

A Rank-based Distance Measure to Detect Polysemy and to Determine Salient Vector-Space Features for German Prepositions

Maximilian Köper Sabine Schulte im Walde

Institut für Maschinelle Sprachverarbeitung
Universität Stuttgart, Germany
{maximilian.koeper, schulte}@ims.uni-stuttgart.de

Abstract

This paper addresses vector space models of prepositions, a notoriously ambiguous word class. We propose a rank-based distance measure to explore the vector-spatial properties of the ambiguous objects, focusing on two research tasks: (i) to distinguish polysemous from monosemous prepositions in vector space; and (ii) to determine salient vector-space features for a classification of preposition senses. The rank-based measure predicts the polysemy vs. monosemy of prepositions with a precision of up to 88%, and suggests preposition-subcategorised nouns as more salient preposition features than preposition-subcategorising verbs.

Keywords: prepositions, polysemy, vector-space model, German

1. Introduction

Vector space models have become a steadily increasing, integral part of data-intensive lexical semantics over the past 20 years (cf. Turney and Pantel (2010) and Erk (2012) for two recent surveys). They have been exploited in psycholinguistic (Lund and Burgess, 1996) and computational linguistic research (Schütze, 1992), to explore distributional properties of target objects and the notion of “similarity” within a geometric setting.

A respective number of individual vector space approaches have been concerned with sense discrimination (cf. Section 2.). Nevertheless, it is still largely unknown how to identify polysemous objects within a vector space model, and which geometric properties characterise the polysemous objects. For example, are polysemous objects

- (a) *outliers in space*, i.e., are the vectors far away from all vectors whose objects are similar in meaning to one of the senses, or
- (b) *close in space to other polysemous objects*, or
- (c) *close in space to semantically similar objects?*

Insights on these spatial properties could foster research on sense discrimination that is interested in the selection and effect of the features underlying the space vectors: A priori, distributional vectors subsume features across word meanings, but the features are clearly more or less salient with regard to specific object senses.

Our current research on spatial vector properties represents a first step to determine the semantic relevance of sense features, (i) by informing us about the effect of features on the spatial location with regard to other vectors, (ii) by informing us about the potential of similarity measures and clustering algorithms to model polysemy, and (iii) by informing us about the effect of the feature types (frequencies and proportions vs. binary values; information at the syntax-semantics interface; etc.).

More specifically, the current paper is part of a larger framework that systematically explores the vector spatial properties of German prepositions. In a data-intensive distributional framework, we rely on high-dimensional vector spaces to determine the degree of polysemy of the prepositions, and to explore salient features to model their actual meanings.

As in many other languages, German prepositions are notoriously ambiguous, cf. the various senses of the German preposition “nach” in the following examples, where “nach” refers to a temporal (*i*), directional (*ii*), or accordance meaning (*iii*):

- (i) *nach drei Stunden* (**after** three hours)
- (ii) *nach Berlin* (**to** Berlin)
- (iii) *nach Meinung* (**according** to the opinion)

In the following, we address the polysemy of German prepositions with regard to two tasks, (i) to decide whether or not a preposition is polysemous, and (ii) to determine which features provide salient information for the semantic classes of a set of prepositions, including polysemous as well as monosemous prepositions. We systematically explore the vector spatial properties of German prepositions, in mutual dependence with the types of vector features. The core instrument in our explorations is a rank-based distance measure.

Section 2. describes related work on the automatic identification of preposition senses, and with regard to sense discrimination in more general terms. Section 3. introduces our rank-based distance measure, and Section 4. introduces our preposition data and the vector features, before Section 5. describes our actual experiments to distinguish polysemous from monosemous prepositions, and to determine salient preposition features.

2. Related Work

Previous approaches to sense discrimination have primarily aimed to identify regions in vector space that correspond to word senses. For example, Schütze (1998) performed

sense discrimination of ambiguous word tokens, based on their second-order co-occurrence distributions. Erk (2009) presented two variants of defining regions of word meaning in vector spaces, a *prototype model* that introduced a region surrounding the target’s type vector as the representation for a target word, and an *exemplar model* that introduced a k-nearest neighbour classification with weighted average features. Erk and Padó (2010) defined a model where polysemous words activated several word vectors. Reisinger and Mooney (2010) introduced a clustering to produce multiple sense-specific vectors for each word type. Boleda et al. (2012b) compared two models of representing regular polysemy, one with multiple class assignments for multiple senses, and one incorporating classes with polysemy properties. Boleda et al. (2012a) built vector representations of semantic classes from monosemous nouns and then modeled sense alternations using the pairs of these sense representations.

