead. On the other hand there are referents that are referred to in the same sentence or the preceding sentence and may even be topics, but are contrasted with other referents, so that the marked demonstrative form signals contrast (cf. again Bosch 1988 for marked personal pronouns).

Clearly, this is only a working hypothesis that still needs further elaboration. But it is sufficient for a first validation with respect to corpus data, provided we can make it operational.

We decided to assume that noun phrases that occur in the nominative are thereby \textit{ceteris paribus} very likely to establish their referent as a default topic for the following sentence, and that noun phrases that are not in the nominative are much less likely to establish topics for the following sentence.

With antecedents that occur in the same sentence as the pronoun we get a slightly more complicated situation. As for the personal pronouns, the relation may not be one of coreference but of a syntactic nature, as mentioned above. Demonstratives could not function the same way and since these non-referential pronouns tell us nothing about referential options of pronouns there is no point comparing them to demonstratives, which are always referential. Since a large proportion of personal pronouns with an antecedent in the same sentence are of this kind, we will exclude these occurrences from our considerations. Demonstratives that have a coreferential NP in the same sentence, on the other hand, are very frequently used contrastively. A typical example is (3) above, but also the following:

(10) ... Überschuss... Der der Ausfuhren übertraf \textit{den} des Imports.  
[... surplus... That of the exports exceeded that of the import.]

Note that this function is not open to personal pronouns. Hence we may do well to exclude sentence-internal uses of personal and demonstrative uses from our comparison altogether: In those sentence-internal cases where both demonstratives and personal pronouns are grammatically admissible, they don't usually show a difference.

\(^6\) Similar considerations were applied in Bosch (1988) to intonationally marked, i.e. stressed, personal pronouns, with the resulting conclusion that the intonational markedness signals the activation of processes of referent search and referent selection.
in function. – This is what we already found in our questionnaire survey mentioned above.

4 Empirical results

We compared the occurrences of the demonstrative with those of the personal pronoun forms in the NEGRA⁷ corpus of written German, with regard to their frequency and with regard to the classification of their antecedents.

4.1 Frequency results

The first observation concerns relative frequency. Demonstratives are much rarer than personal pronouns: We counted 1436 instances of personal pronouns and only 180 demonstratives. Perhaps this proportion cannot straightforwardly be generalized but is typical rather for written discourses and those that typically show topic continuity, as newspaper articles do. We find an entirely different picture in the spoken Verbmobil corpus that has fairly short dialogues in which dates, times and places of business meetings are negotiated between two partners. Here we find nearly the opposite proportion: 565 demonstrative pronouns and a mere 157 personal pronouns.

4.2 Preferred antecedents

The main result we can report concerns anaphor-antecedent relations for demonstratives and personal pronouns. The relevant figures for the NEGRA corpus are given in tables 3 and 4, using the break-up discussed in Section 3 above. The first difference between personal and demonstrative pronouns can be noted immediately: The antecedent of a demonstrative in over 90 % of all cases is found in the same or the preceding sentence, while personal pronouns refer to objects that were introduced earlier in discourse about twice as often (17.7% vs. 8.9%). This figure on its own already gives initial support to the hypothesis that personal pronouns more likely than demonstratives pick up referents that have the status of a discourse topic.

Tables 3 and 4 also show that there is an about equal proportion of occurrences for both pronoun classes that have a sentence-internal antecedent. And we can also clearly see where antecedents are preferentially located: for personal pronouns in nominative NPs in the preceding sentence and for demonstratives in non-nominatives. This would confirm our hypothesis regarding the complementarity of personal and demonstrative pronouns.

<table>
<thead>
<tr>
<th>antecedent position</th>
<th>preceding sentence</th>
<th>earlier in discourse</th>
<th>same sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominative</td>
<td>Non-nominative</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14.5%</td>
<td>46.7%</td>
<td>8.9%</td>
<td>30.0%</td>
</tr>
</tbody>
</table>

Table 3 Position of antecedents of demonstratives

<table>
<thead>
<tr>
<th>antecedent position</th>
<th>preceding sentence</th>
<th>earlier in discourse</th>
<th>same sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominative</td>
<td>Non-nominative</td>
<td></td>
<td></td>
</tr>
<tr>
<td>48.0%</td>
<td>7.3%</td>
<td>17.7%</td>
<td>27.2%</td>
</tr>
</tbody>
</table>

Table 4 Position of antecedents of personal pronouns

As we argued in Section 3, the figures in tables 3 and 4 are not ideal for a comparison, because sentence-internal personal pronouns are largely non-referential and demonstratives are largely contrastive – functions that are inaccessible to the demonstratives and personal pronoun forms respectively – and antecedents earlier in discourse are much less common for demonstratives. The comparison of referential options had therefore better focus on the area where personal pronouns and demonstratives actually compete: when they have antecedents in the previous sentence. If we discount antecedents earlier in discourse and in the same sentence and take the uses with an antecedent in the previous sentence as 100%, the figures look as in tables 5 and 6.

<table>
<thead>
<tr>
<th>antecedent in preceding sentence</th>
<th>nominative</th>
<th>non-nominative</th>
</tr>
</thead>
<tbody>
<tr>
<td>23.6%</td>
<td>76.4%</td>
<td></td>
</tr>
</tbody>
</table>

Table 5 Classification of antecedents of demonstrative pronouns when in previous sentence

⁷ NEGRA is a POS-tagged and syntactically annotated corpus of German of 355,000 words, a subset of the Frankfurter Rundschau corpus, which is also available as a tree bank. See http://www.coli.uni-sb.de/sfb378/negra-corpus/negra-corpus.html.
<table>
<thead>
<tr>
<th>antecedent in preceding sentence</th>
<th>nominative</th>
<th>non-nominative</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>86.7%</td>
<td>13.2%</td>
</tr>
</tbody>
</table>

Table 6 Classification of antecedents of personal pronouns when in previous sentence

These tables show the complementary function of personal and demonstrative pronouns clearer than tables 3 and 4. Personal pronouns have antecedents in the nominative in the overwhelming majority of cases, while demonstratives rather look for non-nominatives. – If our assumption that the nominative signals topicality of the referent of an NP is right and non-nominatives signal non-topicality, then these results have a clear theoretical interpretation that follows our discussion in Section 3. The complementarity is not perfect, however. We may assume that for demonstratives some of the references to topics of the preceding sentence may be of a contrastive nature. Since this option is not open to unstressed personal pronouns, this hypothesis may partly explain the imbalance. Also, there are factors that would make personal pronouns prefer non-topics as referents, e.g. strategies of parallel interpretation. – But such further ideas still need investigation.

5 Conclusion and further work

From the data presented it would appear that our Complementarity Hypothesis is well supported: while personal pronouns prefer topic referents, demonstrative pronouns prefer referents that are not topical. Demonstratives thus behave like German (or English) stressed personal pronouns. To what extent we can also predict for demonstratives the full search and selection functionality of marked German personal pronouns (cf. Bosch 1988) is as yet unclear. More theoretical and empirical work is required.

Further work, part of it already in progress, will include a psycholinguistic study of referent accessibility by demonstratives and, finally, work on reference resolution algorithms for demonstratives.

Acknowledgements

The work reported here is being carried out in a study project of the Cognitive Science Master Programme at the University of Osnabrück. Project participants are Phil Cummins, Boris Gutbrod, Kyoung-ho Park, Tom Rozario, Yufan Zhao and as staff advisers Peter Bosch, Graham Katz, Carla Umbach, and Jaqueline Griego. The authors gratefully acknowledge the contributions of all participants and advisers.

References