The annotation of preposition senses in German

Antje Müller, Claudia Roch, Tobias Stadtfeld, Tibor Kiss

1. Introduction

1.1. Prepositions and their polysemy

While there is general agreement that prepositions (especially simple prepositions) are highly polysemous, there are two competing views on how polysemy might occur: the derivational view assumes that simple prepositions derive from and retain the basic meanings of adverbs, such as temporal, spatial or modal relations. In line with this, it has been assumed that preposition senses can be derived from primordial protosenses, typically spatial senses (cf. Tyler and Evans 2001). Alternative proposals on polysemy, however, suggest that different senses of words only share their phonological form and have emerged arbitrarily and independently at some point in time (cf. Croft 1998).

We favor the second approach and assume a relational view on prepositions.1 Accordingly, prepositions are not specified as bearing a prototypical meaning. Instead, we assume a set of (sometimes arbitrary) relational preposition senses that are associated with preposition lexemes. Consequently, we avoid speaking of *spatial prepositions,* etc. and prefer to speak of preposition senses and their mappings to preposition lexemes.

Of course, it cannot be denied that most simple prepositions in German have spatial interpretations. But neither can it be denied that almost as many prepositions show temporal or conditional interpretations, while other senses are distributed to a lesser degree among prepositions. For instance, the interpretation PRESENCE is expressed only with *mit* ('with') and *ohne* ('without') in German.

(1) *Kommentarband* *mit* *einem Beitrag* *von Albert Knoepfli*
Commentary volume with an article by Albert Knoepfli
'Commentary volume including an article by Albert Knoepfli'

Typically, senses like PRESENCE are investigated to a lesser extent. In general, we deem it implausible to restrict investigations of preposition senses

to a few, possibly prototypical cases, since a complete description of possible preposition senses is thus a requirement of an analysis of prepositional polysemy. This demand becomes even more important as we try to determine the distribution of preposition senses in large corpora. The definition of an annotation scheme is a prerequisite for an analysis of the distributional patterns of preposition senses, and for their automatic classification.

1.2. Existing literature on preposition senses

Preposition senses seem to be well explored. On a closer inspection, however, it turns out that we still lack a structured understanding of preposition senses. As far as we know, there is no account that provides every piece of information required for an annotation scheme of preposition senses. In dictionaries the description is quite often limited to simple – often handcrafted – examples and some more or less informative sense labels (Duden 2002; Kempcke 2000). Criteria for the distinction of senses are not provided and the interpretation of sense labels varies across the prepositions described. One could imagine that this is due to a lack of space in print dictionaries, due to a tension between page limitations and comprehensiveness. But electronic dictionaries, like the DWDS2, do not provide a more detailed analysis either.

We sometimes find dictionaries or collections that focus on prepositions, and attempt to systematize the spectrum of preposition senses. Schröder's *Lexikon deutscher Präpositionen* (Schröder 1986) stands out in including a fine-grained feature-based analysis of preposition senses in German. Making use of over 200 binary features, it is, however, too complex to be feasible for corpus annotation.

Reviewing the existing grammars and dictionaries, it is conspicuous that the examples used in reference works are repeated over and over again, rarely being revised or replaced. If the characterization of possible senses is carried out in neglect of large data sets then senses tend to be ignored that do not come to mind immediately.

Some work selects a subgroup of prepositions with a common subsense (e.g. Retz-Schmidt 1988; Wiese 2004). These publications are mostly concerned with spatial or temporal prepositions, describing the options of realization of space and time in language, and seldom touch on the other subsenses of the prepositions.

2. Preposition senses and preposition-noun combinations

In a preposition-noun combination (PNC), a preposition governs a determinerless nominal projection, the head of which must be a singular count noun. PNCs should be 'convertible' into PPs by adding a determiner. Some German examples are provided in (2):

(2) *auf Anfrage* ('on being asked'), *unter Androhung* ('under threat'), *mit Vorbehalt* ('with reservation')

PNCs can be prenominally modified (3), and postnominal complementation is licit as well (4).

(3) *auf parlamentarische Anfrage* ('after being asked in parliament'), *mit beladenem Rucksack* ('with loaded backpack')

(4) *Er wehrt sich gegen die Forderung nach*
He defies REFL against the demand for
Stilllegung einer Verbrennungsanlage.
closedown an incineration plant
'He defies the demand for closing an incineration plant.'

Yet PNCs present a crosslinguistically attested anomaly (cf. Himmelmann 1998): They violate the rule that singular count nouns have to appear with a determiner, as it is stipulated in the Duden rule 442 for German: *"Substantive mit Merkmalkombination ‚zählbar' plus Singular haben [...] grundsätzlich immer ein Artikelwort bei sich, und wenn es als letzte Möglichkeit der indefinite Artikel ist."* [a determiner must accompany singular count nouns, even if this is the indefinite article as a last resort] (Duden 2005).

For some time, PNCs have been treated as exceptions, as can be illustrated with the Duden Grammar for German (Duden 2005). As an amendment to rule 442, it offers rule 395 to cover PNCs. But this rule is a mere list of exceptions ascribing PNCs to special registers. Addressing such claims, recent research has shown that PNCs are indeed productive (Stvan 1998; Baldwin et al. 2006 for English; Dömges et al. 2007 for German).

