

# P & R A D I G M

preference and preference : analysing demonstrative pronouns in german

## Motivation

The aim of the project is to investigate how people process language with respect to personal and demonstrative pronouns (ER vs. DER pronouns). In particular, given two perfectly grammatical sentences :

1. **Der Oberarzt untersucht den Patienten.**
2. (a) **Er ist nämlich Herzspezialist.**  
(b) **Der ist nämlich Herzspezialist.**

why is there a preference for one pronoun form over the other?

## Research Question

The formal task set out was to examine how the DER / ER pronouns are used in natural language and to produce a theoretical framework that accounts for speakers' preference by characterising the syntactic, semantic, pragmatic and processing differences in the use of the DER and ER pronouns in everyday German discourse.

## Sentence Internal Features

Two corpora were made available to the project group: the Frankfurter Rundschau Corpus (~ 4 million tokens) and the NEGRA Treebank (~ 0.35 million tokens). Sentences containing DER & ER pronoun forms were extracted using regular expressions. The pie chart below is a comparison of the number of occurrences of each pronoun form in the FR Corpus.

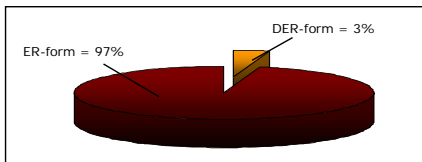


Figure 1 : Total no. of tokens of ER & DER-forms in FR Corpus

The pronoun forms were subsequently classed under two types of divisions: the Subject/Object division and the Vorfeld/Mittelfeld division.

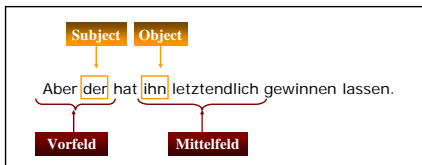


Figure 2 : Sentence-internal features chosen for further investigation

In isolating these sentence-internal features, it was hoped that distinct patterns of distribution for each pronoun type would emerge with respect to these features and as the chart below shows, this turned out to be the case.

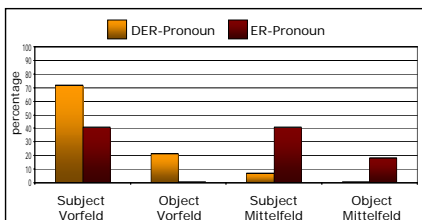


Figure 3 : Distribution of DER & ER pronouns in FR Corpus with respect to sentence-internal features chosen for investigation

The statistics above suggest that the DER pronoun prefers a fronted position in the sentence in which it occurs. We wanted to test whether native speakers were indeed sensitive to this preference and designed our first psycholinguistic experiment around this goal.

## Experiment One

The idea behind this experiment was to take a closer look at those sentence internal features described above (Subject/Object & Vorfeld/Mittelfeld) and to test how varying these features would affect the preferences underlying the choice of the DER and ER pronouns.

Thirty participants were given a self-paced reading task with forty sentences in which the position of the pronoun (either ER- or DER-pronoun) varied with respect to case (nominative/non-nominative) & position in the sentence (Vorfeld/Mittelfeld). Subsequently they were given a questionnaire containing sentence pairs (both ER and DER forms) and asked to choose which of the two they felt to be the most appropriate.

Figure 4 below shows the reading time of a pair of sentences presented during the experiment.

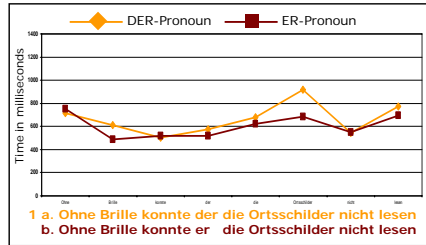


Figure 4 : Reading time course for sample test sentence

We observed differences in some of the sentence pairs but these differences were not consistent. This could have been caused by the self-paced reading method (effects may have been the subject of noise) or the stimuli may have been sub-optimal for the task.

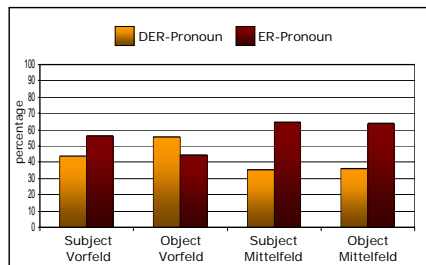


Figure 5 : Results of questionnaire comparing appropriateness of pronoun form for a given sentence

We found that in Subject and Object Vorfeld the participants had no preference for either pronoun. In the Mittelfeld we found a significant dispreference for the DER pronoun (below 40%).

To summarise, in the corpus work we observed that the DER pronoun occurred approximately 90% of the time in the Vorfeld position. In experiment one we found that, given a choice between the two pronoun forms in the Mittelfeld, there was a clear dispreference for the DER pronoun.

