

Reference Determination for Demonstrative Pronouns

Peter Bosch & Carla Umbach

University of Osnabrück, Germany

This paper discusses results from a corpus study of German demonstrative and personal pronouns and from a reading time experiment in which we compared the interpretation options of the two types of pronouns (Bosch et al. 2003, 2007). A careful review of exceptions to a generalisation we had been suggesting in those papers (the Subject Hypothesis: "Personal pronouns prefer subject antecedents and demonstratives prefer non-subject antecedents") shows that, although this generalisation correctly describes a tendency in the data, it is quite wrong in claiming that the grammatical role of antecedents is the relevant parameter. In the current paper we argue that the generalisation should be formulated in terms of *information-structural properties of referents* rather than in terms of the *grammatical role of antecedent expressions*.

1 Introduction

Personal pronouns in English like in many other languages are notoriously ambiguous.¹ In the following sequence, for instance, *he* may equally well refer to Peter or to Paul.

- (1) Peter_i wanted to play tennis with Paul_k. But he_{i,k} was sick.

Dutch and German have an additional repertory of demonstrative pronoun forms that can sometimes reduce the ambiguity in such cases: If we translate (1) as in (2a) or (3a) the clearly preferred interpretation of the personal pronoun *er/hij* would be Peter; and if we translate (1) as in (2b) or (3b), using the demonstrative *der/die*, Paul is indeed the *only* possible referent.

¹ We are grateful to Graham Katz, Phil Cummins, Boris Gutbrod, Kyoung-ho Park, Tom Rozario, and Yufan Zhao who participated to an important extent in the research discussed in this paper.

- (2) a. Peter_i wollte mit Paul_k Tennis spielen. Doch er_{i,k} war krank.
b. Peter_i wollte mit Paul_k Tennis spielen. Doch der_k war krank.
- (3) a. Peter_i wilde met Paul_k gaan tennissen. Maar hij_{i,k} was ziek.
b. Peter_i wilde met Paul_k gaan tennissen. Maar die_k was ziek.

No similar contrast is found among English pronoun forms, at least not with pronouns referring to persons. But even though the difference between the interpretations of the Dutch and German personal and demonstrative pronouns is clear in the case at hand, it has proven difficult to get a clear picture of what exactly the difference in the constraints on the interpretation of demonstrative and personal pronouns is (cf. Comrie 1997, Zifonun et al. 1997, Abraham 2002, Kaiser & Trueswell 2004).

We carried out a number of corpus studies and a reading time experiment in order to learn more about this difference specifically in German (Bosch et al. 2003, 2007). In the current paper we will be fairly brief about these empirical studies and instead will re-discuss the conclusions to which we were lead in those papers.

2 Subject vs. Non-Subject Antecedents

2.1 *Corpus frequency*

An early result from our corpus studies (Bosch et al. 2003) was that personal pronouns in the large majority of cases had antecedents that were in the grammatical role of the subject (operationally the nominative NP) in the preceding sentence, while demonstratives in the majority of cases had non-subjects (non-nominatives) as their antecedents. This result is seen from Figure 1.

2.2 *Self-paced reading*

2.2.1 *The experiment*

This observation of different frequencies in the referent choice of the two pronoun forms is a good starting point and seems to indicate a clear tendency; it also fits in well with our intuitive judgements about the interpretation of the discourses in (2). But what causes the observed differences in frequencies?

First of all we wanted to know how strong these preferences in referent choice are. How do they behave in relation to preferences caused by mundane plausibility? Suppose you know already, when you read the discourses in (2), that either Peter or Paul is a notorious malingerer. Would that change the preferred interpretation?