Most researchers assume that the semantics of the preposition plays a major role in determining the grammaticality of PNCs. Baldwin et al. (2006), arrive at the conclusion that preposition senses in English are more restricted in PNCs than in PPs in general. Initial investigations have shown that certain senses seem to be blocked in German PNCs, too. The preposition *unter* ('under'), e.g., does not allow spatial interpretations in PNCs, unless it appears in the context of newspaper headlines.

We propose an analysis for PNCs that is based on corpus annotation, and logistic regression modeling (Harrell 2001). Annotation mining (cf. Chiarcos et al. 2008) is a recent extension of text mining, where linguistically relevant generalizations and correlations are derived in a bottom-up fashion from a suitably annotated corpus with multiple layers of information. In the case at hand we include several annotation layers that enrich our data with part of speech tags, shallow and deep syntactic analysis, sense annotation, and complex conceptual annotations derived from resources such as HaGenLex (Hartrumpf, Helbig, and Osswald 2003) and GermaNet (Kunze and Lemnitzer 2002). Logistic regression modeling is a suitable tool to characterize determiner omission, as logistic regression identifies features (i.e. annotations) that are relevant for a binary distinction – realization or omission of the determiner in the present case. The annotations thus serve to induce licensing conditions for the omission of a determiner in PPs.

But not every preposition can appear in a PNC. We therefore limit our analysis to (simple) prepositions that are allowed in PNCs and typically take NP complements in the singular when appearing in PPs. For instance, we do not consider *bis* ('until') that only allows NP complements if the NP refers to a temporal entity, but typically selects adverbial phrases as complements, *zwischen* ('between') that requires a coordination of two NPs or a plural NP as complement, or *per* ('per') that always joins with a bare nominal in German. We have also excluded secondary prepositions, as they never occur in PNCs, and prepositions that do not govern a case.

Starting with the description of prepositions in traditional grammars (cf. Helbig and Buscha 2001), we arrived at the following set of prepositions for investigation: *an, auf, bei, dank, durch, für, gegen, gemäß, hinter, in, mit, mittels, nach, neben, ohne, seit, über, um, unter, vor, während,* and *wegen*.

The development of an annotation scheme for preposition senses should not only be useful for the task at hand. It will also allow the definition of a gold standard for automated preposition sense annotation. The resource will be considerably larger than available corpora that are fully annotated but smaller, and thus not sufficient for our needs. Previous investigations have shown that between 2,500 and 5,000 PNCs/PPs per preposition are necessary to provide reliable regression models, and we assume this order to be applicable to automated preposition sense annotation as well.

3. Building an annotation scheme for preposition senses

3.1. The reference corpus

The Swiss-German newspaper "Neue Zürcher Zeitung" (vols. 1993-1999) forms the basis for the annotation, comprising approx. 230 million words. The annotation employs an XML-stand-off format.3 The annotation tool MMAX2 (Müller and Strube 2006) is used for manual annotation.

For each preposition, we consider three datasets that enter into investigation: PNCs (app. 91,000 instances), corresponding PPs (approx. 320,000 instances) with the same count noun, and PPs (approx. 122,000 instances) containing count nouns not appearing inside PNCs.

The identification of count nouns is based on the combination of two classifiers, a decision tree classifier based on the C4.5 algorithm (Quinlan 1986) and a Naïve Bayes classifier (Witten and Frank 2005). The analysis of the classifiers resulted in 4,431 fully countable nouns that are considered in the three datasets. The following annotations are provided for each dataset:

- Lexical level: Part of speech, inflectional morphology, derivational morphology of nouns, interpretation of nouns, interpretation of prepositions, noun compounding.
- Syntactic level: Mode of embedding of the phrase (adjunct or complement), syntactic dependents of the noun, modification of the noun.
- Global level: Is the phrase contained in a headline, title, or quotation? Is the phrase idiom-like?

Headlines, titles, and quotations are particularly prone to text truncation. Similarly, PNCs and PPs in idioms might follow combination rules that differ from general modes of combination.

For automatic annotation the following tools are applied: The Regression Forest Tagger (Schmid and Laws 2008) for POS tagging and morphological analysis (the tagger contains the SMOR component for morphological analysis, cf. Schmid, Fitschen, and Heid 2004), the Tree Tagger (Schmid 1995) for chunk parsing, and the Malt-Parser (Nivre 2006) for syntactic dependencies.

To determine noun meanings we make use of two resources. The first resource is GermaNet (Kunze and Lemnitzer 2002), the German version of WordNet. We employ 23 top-level categories4, and each noun is annotated with every top-level category it belongs to.5 Secondly, we use the computer

lexicon HaGenLex (Hartrumpf, Helbig, and Osswald 2003), which offers specific sortal information derived from a formal ontology for each noun.

As the current set-up is concerned with the creation of a reference corpus, annotations on the global level and preposition sense annotation are carried out manually.