The above approaches are all different to this work, as we do not attempt to distinguish the various senses of a word (vector) yet, but rather aim to identify the spatial positions of polysemous objects in vector space. In previous work (Springorum et al., 2013), we already presented a methodology to identify polysemous German prepositions by exploring their vector spatial properties. We applied two cluster evaluation metrics (the *Silhouette Value* (Kaufman and Rousseeuw, 1990) and a fuzzy version of the *V-Measure* (Rosenberg and Hirschberg, 2007)) as well as various correlations, to exploit hard vs. soft cluster analyses based on Self-Organising Maps. In contrast, the current paper defines and applies a *rank-based distance measure* to explore the preposition vector spaces.

Regarding previous work in computational semantics towards preposition senses, the research on prepositions has in general been enforced by the ACL Special Interest Group on Semantics (ACL-SIGSEM). The SIG has organised a series of workshops on prepositions (Saint-Dizier, 2003; Saint-Dizier, 2006b; Kordoni and Villavicencio, 2005; Arsenijevic et al., 2006; Costello et al., 2007), and a special issue in *Computational Linguistics* (Baldwin et al., 2009). Large-scale projects that were interested in the definition and annotation of preposition senses exist for several languages, most prominently English (*The Preposition Project (TPP)*¹ (Litkowski and Hargraves, 2005)), French (*PrepNet* (Saint-Dizier, 2005; Saint-Dizier, 2006a)) and German² (Müller et al., 2010; Müller et al., 2011; Müller et al., 2012; Müller, 2013)). *The Preposition Project* has also led to the formulation of a *SemEval* task on word sense disambiguation for prepositions (Litkowski and Hargraves, 2007).

Distributional approaches towards preposition meaning and sense distinction have recently started to explore salient preposition features. For example, Baldwin (2006) was one of the first to explicitly use distributional information to determine preposition similarity. Relying on nine English prepositions and a standard vector space (target–content word matrix, based on a window of 5 words) reduced to

100 dimensions by Singular Value Decomposition, he compared the distributional similarity predictions to the preposition lexicon by Dorr (1997) and Roget’s thesaurus classes. Tratz and Hovy (2009) compared several supervised classifiers to assign preposition tokens to their senses. They relied on parse-based features, including the preposition-subcategorising word and the preposition-subcategorised word, and other sentence-relevant functions such as subject, object, other prepositional phrases. They did not only use the lemmas but also WordNet synonyms and hypernyms. Their best results relied on a maximum entropy classifier, outperforming previous approaches on the SemEval 2007 preposition data. Hovy et al. (2010) relied on the best classifier from Tratz and Hovy (2009) and explored context types (fixed windows size vs. selective), the influence of the words in that context (word-based vs. WordNet-based), and the preprocessing method (heuristics vs. parsing) on coarse and fine-grained preposition disambiguation. They found that selective context is better than fixed window size; the governor and the object of the preposition as well as the word directly to the left of the preposition have the highest influence; and that combining different extraction methods works better than either one in isolation. In contrast to these approaches, we focus on two standard sets of features for prepositions, the preposition-subcategorising verbs and the preposition-subcategorised nouns, to demonstrate how the rank-based distance measure determines salient features.

3. Rank-based Distance Measure

The core instrument in this paper to explore the geometrical properties of our high-dimensional vector spaces is a *Rank-based Distance Measure*. More specifically, for a set of 49 German prepositions (cf. Section 4.), we computed the pair-wise distances for each pair of prepositions, as illustrated by Figure 1. The prepositions are represented by high-dimensional vectors, and as measure for calculating the vector similarities/distances we used the standard measure *cosine*.

In a second step, we abstracted the distances to “ranks”, i.e., we determined for each preposition the most similar preposition, the second most similar preposition, etc. The resulting rank information can be visualised by a 49×49 matrix as in Table 1, where “*ab*” is the 27th closest preposition to “*an*”, while “*an*” is only the 42nd closest preposition to “*ab*”. The smaller the rank, the closer are two prepositions and the more similar are their vector representations.

