

# Anaphora Resolution – What Helps in German

Holger Wunsch

SFB 441 – University of Tübingen, Germany

wunsch@sfs.uni-tuebingen.de

## 1 Introduction

Although anaphora resolution has been a very active research area throughout the last decades, there exist only a few studies that focus on German anaphora. Strube and Hahn (1999) present a system for anaphora resolution for German based on an extension of Centering Theory. Müller et al. (2002) and Kouchnir (2003) use co-training and boosting, respectively. Hinrichs et al. (2005) employ a hybrid approach combining a rule-based morphological pre-filter with a memory-based resolution module. With the exception of Schiehlen (2004), who gives a comprehensive survey of important factors for German anaphora resolution, these approaches assume essentially the same core set of features for German that the linguistic literature has come to agree upon for English.

It is however common knowledge that the way a linguistic phenomenon is actually realized crucially depends on the type of text it occurs in, and, of course, the language in which a text is composed. The present study aims to infer how helpful features that are commonly used are for the task of resolving anaphora in the domain of German newspaper text. The assumption is that features that are more important (or helpful) improve the performance of the anaphora resolution system when given higher weight over other features, and impair performance when given lower weights.

Following this paradigm, the results indicate that in the domain of German newspaper text, syntactic features (specifically those derived from an NP's grammatical function) should be considered more influential in the resolution process than positional features like sentence recency and the relative position of the antecedent or postcedent and the pronoun.

## 2 Resolution System

For this study, the “Resolution of Anaphora Procedure” (RAP) by Lappin and Le-ass (1994), a knowledge-rich, rule-based algorithm for the resolution of third person pronouns, has been re-implemented for German. In contrast to other approaches that

<b>Saliency factor</b>	<b>Weight</b>
<b>Syntactic features</b>	
Subject emphasis	170
Accusative object emphasis	70
Dative object emphasis	50
Genitive object emphasis	50
Head noun emphasis	80
Parallelism reward	35
<b>Positional features</b>	
Short distance cataphora penalty	-80
Long distance cataphora penalty	-175
Current sentence reward	20

Table 1: Saliency hierarchy used in the German version of RAP

are frequently based on machine learning strategies, linguistic principles are explicitly encoded in this rule-based system. Thus, the effects of linguistically motivated modifications to rules can directly be observed and evaluated.

RAP relies on measures of saliency derived from syntactic structure and a dynamic model of attentional state to select the NP antecedent of a third person pronoun. The features used by RAP can be grouped in two classes. “Positional features” (like the distance of a pronoun to its antecedent) depend on the relative position of a pronoun and a potential antecedent. “Syntactic features” (for instance the grammatical function of an NP) are determined by the syntactic configuration of the pronoun and/or a candidate antecedent.

Anaphora resolution in RAP is a two-step process. Since the search space of potential antecedents of a pronoun at first contains all noun phrases that precede or follow the pronoun (which can be a very large number), RAP applies a morphological and a syntactic filter in the first step that together substantially cut down the size of the search space. The morphological filter rejects all potential antecedents that do not agree with the pronoun in person, number or gender. The syntactic filter removes candidate antecedents that would violate binding constraints when paired with a reflexive or personal pronoun. The second step is the actual resolution step. RAP uses a saliency hierarchy that assigns weights to grammatical functions and a number of positional and syntactic configurations (see table 1). The system then sums up all saliency weights that correspond to appropriate features of a potential antecedent to compute its saliency value. RAP finally selects the candidate with the highest saliency value as the antecedent of the pronoun.

### 3 Data

The present study uses the first release of the “Tübingen Treebank of Written German” (TüBa-D/Z, Telljohann et al. (2003, 2004)), a large treebank of German newspaper text comprising 15.260 sentences that has been manually annotated with constituent structure and grammatical functions. Recently, coreference relations have been added to the treebank (Hinrichs et al., 2004).

### 4 Experiments and Results

The performance of the re-implementation of Lappin and Leass’ system with feature weights optimized for best performance on the TüBa-D/Z data was taken as a baseline. The optimized feature weights are listed in table 1.

In this configuration, the system achieved precision of 76.6% and recall of 76.5%, resulting in the F-measure of 76.6%. To obtain a measure of the relative importance of the features used by the algorithm, multiple experiments were run, with either the weight of one feature set to zero (thereby ignoring the information provided by this feature), or to a very high value (thereby emphasizing the information provided by this feature). The performance of the system was measured after each experiment. We assume that the greater the difference to the baseline, the more the performance of the algorithm is affected by the presence or absence of the information related to a feature.