3.2. An inventory of preposition senses

There is as yet no standardized inventory of preposition senses for German. Ideally, one would employ a universally applicable scheme, but such a scheme would require language-specific mappings from senses to prepositions and vice versa. The Preposition Project (cf. Litkowski and Hargraves 2005) as well as PrepNet (Saint-Dizier 2005) offer categories for preposition senses in English and Frensh respectively. Yet, we did not make use of these resources, as mappings from senses to lexemes are inherently language-specific. The specificity of the mapping can be illustrated with the prepositions *über* ('above', 'over') and *nach* ('to', 'after') as discussed in section 3.3.1. If we started with an English inventory here, a mapping to two different lexemes would have to be taken into account, as well as the polysemy of the two lexemes, which is only partially present in the German lexeme. We envisage a comparison of the present scheme with the existing resources to identify cross-linguistic mappings.

We started the development of the present scheme by consulting the German grammar by Helbig and Buscha (2001) and the dictionary *Duden Deutsch als Fremdsprache* (Duden 2002), as well as Schröder's dictionary of German prepositions.

Dictionaries quite often do not seek complete coverage of preposition senses. So we did not only rely on the reference works and determined our scheme with the help of examples collected there, but defined the scheme iteratively, testing it against corpus data and filling in missing senses if required.

The scheme makes use of three characteristic features: First, it is hierarchically organized, allowing for the inclusion of taxonomies of subsenses, or the representation of decision trees to arrive at specific subsenses for individual prepositions. Secondly, it allows multiple sense assignments if further disambiguation proves to be infeasible. Thirdly, certain general properties of senses are extracted as cross-classifying features, allowing a compact representation of decision trees and taxonomies, as, e.g., for the ubiquitous distinction between local and directional senses.

As was already mentioned, the scheme is based on a hierarchical tree-like structure. Beginning with a root node, types of preposition meanings branch to subtrees for different classes (e.g. spatial, temporal or conditional) with differing depths or to individual, unary branches.

We have identified 27 different top-level senses for our restricted inventory of 22 prepositions (cf. (5) below).6 The senses SPATIAL and TEMPORAL are mapped to decisions trees. The senses CONDITIONAL and MODAL are mapped to subhierarchies of senses. This holds for the sense PRESENCE as well – which is only employed by the prepositions *ohne* and *mit* (cf. section 5) – and takes two subcategories, viz. ANALYTIC and SYNTHETIC. For the latter hierarchies, the scheme is agnostic as to the ontological status of senses. Assignment of a supersense instead of a most specific subsense may thus imply that a more specific interpretation is conceivable but cannot be derived from the criteria at hand. While the scheme includes an automatic mapping from subsenses to supersenses, it is naturally desirable that the most specific annotation is provided whenever possible. As spatial and temporal senses are represented as decision trees, sense annotation will always yield a most specific subsense (cf. next section).

The majority of senses do not show subsenses. The senses are typically only instantiated by few prepositions. A complete list of the top-level senses is given in (5).

(5) *conditional, spatial, temporal, modal, adversative, affiliation, agent, centre of reference, communality, comparison, copulative, correlation, distributive, exchange, extension, hierarchy, inclusive, order, participation, presence, realization, recipient, state, statement, substitute, theme, transgression.*

The senses in (5) were derived from the description of preposition senses in the three resources mentioned above. We compared descriptive and illustrative interpretations, and developed abstractions. This led to the unification of several senses, as well as to a further differentiation of other senses.

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

(6)	*Feuer*	*nach [temporal/conditional-causal]*	*Blitzschlag*
	Fire	after/because-of	lightning strike

It is for cases like (6) that the scheme allows the assignment of multiple senses to individual instances.7

3.3. Properties of spatial and temporal decision trees

It seems obvious that sense labels like TEMPORAL or SPATIAL are not particularly revealing. Yet examples in dictionaries often just provide this kind of characterization, leaving open questions e.g. as to why two prepositions with spatial senses and identical case requirements cannot be substituted *salva veritate.* Instead, we assume decision trees for SPATIAL and TEMPORAL interpretations, which encode fine-grained distinctions between individual spatial and temporal senses, respectively. Hence the assignment of specific subsenses is guided by the application of specific criteria, which also facilitate the annotation process.

3.3.1. Spatial interpretations

We do not classify spatial prepositions as such. Instead, the present scheme classifies spatial senses that are associated with the respective prepositions. Criteria, like the differentiation between topological and projective (e.g. Wunderlich 1986), were used to distinguish different senses.

Topological prepositions typically locate one object (the localized object, LO) in a neighbouring region of another object (the reference object, RO). They can be distinguished from projective prepositions whose semantics have to include a directional vector or reference axes. As Wunderlich proposed, the combination with spatial measurements like *zwei Meter* ('two metres') is only possible with projective but not with topological prepositions.

In addition to the topological interpretation of *in* ('in'), *an* ('at', 'by'), *auf* ('on', 'at') and *bei* ('at', 'by', 'near') we have added topological senses for *unter* ('under'/'below') and *über* ('over'/'above'). A topological sense is illustrated for *über* in (7), where a vertical axis plays no role in the interpretation.8

(7) *Das Bild hängt über dem Loch.*
The picture hangs above the hole
'The picture hides the hole.'