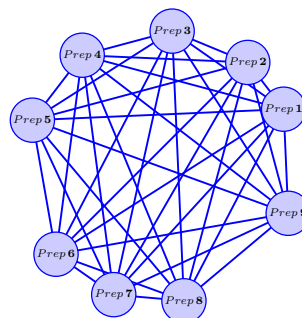


Figure 1: Computing the rank of a preposition requires comparison with all prepositions.

¹<http://www.clres.com/prepositions.html>

²<http://www.linguistics.ruhr-uni-bochum.de/prepositions/>

	<i>ab</i>	<i>an</i>	<i>auf</i>	...
<i>ab</i>	0	27	13	...
<i>an</i>	42	0	23	...
<i>auf</i>	3	17	0	...
...
overall rank	$\frac{42+3+\dots}{48}$	$\frac{27+17+\dots}{48}$	$\frac{13+23+\dots}{48}$...

Table 1: Preposition rank matrix.

	Class	Size	Subclasses
lokal	'local'	27	6
modal	'modal'	24	10
temporal	'temporal'	21	7
kausal	'causal'	5	0
distributiv	'distributive'	6	0
final	'final'	4	0
urheber	'creator'	3	0
konditional	'conditional'	3	0
ersatz	'replacement'	2	0
restriktiv	'restrictive'	2	0
partitiv	'partitive'	2	0
kopulativ	'copulative'	2	0

Table 2: Preposition classes.

In the final step, an overall rank was calculated for each preposition. This rank corresponds to the mean position of a specific preposition in the distance-based sorted lists across all prepositions. Note that the main diagonal in the matrix is always zero and is not considered for the average rank calculation since it corresponds to a comparison of a preposition with itself.

We decided to use the rank information instead of plain distances, because a change in the feature set (vector representation) might increase or decrease individual distances. But to compare distances and average distances across models (feature sets), we convert the distances into ranks.

4. Data

Prepositions and Preposition Senses Our gold standard in terms of preposition senses is the German grammar book by Helbig and Buscha (1998).³ Starting with their class hierarchy, we selected the semantic classes of prepositions that contained more than one preposition. We deleted those prepositions from the classes that appeared less often than 10,000 times in our web corpus containing 880 million words (see below). This selection process resulted in 32 semantic classes covering between 2 and 12 prepositions each (cf. Table 2). The included prepositions exhibit ambiguity rates of 1 (monosemous) up to 6 (cf. Table 3). Each preposition that occurred in more than one class is regarded as polysemous according to the gold standard. In total, 23 out of the 49 preposition types are polysemous (46%).⁴

³Helbig and Buscha (1998) was selected as our semantic resource in accordance with closely related work on German preposition senses (Müller et al., 2012; Müller, 2013).

⁴The preposition classification is identical to that in Springorum et al. (2013) but more fine-grained: We split the three large classes 'local', 'modal' and 'temporal' into 23 sub-classes.

#Senses	#Prepositions
6	1
5	3
4	3
3	11
2	6
1	23

Table 3: Degrees of preposition ambiguity.

Corpus and Distributional Features The distributional features for the German prepositions were induced from the *SdeWaC* corpus (Faaß and Eckart, 2013), a cleaned version of the German web corpus *deWaC* created by the *WaCky* group (Baroni et al., 2009). The *SdeWaC* contains approx. 880 million words and can be downloaded from <http://wacky.sslmit.unibo.it/>.

We focus on two specific feature sets that are expected to provide salient properties towards preposition meaning, (1) the nouns that are subcategorised by the prepositions, and (2) the verbs that subcategorise the prepositions. In examples (i)–(iii), the subcategorised nouns of the preposition *nach* would be *Stunden*, *Berlin*, *Meinung*, and the subcategorising verbs would be *gingen*, *flogen*, *fragten*.

- (i) *Wir gingen nach drei Stunden.* – We left after three hours.
- (ii) *Wir flogen nach Berlin.* – We flew to Berlin.
- (iii) *Wir fragten nach seiner Meinung.* – We asked for his opinion.