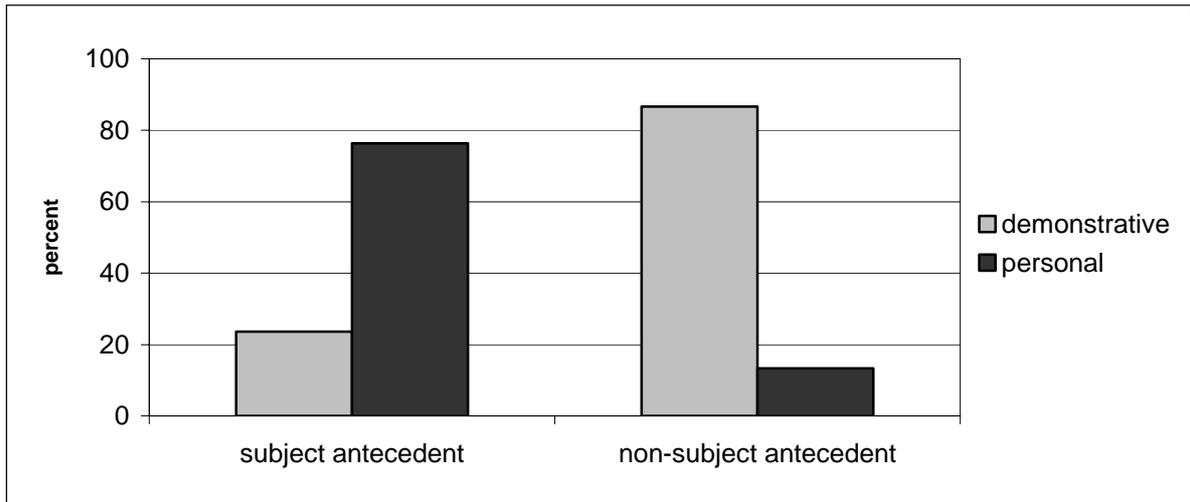


Figure 1: Distribution of antecedents of personal and demonstrative pronouns in preceding sentence (NEGRA Corpus, 355.000 words, 1436 personal pronouns, 180 demonstratives)

And, perhaps more interestingly, is the *grammatical role* of the antecedent indeed the relevant parameter? Could it not be that personal pronouns simply prefer referents that are mentioned first and demonstratives those that are mentioned later in the preceding sentence? Our corpus study could of course not tell the difference, because subjects, due to preferred SVO surface word order in German, just happen to occur most frequently sentence-initially.

headline		
(i)	Im Krankenhaus	[In hospital]
context sentence		
(ii)	(a)	Der Chefarzt untersucht den Patienten. (SVO)
	(b)	Den Patienten untersuchte der Chefarzt. (OVS)
		[The head doctor is examining the patient.]
target sentence		
(iii)	(a)	{ <u>Er</u> / <u>Der</u> } ist nämlich Herzspezialist. subject bias
		[He is a heart specialist.]
	(b)	{ <u>Er</u> / <u>Der</u> } muß sofort operiert werden. object bias
		[He must be operated on immediately.]
	(c)	{ <u>Er</u> / <u>Der</u> } ist gerade erst gekommen. unbiased
		[He has just arrived.]
completion task		
(iv)	(a)	Der _____ ist nämlich Herzspezialist.
	(b)	Der _____ muß sofort operiert werden.
	(c)	Der _____ ist gerade erst gekommen.

Box 1: Stimulus materials for self-paced reading

We investigated the two questions in a reading time study plus completion task with stimulus sets as in Box 1. Each stimulus set started with a lead-in headline (i) and was followed by a context sentence introducing two protagonists (ii). This sentence came in two versions: One with SVO word order and the other with the slightly marked OVS order. The target sentence followed in the third position (iii) and was varied for plausibility of content with a "subject bias", meaning that the content is biased for an interpretation that takes the subject of the preceding sentence as the antecedent of the pronoun, an "object bias" where plausibility prefers the object of the preceding sentence as the pronoun's antecedent, and an "unbiased" variant. Each of the target sentence variants was presented either with a demonstrative or a personal pronoun form.

These materials were presented on a computer screen word by word, the next word appearing when the subject pressed a button. Time lags between button presses were measured ("reading times" for the relevant words).

After each such story subjects were presented with a completion task in the form of a gap sentence, as in (iv) in Box 1, and were asked to type in the noun (e.g., "Chefarzt" or "Patient" in the (a) version) that made the sentence coherent with the preceding story. The (a) versions of sentence (iii) were followed by (a) versions of the comprehension question, and the (b) and (c) versions correspondingly by (b) and (c) versions. The sentences in (iv) are identical to those in (iii) except for the fact that we have definite NPs with a missing noun in (iv) where we have a pronoun in (iii).