In the same vein, we do not assume that *auf* is a purely topological preposition, but is ambiguous between topological and projective readings.

Topological and projective preposition senses can be distinguished from shape-related *um* ('around'), an interpretation of *nach* ('after'/'behind'), and some directed senses.

Topological preposition senses can be further divided into the ones locating the LO in a region inside the RO, the ones locating the LO exterior to the RO, and the ones used for a traversal of the RO by the LO. As can be seen in Figure 29, the criteria for identifying the pertinent senses for localizing a LO within a RO are identical to the ones for localizing a traversal, except that they map to different prepositions. This captures the consistent behaviour of the path prepositions *über* ('across') and *durch* ('through') and the topological prepositions *auf* and *in* with respect to dimensionality. *In* and *durch* require their internal argument, the RO, to have at least the number of dimensions of the LO (cf. Kaufmann 1993).

(8) *Schweizer Truppen auf deutschem Gebiet*
'Swiss troops on German territory'

(9) *Schweizer Truppen gehen über deutsches Gebiet.*
'Swiss troops walk across German territory.'

(10) *Er liegt im Wald.*
'He lies in the woods.'

(11) *Er geht durch den Wald.*
'He walks through the woods.'

Directionality does not affect the basic criteria for the identification of spatial senses with regard to regions and axes. Therefore the scheme in Figure 2 does not distinguish between directional and local interpretations. Instead, the feature [±DIR] will be added to a sense after the classification in Figure 2 has been traversed, and is not listed as a separate feature in Figure 1 and Figure 2 below.[10]

Directional preposition senses must be kept apart from preposition senses that are directed. Directed preposition senses express an inclination or alignment. They split up into the target-orientated interpretations of *nach* ('to', but 'after' with other senses) and *gegen* ('to') (12), and the interpretation 'in line with' of *mit* ('with') and its counterpart *gegen* ('against') (13).

(12) *Das Pendel schlug nach der Seite/gegen die Seite aus.*
'The pendulum swung to the side.'

(13) *Ernst fotografiert mit dem Licht./gegen das Licht.*
'Ernst takes a picture with the light/against the light.'

We distinguish the aforementioned localizations (Figure 2) from localizations that impose selectional restrictions on the syntactic object of the prep-

osition (Figure 1). The latter are typically excluded from systematic classifications of spatial prepositions, and cannot be compositionally derived from the more ordinary spatial senses. The PP in (14) does not denote the property of being localized in the proximal region of Herbert's parents, but implies that Herbert is visiting his parents' home, regardless of his parents being there at all. Such interpretation shifts are only possible if the object of the preposition meets the relevant restrictions.[11]

(14) *Herbert* *ist* *bei* *seinen Eltern.*
Herbert is at his parents
'Herbert is at his parents' home.'