The two types of subcategorisation information were extracted from a parsed version of the *SdeWaC* corpus using Bohnet’s MATE dependency parser (Bohnet, 2010).

In the distributional vector spaces in the experiments to follow (cf. Section 5.), each preposition will be associated with a feature vector containing either the subcategorised nouns, or the subcategorising verbs, or the concatenation of the two sets. We restricted the features to the 10,000 most frequent nouns/verbs, that co-occurred with prepositions in the corpus, and compared frequencies and binary values.⁵ The binary values are a function $f(x)$ that maps a frequency x to 1 or 0:

$$f(x) = \begin{cases} 1 & \text{if } x \neq 0 \\ 0 & \text{if } x = 0, \end{cases}$$

5. Experiments

In the following experiments, we apply the rank-based distance measure and address the polysemy of German prepositions with regard to our two tasks, (i) to decide whether or not a preposition is polysemous (Section 5.1.), and (ii) to determine which features provide salient information for the semantic classes of a set of prepositions (Section 5.2.).

⁵We also used *local mutual information (LMI)* values as alternative to frequencies (Evert, 2005), but the results were very similar to the frequency-based results, so we will disregard them.

5.1. Detecting Polysemous Prepositions in Space

The first question we ask is whether we can distinguish polysemous prepositions from monosemous prepositions in vector space. We apply and compare two approaches to this task: (1) our rank-based distance measure, as defined in Section 3.; and (2) a standard k-Means clustering. Before Sections 5.1.2. and 5.1.3. present the classification experiments, Section 5.1.1. motivates the distinction into polysemous vs. monosemous prepositions from a visual point of view.

5.1.1. Visualisation of Prepositions and Polysemy

In order to explore the different contributions of frequency vs. binary feature values, we decided to visualise the preposition positions in two-dimensional feature spaces. We used *t-Distributed Stochastic Neighbor Embedding* (van der Maaten and Hinton, 2008), which aims to preserve both the local and the global distances between all data points while reducing the dimensionality at the same time.

Figures 2 to 4 show how the prepositions are located in different feature spaces. We compare the noun features, the verb features and the concatenation of the noun and verb features, relying on frequencies (left-hand side (a)) vs. binary values (right-hand side (b)). Each dot corresponds to a preposition, and its size corresponds to the degree of polysemy, creating large blue dots for highly polysemous prepositions. Monosemous prepositions are marked red and have a constant size.

The plots suggest that the binary values provide a better basis for the polysemy-monosemy distinction: In the (b) spaces, most of the polysemous prepositions are in one corner of the plot, while most of the monosemous prepositions are in another corner of the plot. This distinction is intuitively stronger for noun features than for verb features, and stronger for verb features than for concatenated features. In the (a) spaces, the polysemous prepositions are also visually located in different areas than the monosemous prepositions, but rather in the central vs. marginal positions of the plot. In sum, we can identify different areas for polysemous vs. monosemous prepositions in space, but the absolute and relative locations of these areas differ with regard to the feature types and with regard to frequency vs. binary values.

5.1.2. Applying the Rank-based Distance Measure

The rank-based distance measure allows us to assign a mean rank score to each preposition, with regard to a specific feature set. In this first experiment, we explore to what extent prepositions with a low mean rank (i.e., prepositions that are close to many other prepositions) are more polysemous than prepositions with a high mean rank (i.e., prepositions that are distant to many other prepositions). The underlying hypothesis is that polysemous prepositions are close to many other prepositions, because of their various senses. Therefore, the mean distance between a polysemous preposition and the other prepositions should be smaller in comparison to a monosemous preposition.

Figures 5 to 7 plot the mean rank values of all prepositions, with regard to our noun feature set (frequencies underlying-

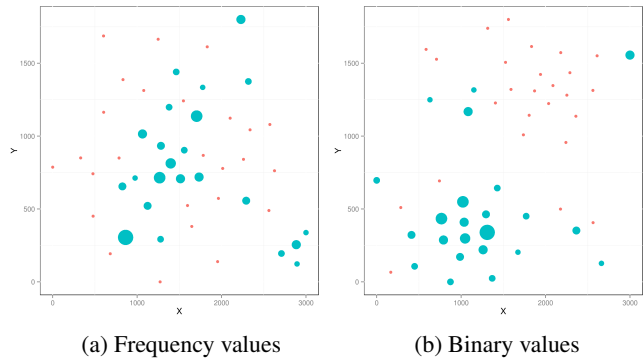


Figure 2: Spaces relying on noun features.