The prediction was that we would find longer reading times where there is a conflict between mundane plausibility and linguistic preference of the pronoun interpretation. E.g., we expected that in a sequence like

- (ii) Der Chefarzt untersucht den Patienten.
- (iii) {Er/Der} ist nämlich Herzspezialist.
[The head doctor is examining the patient.
{PPro/DPro} is a heart specialist.]

the target sentence (iii) would take longer to read in the second variant (with a demonstrative pronoun), where plausibility requires the demonstrative to refer to the head doctor while the linguistic preference of the demonstrative would be for referent of the non-subject antecedents, i.e. the patient.

In the comprehension test we expected that more errors would occur in the case of a conflict between the linguistic preference of the pronoun and mundane plausibility, i.e., that the pronoun's referential preference would cause more answers deviating from the answer that would be plausible in the context of our story. For the unbiased cases we expected that the preference of the pronoun type would be straightforwardly reflected in the responses.

Finally, as for the SOV vs. OVS variation we had no predictions, because we simply had no theory that would prefer either the parameter grammatical role or the parameter order of mention.

2.2.2 The results

The reading time prediction came out significant for both SVO and OVS word order in the conditions where the plausibility brought in a subject bias. Here the demonstrative variants took significantly longer to read than the personal pronoun variants.

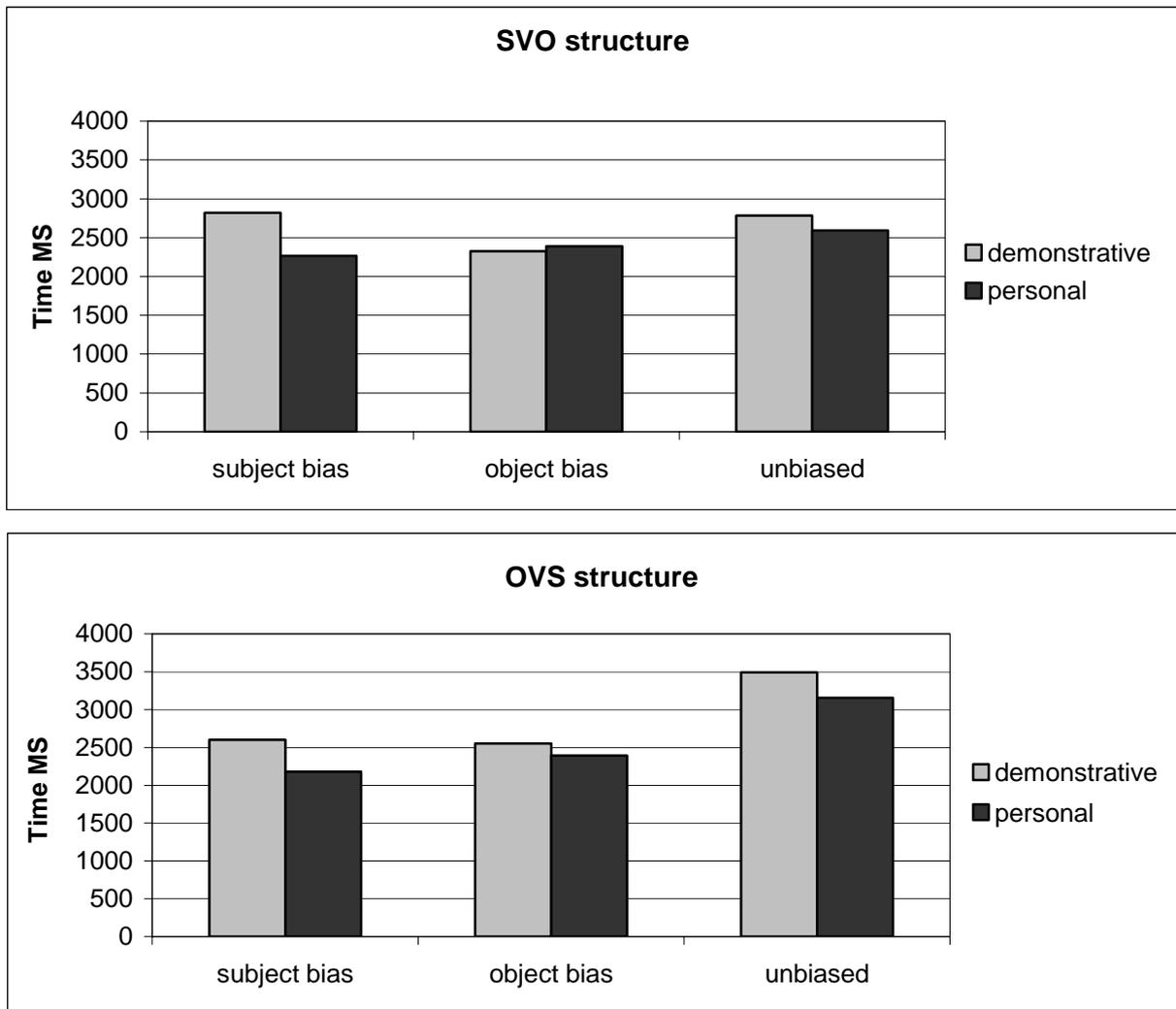


Figure 2: Reading times for target sentence when context sentence word order is either SVO or OVS. The difference between demonstratives and personal pronouns was significant only in the subject bias condition, both for SVO and OVS word order.

There was no significant result though for the conditions where the mundane plausibility of the story suggested a preference for the object antecedent, neither in the SVO nor the OVS variant; nor did the reading times indicate any signifi-

cant preferences in the unbiased conditions. These results are summarised in Fig. 2.

The results for the completion task showed significantly more errors where we had a subject bias in the story: The demonstrative caused more errors than the personal pronoun. However this result became significant only for SVO word order, not for OVS. Nor did we have significant results for the object bias conditions, in neither word order condition. The prediction for the unbiased cases was significant in both SVO and OVS, but only for the demonstrative pronoun.