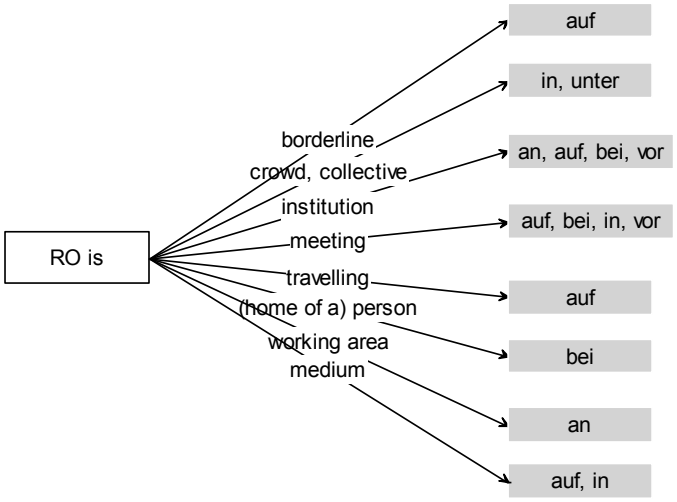

Figure 1. Selectional Restrictions on RO

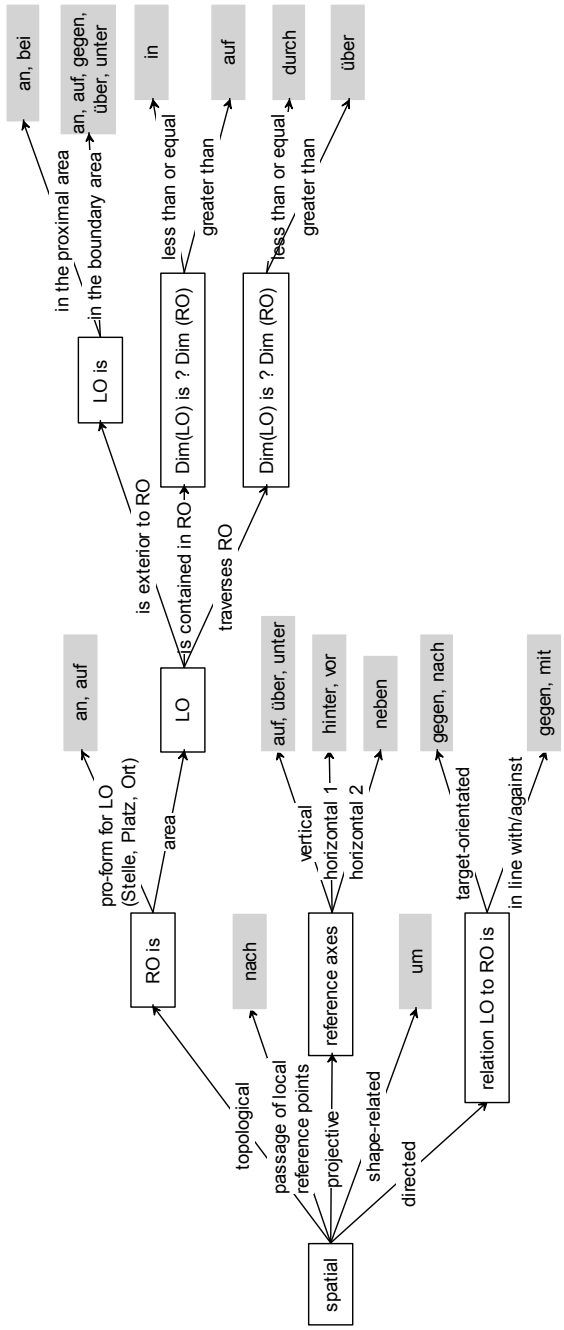

Figure 2. Spatial Relations

3.3.2. Temporal interpretations

Here, we have been able to build on a decision tree for temporal interpretations of German prepositions developed in Durrell 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, and the identification of the temporal relationship between them.

Where Durrell and Brée wanted to offer guided choices from senses to lexemes, we wanted to obtain a useful feature space to annotate the temporal interpretations of prepositions as they are used in PNCs or PPs. As there was no need for the included features to be distinct for all prepositions, we could prune the tree of Durrell and Brée at certain points. But we also had to amend Durrell and Brée's tree because certain interpretations were missing.

The tree itself is divided into two parts: One for prepositions building a time measure phrase with their complements, and one for prepositions that do not. A time measure phrase can either frame the duration of an eventuality (the matrix eventuality) (15) or relate the time in which the matrix eventuality is taking place to a reference time (16) (which is equal to the utterance time in this example).

(15) *Sie wohnt seit drei Jahren in Hamburg.*
She lives for three years in Hamburg
'She has been living in Hamburg for three years now.'

(16) *Vor einer Woche haben die Kurse angefangen.*
Ago one week have the courses started
'The courses started one week ago.'

Non-time-measure readings are found in the upper half of the tree in Figure 3. They relate the matrix eventuality to a so-called subeventuality. It defines whether both occur at the same time or in sequence, and whether they label points in time or periods.

In example (17) the matrix eventuality is *gehen wir schlafen* ('we go and sleep') and the sub eventuality is *dem Essen* ('the meal'). They do not occur at the same time but the subevent took place before the matrix event.

(17) *Nach dem Essen gehen wir schlafen.*
'After the meal we go and sleep.'