The results are interesting. The most surprising finding concerns the “current sentence reward”, which increases the salience of a potential antecedent if it occurs in the same sentence as the pronoun. While both decreasing and raising the weight of this feature lowers performance, the loss is notably larger with higher weights. Potential cataphoric relations are penalized in RAP since they occur less frequently in discourse than anaphoric relations. It turns out that for the German version, performance improves when this strategy is relaxed in such a way that potential postcedents in the same sentence as the pronoun are less penalized than potential postcedents in following sentences. Both the current sentence reward and the cataphora penalty are based on the relative position of a pronoun and a potential antecedent. The finding that RAP’s performance improves when giving less prominence to positional features indicates that resolution can not benefit to a large degree from using these features on newspaper text. It is noteworthy that in Lappin and Leass’ original experiments on a corpus of English computer manuals, positional features turned out to be crucial for correct resolution.

Information provided by syntactic features on the other hand does seem to be helpful. Turning off the subject feature, which is assigned to potential antecedents that are in subject position, results in a dramatic loss in performance of more than 10 percentage points. Even though the literature agrees on the fact that subject NPs are likely to

be antecedents, the substantial negative effect of ignoring the subject feature in the given domain is remarkable. Ignoring or emphasizing other features related to other grammatical functions however yields only moderate effects.

## 5 Conclusion

The present study uses a rule-based approach to anaphora resolution to assess the importance of features for anaphora resolution based on the assumption that when ignoring very important features the resolution system's performance is worse than when ignoring less important features. The results indicate that positional features are less helpful for anaphora resolution in the domain of German newspaper text than in other domains. On the other hand, the resolution process does benefit from syntactic features, with the subject feature being of special significance.

The answer to the question why features behave differently in the newspaper domain than in other domains is subject to future research. It might be due to a specific style of writing in newspapers that allows antecedents to be distributed more freely over the text than in other domains. One might also speculate that this has to do with the more general property of German being a highly-inflectional, free word-order language. In such a language, the actual position where antecedents can be found may vary much more than in a fixed word-order language like English, which means that coreference is less determined by position-related effects.

## 6 Acknowledgments

Acknowledgments go to Erhard Hinrichs, Sandra Kübler, Yannick Versley and Heike Zinsmeister for their valuable comments and encouragement, as well as to the anonymous reviewers for their very helpful remarks.

## References

- Hinrichs, E., S. Kübler, K. Naumann, H. Telljohann, and J. Trushkina (2004). Recent developments in linguistic annotations of the TüBa-D/Z treebank. In S. Kübler, J. Nivre, E. Hinrichs, and H. Wunsch, eds., *Proceedings of the Third Workshop on Treebanks and Linguistic Theories*. Tübingen, Germany.
- Hinrichs, E. W., K. Filippova, and H. Wunsch (2005). A Data-Driven Approach to Pronominal Anaphora Resolution for German. In *Proceedings of Recent Advances in Natural Language Processing (RANLP)*. Borovets, Bulgaria.
- Kouchnir, B. (2003). *A Machine Learning Approach to German Pronoun Resolution*. Master's thesis, School of Informatics, University of Edinburgh.

- Lappin, S. and H. J. Leass (1994). An algorithm for pronominal anaphora resolution. *Comput. Linguist.*, **20**(4):535–561. ISSN 0891-2017.
- Müller, C., S. Rapp, and M. Strube (2002). Applying Co-Training to Reference Resolution. In *ACL02*, pp. 352–359.
- Schiehlen, M. (2004). Optimizing algorithms for pronoun resolution. In *Proceedings of the 20th International Conference on Computational Linguistics*. Geneva.
- Strube, M. and U. Hahn (1999). Functional Centering – Grounding Referential Coherence in Information Structure. *Computational Linguistics*, **25**(3):309–344.
- Telljohann, H., E. W. Hinrichs, and S. Kübler (2003). Stylebook for the Tübingen Treebank of Written German (TüBa-D/Z). Technical report, Seminar für Sprachwissenschaft, Universität Tübingen, Tübingen.
- Telljohann, H., E. W. Hinrichs, and S. Kübler (2004). The TüBa-D/Z Treebank – Annotating German with a Context-Free Backbone. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC 2004)*. Lisbon, Portugal.