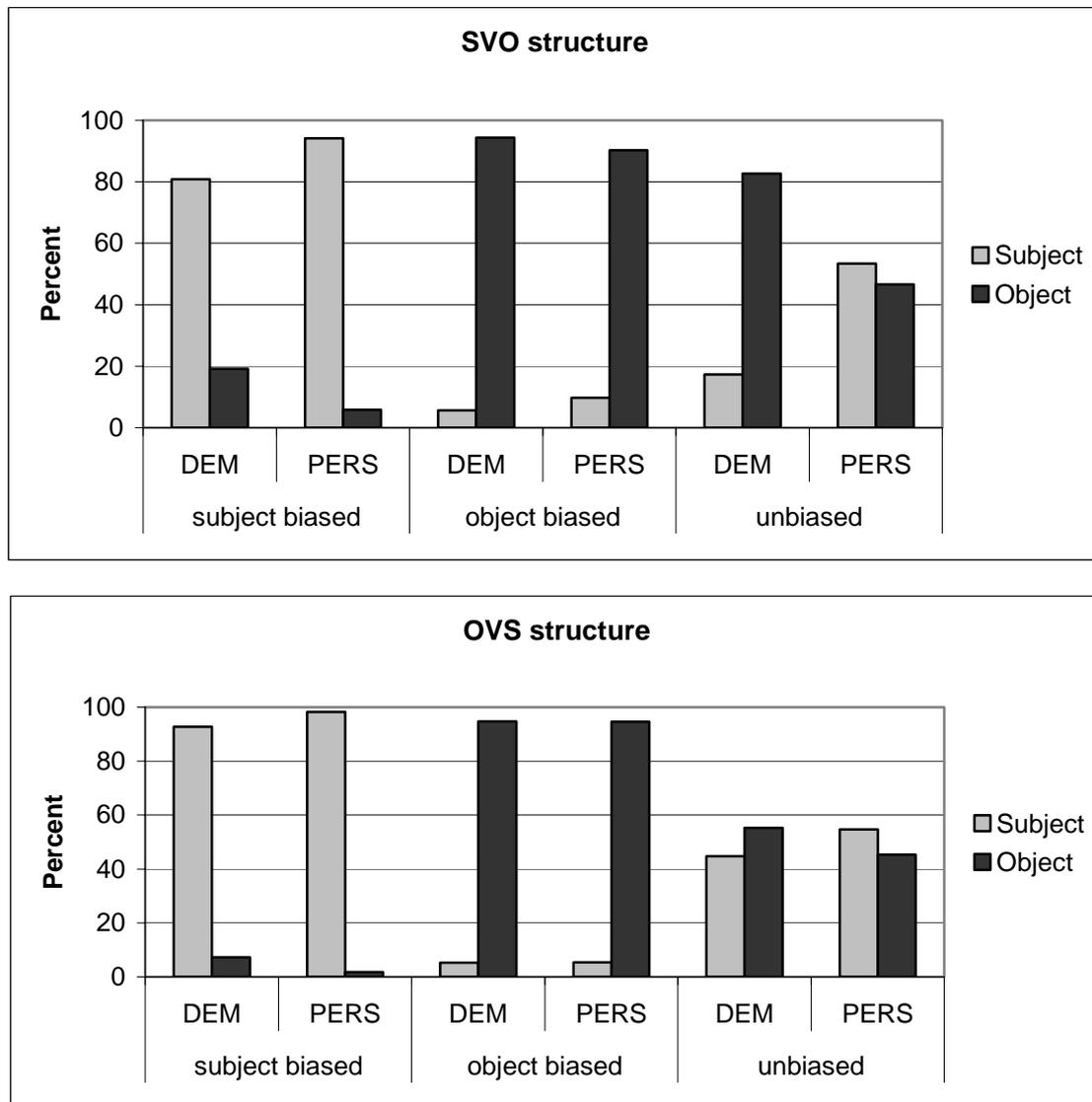


Figure 3: Completion task. Choice of referent for the target pronoun (either demonstrative or personal) is either the referent of the subject or the object expression of the context sentence.

This makes it look as if our predictions were basically on the right track, and that the relevant antecedent parameter would be rather *grammatical role* than

order of mention. – None of the significant results are in contradiction with this. Unfortunately though, as already mentioned, the predications did not come out significant in *all* the conditions. Here's a list of the significant and (in square brackets) the non-significant results.

reading time:

der + subject bias > er + subject bias	in SVO & OVS
[er + object bias > der + object bias	in SVO & OVS]
[der + unbiased: preference for the object	in SVO & OVS]
[er + unbiased: preference for the subject	in SVO & OVS]

completion task errors:

der + subject bias > er + subject bias	in SVO [&OVS]
[er + object bias > der + object bias	in SVO & OVS]

completion task preferences:

der + unbiased: preference for the object	in SVO & OVS
[er + unbiased: preference for the subject	in SVO & OVS]

In the next section we want to discuss some assumptions that we made and see how they may explain why our results did not become significant in some of the conditions.

3 Discussion

In this section we want to reflect upon two questions:

- a. Is there true complementarity in the behaviour of demonstrative and personal pronouns?
- b. Is grammatical role of the antecedent indeed the relevant parameter?

3.1 *Is there true complementarity of the pronoun types?*