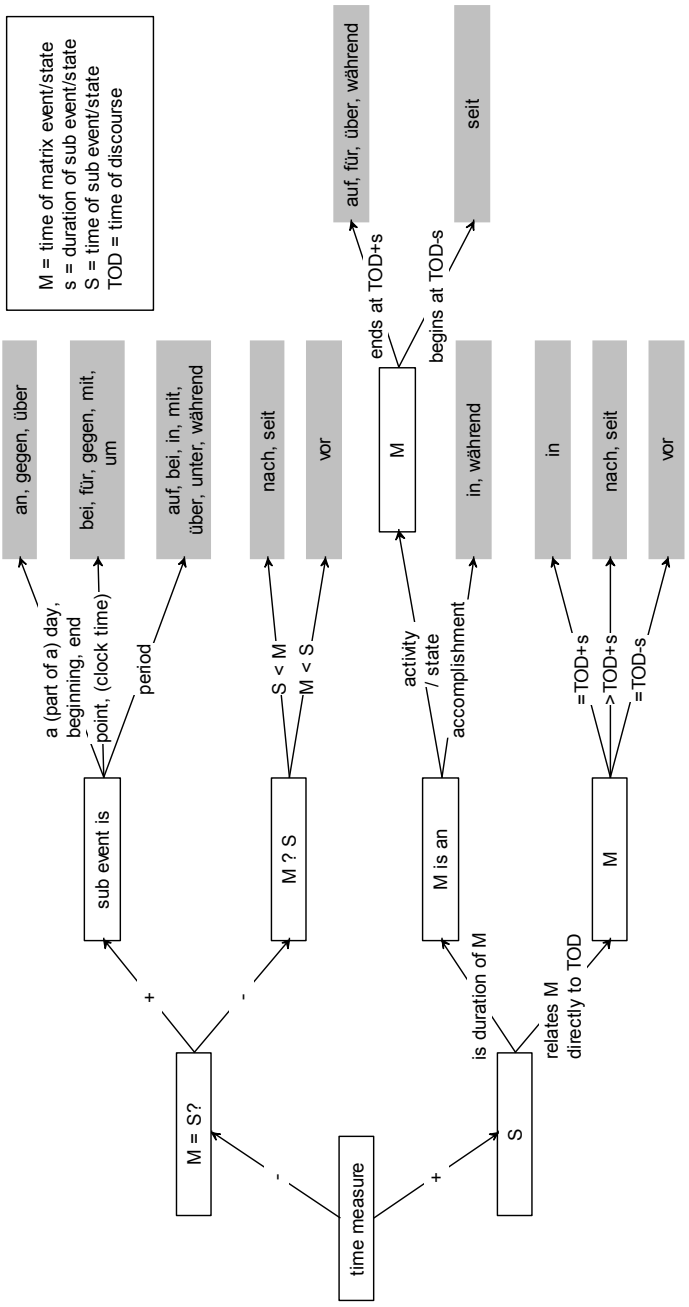

Figure 3. Temporal decision tree

4. A closer look – the preposition *ohne* ('without')

The preposition *ohne* ('without') is peculiar in that it appears more often in PNCs than in PPs – an observation that applies to other languages as well, as has been reported for Dutch, French, and English in de Swart, Le Bruyn, and Zwarts (2010). The preposition shows six discernible top level senses (with 13 subsenses), but two of the six belong to the core interpretation of *ohne,* and are not found with other prepositions apart from *ohne*'s cognate *mit* ('with'). These two senses are (lack of) PRESENCE and (lack of) PARTICIPATION. It should be noted that *ohne* in general expresses a privative meaning, i.e. it expresses the absence or lack of something that is given or present in the analogue interpretation of *mit*. Hence, all senses of *ohne* are to be interpreted as negations of the usual senses. In addition to the two aforementioned interpretations, the following senses of *ohne* can be identified: MODAL (MANNER, CONCURRENT CIRCUMSTANCE, MEDIAL, INSTRUMENTAL), CONDITIONAL (CAUSAL, CONDITIONAL), and (not) INCLUSIVE.

Modal interpretations describe the mode of an event in the broader sense. This might be the way of doing something or a concomitant circumstance of an event. Instrumental interpretations also fall in the modal category – using an instrument is just one manner of doing something.

The subcategory MODAL – MANNER describes the manner of an event (18). It includes intention and is influenceable by an agent. In contrast to this, CONCURRENT CIRCUMSTANCE describes external circumstances or a mode of an event that is not controlled by an agent. There is no intention included in the event (19).

(18) *Eine Mofalenkerin, die ohne Helm unterwegs war,…*
a motorcyclist who without helmet under way has been
'A motorcyclist, riding without a helmet…'

(19) *Das nachgebaute Skelett steht aufrecht und ohne*
The reconstructed skeleton stands upright and without
Makel in der Landschaft.
stain in the landscape
'The copy of the skeleton is standing in the landscape upright and without stain.'

MEDIAL interpretations describe an aid for an event but no instrument in the narrower sense (20). Instruments in the narrower sense are instruments intentionally used by an agent. They must be concrete objects with a clear instrumental character (21). Medial interpretations, however, cover instru-

mentalized events as well as all kinds of means to an end. Often, a replaceability with *via*, *by means of* or *with the aid of* is given.

(20) *Der Text beschreibt einen Geist der sich*
The text describes a mind who REFL
– ohne das äussere Korrektiv allgemeiner
– without the external corrective universal
Werte – die Welt neu erklärt.
values – the world newly defines
'The text describes a mind who newly defines the world for himself without the external corrective influence of universal values.'

(21) *Der italienische Sänger demonstriert, dass seine*
The Italian singer demonstrates that his
sonore Stimme auch ohne Mikrophon wirkt.
sonorous voice also without microphone is effective
'The Italian singer demonstrates that his sonorous voice is also effective without a microphone.'

The interpretation PRESENCE describes the lack of an object or person or of an attribute or characteristic of an object or person (22). Its two sub interpretations are ANALYTIC and SYNTHETIC.[12] Analytic interpretations describe mereological relations, for *ohne* a part of a larger structure that is missing (23). The interpretation SYNTHETIC (24) describes the lack of a defining but not essential attribute. This can be paraphrased with *exclusive*.

(22) *Jugendliche ohne Ausbildungsplatz*
Teenager without apprenticeship position
'Teenager without an apprenticeship position'

(23) *"No See TV" – das Fernsehen ohne Bild*
"No See TV" – the television without picture
'"No See TV" – television without a video signal'

(24) *Messe ohne Predigt*
Mass without sermon
'Mass without a sermon'

The interpretation PARTICIPATION describes the lack of an active or passive participation in an action or event.

(25) *Verhandlungen ohne den Angeklagten*
Hearings without the accused
'Hearings without the accused'

Conditional senses express relations in implication chains, like condition, reason, cause, purpose, and concession. The preposition *ohne* instantiates two conditional subsenses: CAUSAL and CONDITIONAL IN THE NARROWER SENSE.