Having investigated sentence internal features, we now focussed our attention on how antecedent features and world knowledge might influence the choice of pronoun.

## Co-reference Relations

Since our corpora were not annotated for antecedent relations, we had to build a subcorpus ourselves. In order to build such a subcorpus, we constructed a user-interface to aid the annotation of the antecedent relations. See figure below. We used the Negra corpus rather than the FR corpus as it is more easily exportable and manageable.

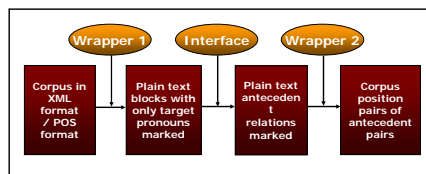


Figure 6 : Diagrammatic representation of antecedent tagging procedure



Figure 7 : Screenshot of the tagging user-interface

After the semi-automatic annotation in the Negra corpus, we conducted a frequency count of the antecedent relations. There are many features of the antecedents that are candidates for investigating the preference. We observed that the DER pronoun often referred to an NP in the previous sentence. Therefore we decided to consider only those features which were a course operationalisation of this observation, namely nominativity of the antecedent. Other factors like definiteness, referent type, voice or embeddedness were not considered.

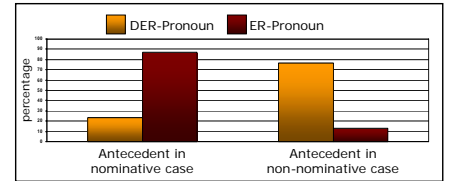


Figure 8 : Distribution of antecedents in previous sentence

Figure 8 shows the comparative distribution of antecedents of the two pronoun forms with respect to nominativity of the antecedent. We formulated a simple hypothesis based on the statistics above, namely that ER pronouns prefer nominative antecedents while DER prefer non-nominative ones. This hypothesis was tested in experiment 2.

## Experiment Two

In this experiment, we wanted to investigate the influence of context on the usage of the DER & ER pronouns, as well as finding evidence to support our hypothesis.

Seventy-two participants were presented eighty-four sets of phrases in a self-paced reading task. After each set, they were asked a comprehension question. On completion of the on-screen tasks, they were asked to complete a questionnaire.

Each trial consisted of a sequence of three lines of text: (1) a headline introducing the context, (2) a sentence introducing two protagonists in SVO order; and (3) a sentence containing a pronoun (DER or ER) whose reference to one of the protagonists was biased by world knowledge. See the sample trial below for details.

(1) Im Krankenhaus  
(2) Der Oberarzt untersucht den Patienten.  
(3a) Er/Der ist nämlich Herzspezialist.  
(3b) Er/Der muß sofort operiert werden.  
(3c) Er/Der ist gerade erst gekommen.

World knowledge supports ...  
reference to subject in sentence 2  
reference to object in sentence 2  
no clear referent

Participants were asked to fill in the gap in the comprehension question.  
(4) Der \_\_\_\_\_ ist nämlich Herzspezialist.  
(4b) Der \_\_\_\_\_ muß sofort operiert werden.  
(4c) Der \_\_\_\_\_ ist gerade erst gekommen.

Figure 9 : Sample test sequence for Experiment Two

The reading time for sentence (3) was measured. We found a significant difference between the DER and ER version of sentence type (3a). In other words, when world knowledge supported reference to the subject in sentence (2), the reading time for the DER pronoun version of sentence (3a) was delayed.

Our hypothesis was that the DER form prefers non-nominative antecedents. In sentences of type (3a), world knowledge forced reference to a nominative antecedent, but the DER pronouns readily picks up non-nominative antecedents. We suggest that the time delay was a result of this conflict.

Analysing how participants answered the comprehension question we found that world knowledge support was always followed. However, for the DER version of sentence type (3a), we found more participants answering sentence (4a) with the object. In addition, for the cases where world knowledge was non-specific, we find a clear preference for the DER form to refer to the object. See fig.10 for details.

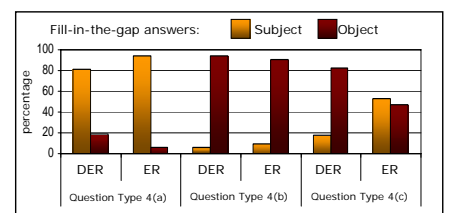


Figure 10 : Distribution of answers given to comprehension questions for SVO test sequence

## Conclusion

The experimental results support the hypothesis generated based on corpus findings: DER pronouns prefer non-nominative antecedents, while ER pronouns, although more flexible, prefer nominative antecedents. One possible interpretation of these facts is that ER pronouns prefer referents that are established as discourse topics, while DER pronouns prefer non-topical referents. Further work would include operationalising the term "topic-hood" using appropriate features and testing these features against native-speaker preferences.