In our initial corpus study it looked as if there was something like a complementary distribution of personal and demonstrative pronouns (cf. Figure 1). But upon closer inspection this impression turns out to be mistaken. Where the pronouns had a subject antecedent the pronoun type was a demonstrative in 23.6% of all cases and a personal pronoun in 76.4%. Where the antecedent was a non-subject, the pronoun was a demonstrative in 86.7% and a personal pronoun only in 13.2% of all cases. This shows clearly that the preference of the demonstrative for a non-subject antecedent is stronger than the preference of the personal pronoun for the subject antecedent. And this frequency fact is also supported by the intuitive judgements on the discourses in (2): The demonstrative in (2b) would *only* refer to Paul, the referent of the non-subject antecedent, while the personal pronoun in (2a), despite a certain preference for the referent of the sub-

ject antecedent, Peter, would not exclude an interpretation with reference to the non-subject antecedent, Paul.

The fact that the demonstrative's preference for the object antecedent became significant in the unbiased condition of the completion task, while the personal pronoun's preference for the subject antecedent did not is thus straightforwardly explained by the difference in the relative strength of the two preferences that we saw in the corpus study.

While the preferences of the two pronoun types for the two antecedent types are clearly of different strength this would of course not cause a reading time delay in the unbiased conditions. In the unbiased condition the reader *does not encounter any conflict* between different preferences the resolution of which would take time.

The failure to find significant results in the comparison of the preferences in the object-antecedent condition in both the reading time and the completion task can again be explained by the asymmetry in strength of the preferences of the two pronoun types, i.e., by the fact that the demonstrative's dislike for subject antecedents is stronger than the dislike of personal pronouns for non-subject antecedents.