The causal interpretation gives a cause or reason for something. It can mostly be paraphrased with *because of*.

(26) *Ohne die absolute Parlamentsmehrheit hat sie*
Without the absolute parliamentary majority has she
noch kein klares Mandat.
yet no clear mandate
'Without the absolute parliamentary majority she does not have a clear mandate yet.'

Conditional interpretations can be paraphrased with *if…then* but one must consider the negation in the case of *ohne*.

(27) *Ohne Basiswissen bleibt die Innovation auf der Strecke.*
Without basic knowledge stays the innovation on the route
'Without basic knowledge, innovation falls by the wayside.'

The interpretation INCLUSIVE is restricted to occurrences with numeral adjectives and gives an exclusion of something or someone in the case of *ohne*.

(28) *Ohne den Lehrer waren sie zu zehnt.*
Without the teacher were they in tens
'Without/Excluding the teacher there were ten of them.'

5. Summary

We have developed a scheme for the annotation of preposition senses for 22 simple prepositions in German PNCs and PPs. The annotation will lead to a reference corpus of preposition senses in PPs and PNCs, which will not only be useful as a resource for the further investigation of preposition senses in PPs and PNCs, but can also be used as a gold standard for training automatic methods for preposition sense tagging.

Notes

1. The work reported herein was supported by the DFG (Deutsche Forschungsgemeinschaft) under grant KI-759/5. We would like to thank three anonymous reviewers for their comments.
2. Digitales Wörterbuch der deutschen Sprache, Berlin-Brandenburgische Akademie der Wissenschaften, D-10117 Berlin. http://www.dwds.de
3. In general, one would have preferred to work with corpora from more than one genre. Hence, the present corpus can be considered representative, yet not balanced.
4. The following list gives the categories used: *Motiv, natGegenstand, Gruppe, Nahrung, natPhaenomen, Form, Tier, Zeit, Gefuehl, Kommunikation, Substanz, Ort, Relation, Tops, Koerper, Pflanze, Besitz, Menge, Attribut, Geschehen, Kognition, Artefakt, Mensch*.
5. Nouns that are assigned to more than one top-level category are presumably homonymous or polysemous. We do not disambiguate the nouns but represent this ambiguity by assigning a top-level probability to the noun. If a top-level category value is given a probability smaller than 1, it indicates that more than one reading of the noun is possible, and only one of these readings will be assigned the appropriate top-level category. We can thus be sure that a significant semantic feature (i.e. membership to a top-level category) will be reflected in the classification.
6. In addition to semantic features, the scheme contains the feature governed for prepositions governed by a lexical head. Governed prepositions are often considered to show light semantics only, if at all. But the assignment of the feature governed does not preclude the assignment of additional semantic features if it turns out that the preposition shows a discernible meaning despite its being governed.
7. With regard to example (6), one could argue that temporal interpretations are always present with causal interpretations, as factual causal relations are anchored temporally. But under closer scrutiny, one finds causal interpretations that are predominant to such an extent that a parallel temporal interpretation diminishes. Further analysis might lead to the eventual conclusion that we are not actually observing an ambiguity in example (6), but an inference. Currently, however, we consider such a conclusion premature.
8. The interpretation of *über* may change depending on the presence of a measure phrase: in (7) the hole is (at least partially) hidden by the picture, but in the following example the picture is located two metres above the hole. This is a projective use of the preposition.

   Das Bild hängt zwei Meter über dem Loch.
   'The picture hangs two metres above the hole.'

9. The prepositions in Figures 1 and 2 are not part of the decision tree, but are listed for illustrative purposes only.
10. It has to be considered that a differentiation between local and directional senses does not apply to every spatial interpretation.
11. Without a context the example (28) is ambiguous between an interpretation in which the LO is localized in the proximal region of the RO, and the interpretation adhered to above, and thus would lend itself to multiple annotations.
12. To distinguish the two subinterpretations it can be helpful to replace *ohne* with *mit*. If there is no further modification of an analytic feature the resulting construction with *mit* sounds somehow odd because the specification given is redundant.
    1) *Hand ohne Daumen* 'hand without a thumb'
    2) *?Hand mit Daumen* 'hand with a thumb'
    3) *Hand mit großem Daumen* 'hand with a big thumb'

References

Baldwin, Timothy, John Beavers, Leonoor van der Beek, Francis Bond, Dan Flickinger, and Ivan A. Sag
2006 In Search of a Systematic Treatment of Determinerless PPs. In: Patrick Saint-Dizier (ed.), *Syntax and Semantics of Prepositions*, 163–179. Dordrecht: Springer.

Chiarcos, Christian, Stefanie Dipper, Michael Götze, Ulf Leser, Anke Lüdeling, Julia Ritz, and Manfred Stede
2008 A flexible framework for integrating annotations from different tools and tagsets. *Traitement Automatique des Langues* 49 (2): 217– 246.

Croft, William
1998 Linguistic evidence and mental representations. *Cognitive Linguistics* 9: 151-174.

de Swart, Henriette, Bert Le Bruyn, and Joost Zwarts
2010 Monolingual, multilingual and comparative explorations in countablity. Talk presented at Countability 2010, Ruhr-Universität Bochum.

