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 distribution and reference of the German demonstrative pronoun *der* and its case and gender variants, in comparison with the behaviour of personal pronouns. After introducing the relevant facts about both pronoun types, we develop our *Complementarity Hypothesis*, which claims that demonstratives differ from personal pronouns mainly in that they prefer less salient referents. The hypothesis is tested against a newspaper corpus and is found to be fundamentally correct. The paper concludes with the discussion of the empirical results and their effect for future work.

1 Introduction

Unlike English, German has demonstrative pronouns that are inflected for number, gender, and case and can refer also to persons as well as other individuals, very much like personal pronouns. Here are some examples:

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

(2) Paul sah eine Frau hereinkommen. {*Sie/Die*} 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) both the personal pronoun *sie* and the demonstrative *die* can be used equally well to refer to the woman.

There is, to the best of our knowledge, no computational nor any explicit formal linguistic account of German demonstratives. – One obstacle is that the same forms are also used as definite articles and relative pronouns, which makes it hard at times to single out demonstrative uses automatically. Some part of speech taggers, even very good ones, sytematically go wrong on demonstratives[1]. A further problem for a systematic account is that demonstratives, as we already saw in (2), overlap in their distribution as well as in their referential options with the personal pronouns and the distinction occasionally looks like no more than a stylistic one.

The main reason that no computational account has been attempted, however, is probably that there seems no urgent need for it: Demonstrative pronouns of the *der* paradigm – different from other German demonstrative pronouns like *dieser, jener, derjenige, derselbe,* etc. – are comparatively rare. 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[2]. Because of its clear deictic potential, similar to English *this* or *that,* one might think that the demonstrative would be more frequent in spoken language. But an unbiased comparison would require written and spoken corpora of the same discourse type, which

---

[1] An example is the tree tagger by Helmut Schmid (1994).
[2] These figures are based on a 41 Mio word corpus of *Frankfurter Rundschau,* a German national daily newspaper.

we have not been able to acquire. So we have to leave this point to speculation[3].

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 of course the motivation from applications, similar as for all other work on reference resolution: A treatment of demonstratives is clearly needed, e.g., for machine translation. Nearly no German demonstrative can be adequately translated into English, or most other languages, without reference resolution.

On the other hand, independent of applications, we believe that an improved understanding of demonstratives may contribute both to a better understanding of processes of pronominal reference and to theoretical accounts of information structure and the dynamics of discourse.

The interesting point about German demonstratives in this connection is that they allow for a fairly direct comparison of pronominal anaphoric and demonstrative reference in a way this comparison is not possible for other languages. In English we could look at differences between *it, this,* and *that*. But here the range of potential referents brings in a number of additional complications (cf. Passonneau 1989, Webber 1991, Eckert & Strube 2000) that are avoided when we compare personal and demonstrative pronouns in German.

This paper takes a few modest steps towards an account of the distribution and reference of demonstratives, but 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 Demonstrative and personal pronouns

At a first glance there is considerable overlap in the distribution of demonstrative and personal pronouns. In many contexts, as already in example (2), either form seems acceptable and semantic, pragmatic, or even stylistic differences are hard to pinpoint. This is also true in the following cases:

(3) Am Dienstag, 16. Juni, können dann Falkensteiner 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.]

(4) 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 indistinguishable from the anaphoric personal pronoun and where the two forms could be substituted for each other without a clear grammatical or semantic difference was confirmed in two psycholinguistic experiments carried out by our project group (see Cummins et al. (in preparation)). Self-paced reading as well as native speakers' acceptability preferences yielded no significant difference.

This 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 show up when not enough information structure is visible – most clearly when we consider odd sentences in isolation, but often also in sentence pairs.

Still, there are occurrences of the two pronoun types for which the difference is quite clear as already in (1), repeated here:

---

[3] 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. However, this comparison remains inconclusive: the Verbmobil corpus consists of dialogues negotiating the time and place of appointments, while the Negra corpus is a regular daily newspaper corpus with obviously vastly different discourse structure.

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

The *er* can only be interpreted with reference to Paul, while *der* must refer to Peter – a difference that is hard to emulate in an English translation. But also a difference in grammaticality can be found. In (5) only the demonstrative is allowed while the personal pronoun would be ungrammatical:

(5) Die Gelder sollen nicht aus dem Etat des Umweltministeriums, sondern aus {*dem*/ **ihm*} 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.]