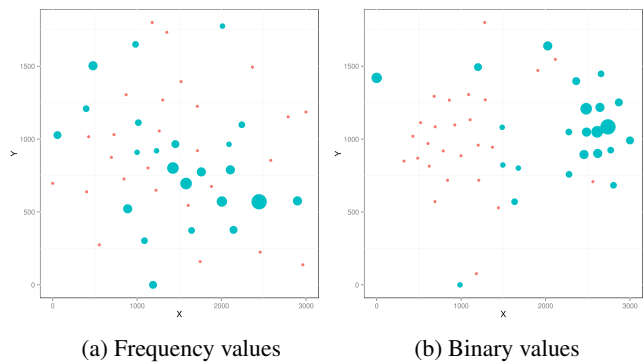


Figure 3: Spaces relying on verb features.

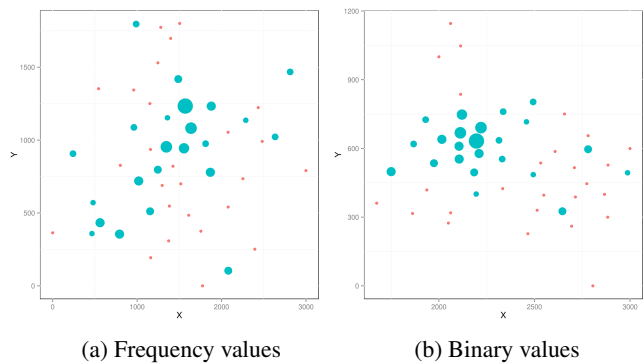


Figure 4: Spaces relying on noun and verb features.

ing Figure 5 and binary values underlying Figure 6) and with regard to our verb feature set (binary values underlying Figure 7). The preposition bars are sorted by the average rank values, and the plot shows polysemous prepositions in white, and monosemous prepositions in dark. The straight blue line represents the median rank (which is suitable to our dataset with an almost 50% split of monosemous vs. polysemous prepositions), and the degree of polysemy of the prepositions (i.e., the number of senses in the gold standard) is indicated by the yellow line.

The figures based on the subcategorised nouns illustrate that prepositions with an above-average rank are indeed likely to be monosemous, while a low average rank indeed seems to be a good indicator for polysemous prepositions. Using binary values seems to improve the separation between monosemous and polysemous prepositions. This

effect is not visible if the prepositions are represented by verb features, neither for plain frequency values (omitted for space reasons) or for binary values (see Figure 7). In sum, the likelihood of a preposition being polysemous decreases with an increasing average rank, thus demonstrating the potential of our rank-based measure to distinguish polysemous from monosemous prepositions. We conclude that prepositions that are close to many other prepositions in vector space are likely to be polysemous.

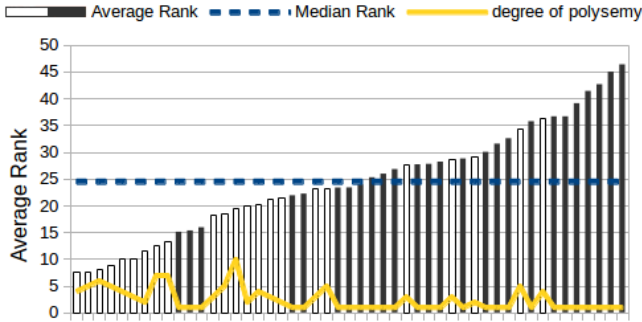


Figure 5: Mean rank values of German prepositions, relying on noun frequency vector spaces.

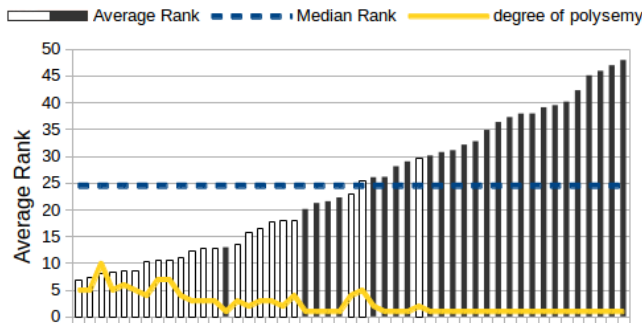


Figure 6: Mean rank values of German prepositions, relying on noun binary vector spaces.