Dömges, Florian, Tibor Kiss, Antje Müller, and Claudia Roch
2007 Measuring the Productivity of Determinerless PPs. *Proc. of the ACL 2007 Workshop on Prepositions*: 31–37. Prague.

Duden
2002 *Duden. Deutsch als Fremdsprache*. Mannheim: Bibliographisches Institut and F.A. Brockhaus AG.

2005 *Duden. Die Grammatik.* Duden Band 4. Mannheim: Bibliographisches Institut and F.A. Brockhaus AG.

Durrell, Martin, and David Brée

1993 German temporal prepositions from an English perspective. In: Cornelia Zelinsky-Wibbelt (ed.), *The Semantics of Prepositions. From Mental Processing to Natural Language Processing*, 295-325. Berlin/New York: De Gruyter.

Harrell, Frank E.

2001 *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. Springer: New York.

Hartrumpf, Sven, Hermann Helbig, and Rainer Osswald

2003 The Semantically Based Computer Lexicon HaGenLex – Structure and Technological Environment. *Traitement Automatique des Langues* 44(2): 81-105.

Helbig, Gerhard, and Joachim Buscha

2001 *Deutsche Grammatik. Ein Handbuch für den Ausländerunterricht*. Leipzig: Langenscheidt.

Himmelmann, Nikolaus

1998 Regularity in Irregularity: Article Use in Adpositional Phrases. *Linguistic Typology* 2: 315-353.

Kaufmann, Ingrid

1993 Semantic and conceptual aspects of the preposition durch. In: Cornelia Zelinsky-Wibbelt (ed.), *The Semantics of Prepositions. From Mental Processing to Natural Language Processing*, 221-247. Berlin, New York: De Gruyter.

Kempcke, Günter

2000 *Wörterbuch Deutsch als Fremdsprache*. Berlin/New York: De Gruyter.

Kunze, Claudia, and Lothar Lemnitzer

2002 GermaNet - representation, visualization, application. *Proc. of the International Conference on Language Resources and Evaluation 2002*, main conference, Vol V., 1485-1491.

Litkowski, Ken, and Orin Hargraves

2005 The Preposition Project. *Proc. ACL-SIGSEM Workshop on "The Linguistic Dimensions of Prepositions and Their Use in Computational Linguistic Formalism and Applications"*: 171-179. University of Essex - Colchester, United Kingdom.

Müller, Christoph, and Michael Strube

2006 Multi-level annotation of linguistic data with MMAX2. In: Sabine Braun, Kurt Kohn, and Joybrato Mukherjee (eds.), *Corpus Technology and Language Pedagogy: New Resources, New Tools, New Methods*, 197-214. Frankfurt am Main: Verlag Peter Lang GmbH.

Nivre, Joakim

2006 *Inductive Dependency Parsing*. Text, Speech, and Language Technology 34. New York: Springer.

Quinlan, Ross
1986 Induction of Decision Trees. *Machine Learning 1* (1): 81-106.

Retz-Schmidt, Gundula
1988 Various Views on Spatial Prepositions. *AI Magazine* Vol. 9 (2): 95-105.

Saint-Dizier, Patrick
2005 PrepNet: a Framework for Describing Prepositions: Prelimary Investigation Results, *Proc. IWCS 6*, Tilburg.

Schmid, Helmut
1995 Improvements in part-of-speech tagging with an application to German. *Proc. of the EACL SIGDAT Workshop*, Dublin.

Schmid, Helmut, Arne Fitschen, and Ulrich Heid
2004 SMOR: A German computational morphology covering derivation, composition, and inflection. *Proc. of the International Conference on Language Resources and Evaluation 2004*, Lissabon, 1263-1266.

Schmid, Helmut, and Florian Laws
2008 Estimation of conditional probabilities with decision trees and an application to fine-grained POS tagging. *Proc. of COLING 2008*, Manchester.

Schröder, Jochen
1986 *Lexikon deutscher Präpositionen*. Leipzig: VEB Verlag Enzyklopädie.

Stvan, Laurel Smith
1998 The Semantics and Pragmatics of Bare Singular Noun Phrases. Ph. D. diss., Northwestern University, Evanston/Chicago, IL.

Tyler, Andrea, and Vyvyan Evans
2001 Reconsidering prepositional polysemy networks: the case of over. *Language* 77(4): 724-765.

Wiese, Bernd
2004 Über Lokalisationssysteme. Zur Struktur des Inventars der deutschen Lokalpräpositionen mit Berücksichtigung finno-ugrischer Lokalkasussysteme. Ms.: IDS Mannheim.

Witten, Ian, and Frank Eibe
2005 *Data Mining. Practical Machine Learning Tools and Techniques*. 2nd Edition. San Francisco: Morgan Kaufmann Publishers.

Wunderlich, Dieter
1986 Raum und Struktur des Lexikons. In: Hans Georg Bosshard (ed.), *Perspektiven auf Sprache. Interdisziplinäre Beiträge zum Gedenken an Hans Hörmann:* 212-231. Berlin: De Gruyter.