The way that demonstratives differ from personal pronouns may in some respects be reminiscent of differences found in English between stressed and unstressed personal pronouns, as documented by Akmajian and Jackendoff (1970). Cf.[4]

(6) When {*he*/*HE*} came home, John was tired.

where the unstressed anaphoric pronoun clearly refers to John and the stressed *HE* cannot be so interpreted. In addition we have cases like those brought up by Lakoff (1971):

(7) John called Fred a Republican. Then he inSULted him.

(7') John called Fred a Republican. Then HE insulted HIM.

Also here the referential options of the pronouns change when the pronouns are accentuated. Observations like these suggest something like a complementarity between stressed and unstressed personal pronouns in English as was claimed in Kameyama's (1999) Complementary Preference Hypothesis: "A focused pronoun [i.e. stressed personal pronoun] takes the complemen-

tary preference of the unstressed counterpart." – where the referents of both stressed and unstressed pronouns come from the current focus of attention of the discourse, the Forward-looking Centre of the preceding utterance, in terms of Centering Theory (Grosz e.a. 1995).

But the distinction between German personal and demonstrative pronouns is not the same. The demonstrative in (1) does not translate into English as a stressed personal pronoun: While *der* in (1) can only refer to Peter, *HE* in (1') is just oddly ambiguous:

(1') Paul wanted to go running with Peter. But *HE* had a cold.

Kameyama (1999) is probably right characterising the function of English stressed personal pronouns in terms of an Alternative Semantics approach to focus (Rooth 1992), and hence in terms of contrast, and an analysis of stressed German and English personal pronouns in terms of contrast and markedness was also put forward in Bosch (1988). The difference between personal pronouns and demonstratives in German, however, cannot be captured along these lines: Even though complementarity plays a role, the category of contrast is here absent.

The basic regularity that we find in the behaviour of demonstratives is that they

i. choose from several antecedents (that obey the relevant number and gender constraints as well as subcategorisation constraints of the relevant predicate) the one whose referent is preferentially *not* at the top salience rank (not a topic); Cf. (1), (3), (4).

ii. pick up a unique antecedent (that obeys the relevant number and gender constraints as well as subcategorisation constraints of the relevant predicate), even though it stands for the most salient referent (the topic). They may also accommodate a generic referent that obeys the subcategorisation constraints of the relevant predicate (cf. (8) below). In this condition their referential options are identical to those of referential anaphoric personal pronouns, except for a certain stylistic markedness; cf. (2).

iii. when accompanied by a pointing gesture they refer to a unique referent that was not salient before.

---

[4] Stressed syllables are marked typographically by capitals.

Condition (iii), even though without direct relevance for written language, probably provides the clearest intuitive understanding for the function of demonstratives: *They highlight a previously non-salient referent*, and in a sense they presuppose the non-salience of their referents. – If their reference defaults to salient discourse entities, as under condition (ii), when no other suitable discourse referents are available, we observe an effect of stylistic markedness: The referent is represented, as it were, as something put at a distance. A similar effect is seen in English when, e.g., a salient female person is referred to as "that woman" or the current head of government as "that prime minister".

For the purposes of this paper we summarize this hypothesis about the function of demonstratives in the following form:

Complementarity Hypothesis
Anaphoric personal pronouns prefer referents that are established as discourse topics, while demonstratives prefer non-topical referents.

Clearly, this is a working hypothesis that needs further elaboration, in particular with regard to the notions of salience and topic. But it is sufficient for a first validation with respect to corpus data, provided we can make it operational.

In line with most research in the area, guided in particular by the Keenan and Comrie's (1977) Accessibility Hierarchy, Givòn's (1984:139) Topicality Hierarchy, and Centering Theory (Grosz e.a. 1995), we decided that noun phrases that occur in the nominative are thereby *ceteris paribus* very likely to establish their referent as a topic for the following sentence, and that noun phrases that are not in the nominative are less likely to establish topics for the following sentence. We disregard in this study all other parameters that may influence salience or topichood, such as definiteness, referent type, voice, or embeddedness.

Before we can check our hypothesis on corpus data, however, we have to deal with a number of smaller problems related to the identification of demonstratives.

3 Identifying demonstrative uses of *der*

The first problem 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?

3.1 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 by the gender and number of a (virtual or actual) antecedent (cf. Bosch 1987:72-73).