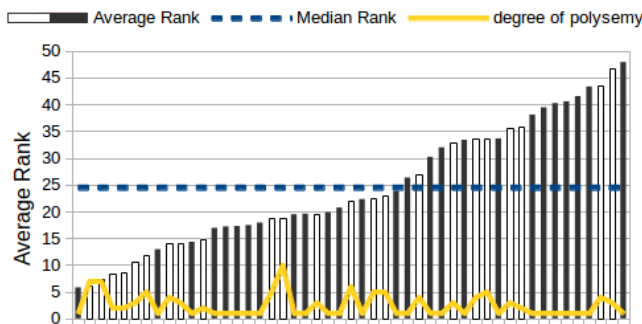


Figure 7: Mean rank values of German prepositions, relying on verb binary vector spaces.

Tables 4 to 6 accompany Figures 5 to 7 and present the precision, recall and f-score values for the two-way classification of the prepositions into polysemous vs. monosemous, when using the median rank as threshold. The results confirm our intuitions: The binary values are more useful to distinguish polysemous from monosemous prepositions, and noun features are more salient than verb features with regard to this task.

	Correct	Wrong	P	R	F
Poly	18	5	78.3%	78.3%	78.3%
Mono	18	8	69.2%	69.2%	69.2%
<i>all</i>	36	13	73.5%		

Table 4: Classification by mean rank (noun frequencies).

	Correct	Wrong	P	R	F
Poly	21	2	91.3%	91.3%	91.3%
Mono	21	5	80.8%	80.8%	80.8%
<i>all</i>	42	7	85.7%		

Table 5: Classification by mean rank (noun binary values).

	Correct	Wrong	P	R	F
Poly	15	8	65.2%	65.2%	65.2%
Mono	12	14	46.2%	46.2%	46.2%
<i>all</i>	27	22	55.1%		

Table 6: Classification by mean rank (verb binary values).

5.1.3. Applying k-Means Clustering

The final approach to distinguish polysemous prepositions from monosemous prepositions applies a k-Means clustering with $k = 2$, to assign the prepositions to the two classes *polysemous* vs. *monosemous*. In order to decide which cluster represents the group of polysemous prepositions, we relied on the mean rank distance information: the cluster with the lower average of the mean rank values was regarded as the class of polysemous prepositions.

We applied the k-Means algorithm provided by the *R Project for Statistical Computing*⁶, and again compared cluster analyses based on the binary noun features, the binary verb features, and their concatenation. We included both the 2-dimensional features from the dimensionality reduction above (cf. Section 5.1.1.), and the original high-dimensional features (10,000/20,000 feature types).

In parallel to the plots, Tables 7 to 9 present the results of the clusterings: We reach a classification precision of up to 87.8%, and all but one setting have precision values $> 80\%$ and significantly outperform the baseline (the majority class), relying on χ^2 with $*p \leq 0.05$. Again, the noun features outperform the verb features. The concatenation of the feature sets achieves the overall best results when using all 20,000 dimensions, but the worst when relying on 2 dimensions only. Across the experiments, the absolute differences are small, varying by a few correct vs. wrong decisions only.

5.1.4. Summary

In this first set of experiments we demonstrated that the rank-based distance measure as well as standard classification approaches are able to distinguish polysemous prepositions from monosemous prepositions. Noun features were more useful than verb features for this task, and binary values were more useful than frequency values. The concatenation of noun and verb features demonstrated an unclear behaviour with regard to the task, and strongly depended on the experiment setup.

⁶<http://www.r-project.org/>

	Correct	Wrong	P	R	F1
Poly (2)	20	5	80.0%	87.0%	83.3%
Mono (2)	21	3	87.5%	80.8%	84.0%
<i>all</i>	41	8	83.7%		
Poly (10k)	19	4	82.6%	82.6%	82.6%
Mono (10k)	23	3	88.5%	88.5%	88.5%
<i>all</i>	42	7	85.7%		

Table 7: Classification by k-Means (noun binary values).