This leaves us with only one apparent oddity in our results: The fact that the greater strength of the demonstrative's dislike for subject antecedents in the completion task became significant only in the SVO, but not in the OVS condition. This observation is indeed interesting. Why should the demonstrative dislike subject antecedents less when they are not in the regular subject position, i.e., in OVS word order? – For the moment let us just book this as a hint that grammatical role may not actually be the decisive factor in the antecedent selection for demonstratives, or at least not the only relevant factor. Demonstratives seem to shy away more strongly from subject antecedents that occur in canonical subject position than from subject antecedents that occur in canonical object position. We shall return to this point below in Section 4.2

In these reflections upon the conditions in which we did *not* have significant results in our experiment we have just seen that the lack of significance of some of the results only does not speak against our hypotheses, but may in fact be counted as evidence in favour of them.

3.2 Is grammatical role of the antecedent the truly relevant parameter?

There is a first indication that we may have been on the wrong track with our classification of antecedents, and indeed that any classification of antecedents could only be indirectly related to the relevant parameters, from the fact that demonstratives as well as personal pronouns can function perfectly well *without* antecedent expressions. Ever since substitutional accounts of pronouns (in the

style of Bloomfield) have fallen out of favour it has generally been agreed that the interpretation of personal pronouns (unbound pronouns to be precise) is more a matter of referent search than a search for coreferential antecedent expressions. Apart from bound pronouns of various kinds – where coreference with the antecedent is not really at issue anyhow – pronouns do not in general require explicit antecedent expressions; and if they have antecedents, they need not be coreferential with them, but the relation may be arbitrarily more complicated. One may think here of Donkey Pronouns (Kamp 1981) or E-type pronouns (Evans 1980), but also of pronouns whose referent must be inferred on the basis of one or more antecedent expressions with which the pronoun is not coreferential (Webber 1979). In particular, and this is the most striking case, there are also occurrences of both demonstrative and personal pronouns, that require no linguistic expression whatsoever that introduces a referent for the pronoun, and where we are still concerned with de-accented anaphoric uses and no physical pointing is required (cf. Bosch 1986):

- (4) (watching someone trying to move a book case full of books):
Wenn du die Bücher nicht rausnimmst, kriegst du {ihn/den} nie von der Stelle.
[If you won't take the books out, you'll never be able to move it {PPro/DPro}]

If pronouns can indeed function without antecedent expressions, it is plain that we cannot capture their interpretation constraints by looking at properties of antecedent expressions. In order to also capture antecedentless pronouns, we have to look at properties of their *referents*. We take this to be a decisive argument against any constraints that are formulated in terms of properties of antecedent expressions, in particular, against our own earlier formulations in terms of grammatical status of antecedent expressions.

4 Towards a Generalisation

The question then is how *referents* of personal pronouns differ from those of demonstrative pronouns. Bernard Comrie suggested, in an investigation of Dutch demonstrative pronouns (Comrie 1997), that demonstratives establish "less expected coreference" – which, for us, would qualify as a discourse property of referents. But he strangely uses expectability only as a way to distinguish between referents of different textually available antecedents. We shall not follow him in this respect and want to suggest instead that expectability of referents is a matter of the *information-structural properties* of the referent at a specific point in discourse – which may indeed be reflected in properties of antecedent expressions, provided there are any.

In order to find out what these referent properties could be and how they differ for the referents of personal and demonstrative pronouns in German, we

shall now discuss cases that form exceptions to our earlier Subject Hypothesis, i.e., cases where the preference of demonstratives for referents of preceding non-subject expressions and the preference of personal pronouns for referents of preceding subjects is violated. Let's see if we can capture these cases by regularities in terms of information structure.

4.1 *Exceptions to the subject hypothesis*

Counter to the Subject Hypothesis we also find personal pronouns that quite naturally pick up referents from non-subject antecedents (as the (a) and (b) examples show in (5) and (6)); and demonstrative pronouns can also co-refer with a subject antecedent, as is shown in the (c) examples – at least in cases where no competing referents are around.

- (5) a. Woher ich das weiß? Ich habe mit Peter_i gesprochen. {Der_i/Er_i} war gerade hier.
[How do I know? I spoke to Peter_i. He {DPro_i/PPro_i} has just been here.]
- b. Woher ich das weiß? Ich habe es von Peter_i gehört. {Der_i/Er_i} war gerade hier.
[How do I know? I heard it from Peter_i. He {DPro_i/PPro_i} has just been here.]
- c. Woher ich das weiß? Peter_i hat es mir gesagt. {Der_i/Er_i} war gerade hier.
[How do I know? Peter_i told me. He {DPro_i/PPro_i} has just been here.]
- (6) a. Woher Maria_k das weiß? Sie_k hat mit Peter_i gesprochen. {Der_i/Er_i} war gerade hier.
[How does Maria_k know? She_k spoke to Peter_i. He {DPro_i/PPro_i} has just been here.]
- b. Woher Maria_k das weiß? Sie_k hat es von Peter_i gehört. {Der_i/Er_i} war gerade hier.
[How does Maria_k know? She_k heard it from Peter_i. He {DPro_i/PPro_i} has just been here.]
- c. Woher Maria_k das weiß? Peter_i hat es ihr_k gesagt. {Der_i/Er_i} war gerade hier.
[How does Maria_k know? Peter_i told her_k. He {DPro_i/PPro_i} has just been here.]