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 (5) above, repeated here.

	sing			plur
	masc	**fem**	**neut**	**m/f/n**
nom	*der*	*die*	*das*	*die*
gen	*dessen*	*deren*	*dessen*	*deren/ derer*[5]
dat	*dem*	*der*	*dem*	*denen*
acc	*den*	*die*	*das*	*die*

Table 1: Forms of the demonstrative pronoun

	sing			plur
	masc	**fem**	**neut**	**m/f/n**
nom	*der*	*die*	*das*	*die*
gen	***des***	***der***	***des***	***der***
dat	*dem*	*der*	*dem*	***den***
acc	*den*	*die*	*das*	*die*

Table 2: Forms of the definite article

---

[5] *derer* only when immediately followed by a relative clause, or as a relative pronoun when preceded by a preposition. But cf. Bærentzen (1995) for details.

(5) 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:

(5') 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 (5) is also a demonstrative is thus born out.

3.2 Demonstrative and relative pronouns

The distinction between occurrences of the demonstrative and occurrences of the relative pronoun is trickier 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. Unfortunately, there still remains an area of overlap that is controversial on theoretic grounds (cf. Gärtner 2001). We have here decided to count all and only those pronouns from the paradigm as relatives that are unambiguously 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 demonstratives.

4 Empirical results

We compared the occurrences of the demonstrative with those of the personal pronoun forms in the NEGRA[6] 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 here 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 in any case for written discourses and those that typically show topic continuity, as newspaper articles do.

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 2 above[7].

antecedent position			
preceding sentence		earlier in discourse	same sentence
nominative	non-nominative		
14.5%	46.7%	8.9%	30.0%

Table 3 Position of antecedents of demonstratives

antecedent position			
preceding sentence		earlier in discourse	same sentence
nominative	non-nominative		
48.0%	7.3%	17.7%	27.2%

Table 4 Position of antecedents of personal pronouns

The possibly most striking difference between personal and demonstrative pronouns is that the antecedent of a demonstrative in over 90 % of all cases is found in the same or the preceding sen-

---

tence[8], while personal pronouns refer to objects that are 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 topic: Topics are usually maintained over more than one sentence.

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.

The figures in tables 3 and 4 are not ideal for a comparison though. For two reasons: (a) they include demonstratives that have their antecedents neither in the same sentence nor in the preceding sentence, and (b) they include personal pronouns that have their antecedents in the same sentence. – The first case poses a problem with regard to antecedents further back in discourse, of which there are practically none. The figure almost exclusively rests on cases where actually no antecedent could not be found in the last three preceding sentences; and these are demonstratives that accommodate their referents. An typical example is the following:

(8) Jetzt, wo hier auf einmal an jedem Feldweg jemand Blumen verkauft, wollen *die* uns vertreiben.
[Now, where there is suddenly someone selling flowers along each country lane, *they* [the authorities] want to chase us away.]

Although this kind of accommodating use is possible also for personal pronoun forms, the dominant case among pronouns referring to an antecedent "earlier in discourse" is just straightforward anaphora. But this means that the parameters governing the class of references to antecedents "earlier in discourse" are quite dif-

---

[8] We must add that we counted cases of accommodated referents and occurrences for which no antecedent could be recovered among those where the antecedent was not in the same or the preceding sentence.

ferent for personal and demonstrative pronouns. Hence this class should be discarded for a quantitative comparison and should be treated separately.

The other set of figures we should ignore in a quantitative comparison are those that have to do with sentence-internal pronouns. Here the point is that personal pronouns frequently do not function referentially when their antecedent is in the same sentence, while demonstrative pronouns are referential in all their occurrences. Personal pronouns, next to their referential occurrences, also have "bound" or "syntactic" occurrences (cf. Reinhart 1983; Bosch 1983) as in the following sentences

(9)   Nobody thought *he* would make it.

(10)   Fred felt sick, when *he* returned.

or their German equivalents

(9')   Niemand dachte, *er* würde es schaffen.

(10')   Fred fühlte sich krank, als *er* zurückkam.

The personal pronouns in these sentences (on the intended interpretation) are not independently referential items and hence their job obviously cannot be done by (referential) demonstrative pronouns:

(9")   Niemand dachte, *der* würde es schaffen.

(10")   Fred fühlte sich krank, als *der* zurückkam.

The occurrences of *der* here cannot be interpreted the same way as the personal pronouns before, but must refer to a person introduced elsewhere in discourse. We find a similar effect in English with stressed personal pronouns:

(9''')   Nobody thought *HE* would make it.

(10''') Fred felt sick, when *HE* returned.