	Correct	Wrong	P	R	F1
Poly (2)	16	3	84.2%	69.6%	76.2%
Mono (2)	24	6	80.0%	92.3%	85.7%
<i>all</i>	40	9	81.6%		
Poly (10k)	18	5	78.3%	78.3%	78.3%
Mono (10k)	22	4	84.6%	84.6%	84.6%
<i>all</i>	40	9	81.6%		

Table 8: Classification by k-Means (verb binary values).

	Correct	Wrong	P	R	F1
Poly (2)	17	10	63.0%	73.9%	68.0%
Mono (2)	16	6	72.7%	61.5%	66.7%
<i>all</i>	33	16	67.3%		
Poly (20k)	19	4	82.6%	82.6%	82.6%
Mono (20k)	24	2	92.3%	92.3%	92.3%
<i>all</i>	43	6	87.8%		

Table 9: Classification by k-Means (noun+verb binaries).

Two-dimensional plots of the dimensionality-reduced feature sets already suggested that polysemous prepositions are in different spatial areas than monosemous prepositions: In frequency spaces, the polysemous vs. monosemous prepositions are visually located in the central vs. marginal positions of the plots. In binary spaces, most of the polysemous prepositions were in one corner of the plot, while most of the monosemous prepositions were in another corner of the plot.

The rank-based distance measure experiments demonstrated that (i) not only it can distinguish polysemous and monosemous prepositions but in addition that (ii) prepositions that are close to many other prepositions in vector space are likely to be polysemous. We interpret this as follows. Since polysemous prepositions subsume the features with regard to several senses, they overlap in many of their features, while monosemous prepositions represent one specific sense represented by a reduced feature set. The k-Means experiments confirmed that (i) the polysemous and monosemous prepositions can be distinguished according to binary noun and/or verb subcategorisation features, and further showed that (ii) the original 10k/20k dimensions are clearly more useful to classify the prepositions than the reduced two-dimensional spaces.

5.2. Exploring Salient Preposition Features

The second question we ask is whether the rank-based distance measure as defined in Section 3. can be used as a tool to explore the impact of varying the preposition features, and to identify the semantically most salient features.

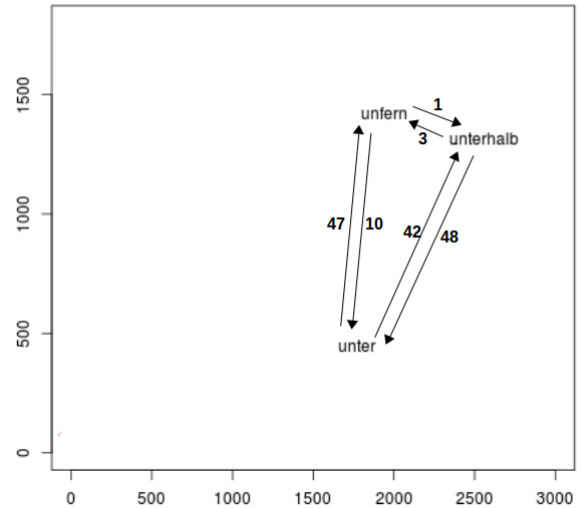


Figure 9: Gold standard class “lokal:Lage” (*local: position*), containing the prepositions “unter” (*under*), “unfern” (*not-far*), and “unterhalb” (*underneath*).

For this task, we do not actually perform a cluster analysis but utilise the gold standard class information about the prepositions to predict the salience of preposition features. I.e., knowing which prepositions belong to which gold standard classes, we computed an average rank score between all prepositions within the same gold standard class, based on a specific feature set. Our hypothesis is that if a feature set represents salient preposition properties, then the average rank distances of those prepositions that belong to the same gold standard class should be low, because prepositions that belong to the same sense class should be highly similar to each other. Our approach is similar to the idea underlying the cluster evaluation metric *Silhouette Value* (Kaufman and Rousseeuw, 1990), which compares the within-cluster distances with the across-cluster distances, to judge about which objects lie well within a cluster and which objects are marginal to a cluster (i.e., potentially misclassified). In our setup, the within-cluster distances are calculated by the rank-distance measure, and comparing the average within-cluster distances allows a direct comparison between different feature sets. We thus utilise our rank-based information measure as a pre-clustering feature analysis.