The same is not true when there are competing referents available. Then we are back with the preference of demonstratives for non-subject antecedents (7a & b). But (7c) is an exception: In (7c) the demonstrative not only accepts, but strongly prefers the referent of the subject antecedent.

- (7) a. Woher Karl_i das weiß? Er_i hat mit Peter_k gesprochen. {Der_k/Er_{j,k}} war gerade hier.
[How does Karl_i know? He_i spoke to Peter_k. He {DPro_k/PPro_{i,k}} has just been here.]
- b. Woher Karl_i das weiß? Er_i hat es von Peter_k gehört. {Der_k/Er_{j,k}} war gerade hier.
[How does Karl_i know? He_i heard it from Peter_k. He {DPro_k/PPro_{i,k}} has just been here.]
- c. Woher Karl_i das weiß? Peter_k hat es ihm_i gesagt. {Der_k/Er_{i,k}} war gerade hier.
[How does Karl_i know? Peter_k told him_i. He {DPro_k/PPro_{i,k}} has just been here.]

Again, in the examples in (8), while the demonstratives in (a & b) reject the subject antecedent as they should, the demonstrative in (c) also rejects the non-subject antecedent, even though there is no competition.

- (8) a. Woher Maria_i das weiß? Sie_i hat mit Peter gesprochen. {[?]Die_i/Sie_i} war gerade hier.
 [How does Maria_i know? She_i spoke to Peter. She {DPro_i/PPro_i} has just been here.]
- b. Woher Maria_i das weiß? Sie_i hat es von Peter gehört. {[?]Die_i/Sie_i} war gerade hier.
 [How does Maria_i know? She_i heard it from Peter. She {DPro_i/PPro_i} has just been here.]
- c. Woher Maria_i das weiß? Peter hat es ihr_i gesagt. {[?]Die_i/Sie_i} war gerade hier.
 [How does Maria_i know? Peter told her_i. She {DPro_i/PPro_i} has just been here.]

4.2 What do these exceptions tell us?

Summarising what we saw in the previous section we may say that

- (A) Personal pronouns, despite a certain preference for subject antecedents, are not fussy about non-subject antecedents, even in competition situations.
- (B) Demonstrative pronouns, although they generally reject subject antecedents (7a,b & 8a,b),
- may also accept subject antecedents (5c, 6c), at least in non-competition situations;
 - may take a subject antecedent, as in (7c), with a strong preference over the non-subject antecedent;
 - and in (8c) the demonstrative rejects the non-subject antecedent, although there is no alternative.

The cases under (B) make it clear enough that the earlier generalisation in terms of grammatical role of the antecedent, although it is still a fair corpus generalization, definitely looks at the wrong parameter. – But what is the right parameter?

Take the hard core of the exceptions from (5)-(8): Demonstratives *accepting subject antecedents* (5c, 6c) and *excluding non-subject antecedents* (7c, 8c).

- (5c) Woher ich das weiß? Peter_i hat es mir gesagt. {Der_i/Er_i} war gerade hier.
 [How do I know? Peter_i told me. He {DPro_i/PPro_i} has just been here.]
- (6c) Woher Maria_k das weiß? Peter_i hat es ihr_k gesagt. {Der_i/Er_i} war gerade hier.
 [How does Maria_k know? Peter_i told her_k. He {DPro_i/PPro_i} has just been here.]

- (7c) Woher Karl_i das weiß? Peter_k hat es ihm_i gesagt. {Der_k/Er_{i,k}} war gerade hier.
[How does Karl_i know? Peter_k told him_i. He {DPro_k/PPro_{i,k}} has just been here.]
- (8c) Woher Maria_i das weiß? Peter_k hat es ihr_i gesagt. {[?]Die_i/Sie_i} war gerade hier.
[How does Maria_i know? Peter_k told her_i. She {DPro_i/PPro_j} has just been here.]