Syntactic occurrences of personal pronouns are in a relation of congruence or agreement with an NP that c-commands them and not in a relation of co-reference (cf. Bosch 1983). Although

not all personal pronouns with an explicit antecedent in the same sentence are syntactic pronouns, a large proportion is. And since this mode of operation is unavailable for demonstratives, we should not include sentence-internal pronouns when we are interested in comparing options for referent choice and had 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.

antecedent in preceding sentence	
nominative	non-nominative
23.6%	76.4%

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

antecedent in preceding sentence	
nominative	non-nominative
86.7%	13.2%

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 is right that the nominative signals topicality of the referent of an NP, or a high rank in the attentional structure of discourse, and non-nominatives signal non-topicality, or a lower rank in the attentional structure, then these results have a clear theoretical interpretation that follows our discussion in Section 2.

The complementarity documented in tables 5 and 6 is not perfect, however. What are the disturbing factors? First, we should recall that we used a pretty coarse operationalization of the notions of salience and topic that ignores a number of potentially relevant parameters, such as definiteness, voice, embeddedness, or reference type. Then, as already mentioned above, there is the use under which a demonstrative refers to the topic of the preceding sentence, which is accompanied by a certain stylistic effect. These are cases where there is only one grammatically suitable antecedent. Thirdly, there are factors that would make personal pronouns prefer non-topics as referents, e.g. strategies of parallel interpretation. And, finally, although this is comparatively rare in newspaper corpora, we find cases in which world knowledge simply dictates an interpretation that runs counter to what the grammar of the pronouns make us expect, i.e. badly written text. – All of these factors require further 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. This raises a number of issues that need further investigation. Among them the question about the interaction of the case parameter for salience with other parameters. This issue is investigated in psycholinguistic experiments we are now running. Further there is a great need to study the intrasentential division of labour between personal and demonstrative pronouns, which involves, in particular, closer attention to the relation between relatives and demonstratives. A large area of investigation, for which we do not currently have suitable corpora, is the question of how the regularity we have found for newspaper texts carry over to other discourse types, in particular spoken discourse. In this context also the parameter of stress needs further research, which plays (nearly) no role in written text, but influences reference options for German demonstratives just as it does for personal pronouns in English and German. – Another, and somewhat different, line of development is the application of results from this line of research to computational reference resolution. We believe that the contrastive investigation of reference options of different lexical forms, as here demonstratives vs personal pronouns, may be particularly useful in developing further constraints also for the reference resolution of the vastly more popular personal pronouns, even when reference resolution for

demonstratives themselves is not seen as representing a particularly urgent task.

Acknowledgements

References

Adrian Akmajian & Ray Jackendoff. 1970. Co-referentiality and stress. *Linguistic Inquiry,* 1.

Per Bærentzen. 1995. Zum Gebrauch der Pronominalformen 'deren' und 'derer' im heutigen Deutsch. *Beiträge zur Geschichte der deutschen Sprache und Literatur,* 117: 299-217.

Peter Bosch. 1983. *Agreement and Anaphora.* Academic Press, London.

Peter Bosch. 1987. Pronouns under control. *Journal of Semantics,* 5:65-78.

Peter Bosch. 1988. Representing and accessing focussed referents. *Language and Cognitive Processes,* 3: 207-231.

Philip Cummins et al. in preparation. *German demonstratives.* Institute of Cognitive Science, University of Osnabrück.

Miriam Eckert & Michael Strube. 2000. Dialogue acts, synchronizing units, and anaphora resolution. *Journal of Semantics* 17:51-89

Hans-Martin Gärtner. 2001. Are there V2 relative clauses in German? *Journal of Comparative Germanic Linguistics,* 3: 97-141.

Talmy Givòn. Syntax. A Functional-Typological Introduction. 1984. Benjamins, Amsterdam.

Barbara Grosz, Aravind Joshi, Scott Weinstein. 1995. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics* 21:67-90

Megumi Kameyama. 1999. Stressed and unstressed pronouns. In: Peter Bosch & Rob van der Sandt. eds..: *Focus.* CUP, Cambridge. 306-321.

Edward Keenan and Bernard Comrie. 1977. Noun phrase accessibilty and universal grammar. *Linguistic Inquiry*, 8:63-99.

George Lakoff (1971) Presupposition and relative well-formedness. In: Danny Steinberg & Jakobovits (eds.): *Semantics.* CUP. Cambridge. 329-344.

Tanya Reinhart. 1983. *Anaphora and semantic interpretation.* Croom Helm, London.

Mats Rooth. 1992. A theory of focus interpretation. *Natural Language Semantics* 1:75-116.

Helmut Schmid. 1994. Probabilistic Part-of-Speech Tagging Using Decision Trees. *Proceedings of International Conference on New Methods in Language Processing.* September 1994

Bonnie Lynn Webber. 1991. Structure and ostension in the interpretation of discourse deixis. *Language and Cognitive Processes* 6:107-135.