

Creating a Feature Space for the Annotation of Preposition Senses in German

Antje Müller, Katja Keßelmeier, Claudia Roch, Jan Strunk, Tobias Stadtfeld, Tibor Kiss

Ruhr-Universität Bochum

mueller@linguistics.rub.de

1 A Reference Corpus for Preposition Senses

We present an annotation scheme for preposition senses in preposition-noun-combinations (PNCs) and PPs in German. PNCs are combinations of prepositions with determinerless nominal projections such as in (1):

- (1) *auf Anfrage* (‘after being asked’), *unter (sanfter) Androhung* (‘under (gentle) threat’), *mit Vorbehalt* (‘with reservation’)

They present an anomaly in the grammar because they violate the rule on the realization of countable singular nouns, viz. that such nouns have to appear with a determiner. For some time, PNCs have been treated as exceptions, but recent research has shown that they are indeed productive and no more idiomatic than other phrasal combinations (Dömges et al., 2007, Kiss, 2007). In search of licensing conditions for PNCs it turns out that not every P can appear in a PNC, so that we limit our analysis to the following prepositions¹:

- (2) *an* (4), *auf* (5), *bei* (5), *binnen* (1), *dank* (2), *durch* (4), *für* (10), *gegen* (5), *gemäß* (1), *hinter* (1), *in* (6), *mit* (10), *mittels* (1), *nach* (6), *neben* (3), *ohne* (4), *seit* (1), *über* (6), *um* (5), *unter* (9), *vor* (4), *während* (1), *wegen* (2)

Baldwin et al. (2006) have claimed that preposition senses in PNCs in English are more restricted than in PPs in general. Initial investigations have shown that senses of prepositions in German PNCs are restricted, too. Yet, the exact nature of the restriction remains unclear. The preposition *unter* (‘under’), for example, does not allow spatial (and temporal) interpretations in PNCs, unless they appear in the context of newspaper headlines, but there is no evidence for a general ban on local (or temporal) interpretations in PNCs. To determine the distribution of preposition senses in language data, it is necessary to develop an annotation scheme for preposition senses, which will facilitate manual annotation. We are planning to develop a reference corpus of preposition senses large enough to allow automatic

¹ The number in parentheses offers the number of possible senses for each preposition. The issue will be addressed in more detail below.

sense annotation eventually. Automatic annotation will not only rely on the preposition sense annotations, but on various annotations on the lexical, syntactic, and conceptual level, comprising particularly a complex conceptual annotation of nominal complements provided by HaGenLex (Hartrumpf, Helbig and Osswald, 2003). Presently, the corpus data are represented in a stand-off-format provided by MMAX2 (cf. Müller and Strube, 2006).

2 Building an Inventory of Preposition Senses

It seems to be common wisdom that prepositions (especially simple prepositions) tend to occur very frequently and are highly polysemous. Most simple prepositions in German have spatial interpretations, from which metaphorical usages have been derived. Temporal interpretations are also common with the majority of simple prepositions, whereas other interpretations (as e.g. *order*) are represented to a lesser extent. In addition, certain interpretations are restricted to single prepositions. In the development of a comprehensive inventory of preposition senses we must keep in mind the bottleneck of human annotation. We thus consulted several reference works, combining our findings in a second step. We selected the German grammar by Helbig and Buscha (2001) and the dictionary Duden Deutsch als Fremdsprache (Duden, 2002), as well as a dictionary of German prepositions (Schröder, 1986). It was necessary to consider additional literature for prepositions with temporal senses because the dictionaries do not provide enough information on these senses.

We have identified 28 different top level interpretations from these dictionaries for our restricted inventory of 23 prepositions. Spatial, temporal, causal, and modal senses are differentiated from the other elements of the sense inventory in that subcategories are defined for these four senses. Annotators, however, are free to either use the most general sense or a subsense in their annotations as the scheme allows an automatic mapping from subsenses to supersenses. Moreover, for spatial and temporal senses, decision trees implemented in MMAX2 guide the annotators (see section 3). The remaining senses do not show subsenses and are typically only instantiated by few prepositions. (3) gives a full list of the top level interpretations together with the number of prepositions that can instantiate them.

- (3) temporal (15), causal (14), spatial (13), modal (11), order (3), comparison (3), point of reference (3), statement/opinion (3), exchange (3), transgression (2), state (2), communality (2), affiliation (2), correlation (2), restrictive (2), theme (2), substitute (2), adversative (2), distributive (2), recipient (1), property (1), inclusion (1), partitive (1), copulative (1), extension (1), participation (1), agent (1), subordination (1)

In some cases, a final distinction between two senses cannot be drawn:

- (4) *Feuer nach [temporal/causal] Blitzschlag*
 Fire after/because of lightning stroke

For cases like (4), the scheme allows the assignment of multiple senses.