A generalisation that would be supported by these cases is that demonstratives *avoid referents that are discourse topics*.² Where by discourse topics we understand referents that are, in Prince's (1992) terms "discourse-old": They were introduced into the discourse before, not though as new referents in the immediately preceding sentence; they must have been discourse topics in the preceding sentence already. This condition is most clearly fulfilled in the case where a referent was referred to in the preceding sentence by an unstressed personal pronoun. – The referents of *ihm* and *ihr* in (7c) and (8c) are clear examples: They have the status of discourse topics in the sentences where *ihm* and *ihr* occur, while the referent of *Peter* in (5c) and (6c) does not have this status.

If this new generalisation is correct, it follows that the earlier generalisation in terms of subjects had to be *roughly correct*: Subjects *typically* refer to discourse topics – not though in the cases we have just been discussing, and less reliably when they do not occur in their canonical sentence-initial position (cf. section 3.1 above).

It also follows that what may well be the primeval use of demonstratives should be acceptable: Their use in physically *pointing to discourse-new referents* that were not previously mentioned and not referred to by antecedent expressions in discourse.

5 Conclusion

We can summarise the results of our discussion by the following characterisation of the interpretation options of German personal and demonstrative pronouns:

The preferred referents of personal pronouns are *discourse topics* which are – under an assumption of referential continuity of discourse – the most expectable referents.

Demonstrative pronouns choose their referents in contrast to the currently most expectable referent and thus *avoid discourse topics* as referents.

² As Comrie (1997) had found for Dutch, even though he spoke in terms of antecedents. Something very close to our proposal also seems intended in Zifonun et al. (1997, vol. 1:558), there formulated in terms of Theme and Rheme. Still Zifonun et al. also insist that both anaphoric personal pronouns and demonstratives are means of "thematic continuation", which would clearly contradict the generalisation we are proposing.

These still tentative results are compatible with all exceptions from our earlier Subject Hypothesis that we discussed and are they are supported by the results from our corpus study and the reading time experiment.

6 References

- Abraham, W. (2002). Pronomina im Diskurs: deutsche Personal- und Demonstrativpronomina unter 'Zentrierungsperspektive'. Grammatische Überlegungen zu einer Teiltheorie der Textkohärenz. *Sprachwissenschaft* 27.4: 447-491.
- Bosch, P. (1986). Pronouns under Control? - A Reply to Liliane Tasmowski and Paul Verluysen. *Journal of Semantics* 5: 65-78.
- Bosch, P., Rozario, T. & Zhao, Y. (2003). Demonstrative Pronouns and Personal Pronouns. German *der* vs. *er*. Proceedings of the EACL 2003. Budapest. *Workshop on The Computational Treatment of Anaphora*.
- Bosch, P., Katz, G. & Umbach, C. (2007). The Non-Subject Bias of German Demonstrative Pronouns. In: M. Schwarz-Friesel, M. Consten & M. Knees (eds.): *Anaphors in Text. Cognitive, formal, and applied approaches to anaphoric reference*. Amsterdam: John Benjamins.
- Comrie, B. (1997). Pragmatic Binding: Demonstratives as anaphors in Dutch. *Proceedings of the Annual Meeting of the Berkeley Linguistic Society* Vol. 23: 50-61.
- Evans, G. (1980). Pronouns. *Linguistic Inquiry* 11: 337-362.
- Kaiser, E., Trueswell, J. (2004). The referential properties of Dutch pronouns and demonstratives: Is salience enough? *Proceedings of Sinn und Bedeutung* 8, Arbeitspapier Nr. 1777, FB Sprachwissenschaft, Universität Konstanz.
- 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*. Amsterdam: Mathematisch Centrum, 277-322.
- Prince, E. F. (1992). The ZPG letter: subjects, definiteness and information-status. In S. Thompson & W. Mann (eds.): *Discourse Description: Diverse Analyses of a Fund Raising Text*. Amsterdam: John Benjamins.
- Webber, B. L. (1979). *A Formal Approach to Discourse Anaphora*. New York: Garland.
- Zifonun, G., Hoffmann, L. & Strecker, B. (1997). *Grammatik der deutschen Sprache*. Berlin: de Gruyter.