In addition to these semantic features, we use the feature *governed* for prepositions governed by a lexical head. Although these prepositions are usually assumed to only show light semantics, if at all, this feature assignment does not preclude the assignment of an additional semantic feature if it turns out that the preposition shows a discernible meaning despite its being governed.

As an initial result of the compilation of preposition senses, it turns out that three types of relations between prepositions and senses can be distinguished: First, a few prepositions (such as *mit*) are highly polysemous in allowing up to 10 discernable senses. Secondly, many of the 23 prepositions exhibit polysemy w.r.t. local, causal, temporal, and modal senses; and thirdly, many other senses are only instantiated by few prepositions which are often not polysemous at all. We can thus conclude that it would be premature to call prepositions highly polysemous *in general*.

3 Properties of Decision Trees

It might be apparent to the reader that no distinction between local and directional readings has yet been made. Instead of adopting a top level distinction between local and directional senses, we are able to capture the difference systematically within a decision tree for spatial interpretations. Specifically, we use one of three cross-classifying features, named *directional*, applicable to all local preposition senses. With the exception of a single preposition (*nach*), the investigated spatial interpretations show a case alternation corresponding to the difference between local and directional interpretations, so that the decision between local and directional can be based on the case of the complement.

The other two cross-classifying features used are *contact* and *tangible/concrete*. *Contact* is a feature applying to all reference planes, signifying whether or not a contact has been established. The feature *tangible/concrete* specifies the concreteness of the planes. As an illustration, consider *Himmel* ('sky') in (5) which can be seen as a top surface but is not tangible.

(5) *Sterne am Himmel* ('stars in the sky')

The tree is based on the descriptions in Schröder (1986). It entails the original spatial interpretations of the prepositions (like localization with regard to a local reference point) as well as some metaphorical extensions, which occur on a regular basis.

With regard to temporal interpretations, we have been able to build on a decision tree for temporal interpretations of German prepositions developed in Durell and Brée (1993). The decisions in the tree are based on the distinction between a matrix and a subordinate eventuality, the characteristics of these eventualities as well as on the identification of the temporal relationship between them: whether they occur at the

same time or in sequence, whether they label points in time or periods and how they are temporally related to the time of discourse.

4 Summary

We have developed a scheme for the manual annotation of preposition senses in German, which will lead to a reference corpus of preposition senses in PPs and PNCs. The corpus will not only be useful as a resource for further investigating the realization of preposition senses in PPs and PNCs, but can also be used as a reference corpus for training automatic methods for preposition sense tagging.

References

- Baldwin, T., J. Beavers, L. van der Beek, F. Bond, D. Flickinger, and I. Sag (2006). In Search of a Systematic Treatment of Determinerless PPs. In P. Saint-Dizier, *Syntax and Semantics of Prepositions*, pp. 163-179. Dordrecht, Springer.
- Dömges, F., T. Kiss, A. Müller, and C. Roch (2007). Measuring the Productivity of Determinerless PPs. *Proceedings of the ACL 2007 Workshop on Prepositions*, pp. 31-37. Prague.
- Duden (2002). *Duden. Deutsch als Fremdsprache*. Mannheim, Bibliographisches Institut and F.A. Brockhaus AG.
- Durell, M. and D. Brée (1993). German temporal prepositions from an English perspective. In C. Zelinsky-Wibbelt, *The Semantics of Prepositions. From Mental Processing to Natural Language Processing*, pp. 295-325. Berlin/New York, De Gruyter.
- Hartrumpf, S., H. Helbig, and R. Osswald (2003). The Semantically Based Computer Lexicon HaGenLex – Structure and Technological Environment. *Traitement automatique des langues* 44(2), pp. 81-105.
- Helbig, G. and J. Buscha (2001). *Deutsche Grammatik. Ein Handbuch für den Ausländerunterricht*. Leipzig, Langenscheidt.
- Kiss, T. (2007). Produktivität und Idiomatizität von Präposition-Substantiv-Sequenzen. *Zeitschrift für Sprachwissenschaft* 26, pp. 317-345.
- Müller, C. and M. Strube (2006). Multi-level annotation of linguistic data with MMAX2. In Braun, S., K. Kohn, and J. Mukherjee (eds.), *Corpus Technology and Language Pedagogy: New Resources, New Tools, New Methods*. Frankfurt a.M., Peter Lang.
- Schröder, J. (1986). *Lexikon deutscher Präpositionen*. Leipzig, VEB Verlag Enzyklopädie.