Figure 8 compares the noun feature set with the verb feature set, across our 32 gold-standard preposition classes. A low average rank of a gold standard class indicates that the respective feature set predicts this class well, as relying on the respective feature set. Overall, it is obvious that the black bars (the noun features) represent more salient features than the white bars (the verb features). Furthermore, the average of the average rank values across all gold standard classes is smaller for nouns than for verbs, confirming our intuitions about the plot: 14.74 vs. 19.90.

We conclude this experiment with an analysis of one class and feature set, by looking at the class “lokal:Lage” (*local: position*) and the noun features. Figure 8 shows that this

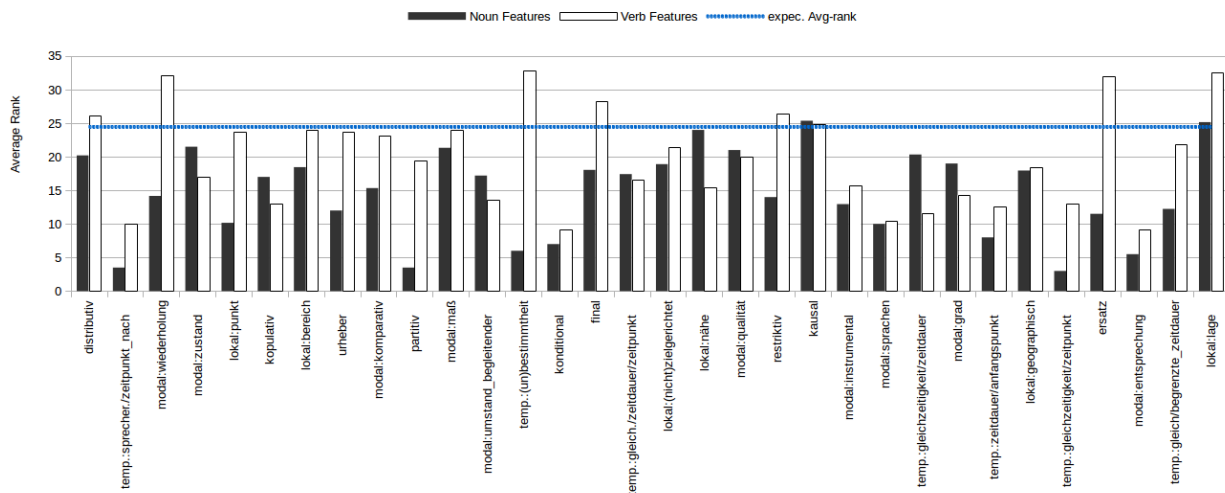


Figure 8: Comparing the salience of noun and verb features using the rank-based distance measure.

class is one of the classes that possess a very high average rank (≈ 25). The class contains only three prepositions according to the gold standard. The feature representation enforces a high similarity between “unfern” (*not far*) and “unterhalb” (*underneath*), but the similarity does not hold for the preposition “unter” (*under*). We can however explain this distance by taking into account that “unter” is polysemous. In addition to the local meaning of “unter” as in the example sentence “*unter dem Tisch – under the table*”, there is also a meaning of “unter” referring to conditions, as in “*unter den gegebenen Umständen – considering the circumstances*”. Figure 9 illustrates the polysemy–monosemy difference very clearly: “unfern” and “unterhalb” are likely to be monosemous while “unter” has a low mean rank, showing its polysemous characteristic. Therefore another semantic class of “unter” is likely to be the reason of the large distance in this case. This could also be the reason why “unfern” \rightarrow “unter” has rank 10 but “unter” \rightarrow “unfern” has only rank 47.

6. Conclusion

This paper presented a rank-based distance measure as an effective way (i) to decide whether or not a preposition is polysemous, and (ii) to determine salient vector-space features for a classification of preposition senses. To this end, we have shown that (i) polysemous prepositions share more features than monosemous prepositions and therefore possess a lower average rank-based distance; we can automatically classify the prepositions into polysemous vs. monosemous, reaching a precision up to 88%; and (ii) preposition-subcategorised nouns are more salient preposition features than preposition-subcategorising verbs.

7. Acknowledgements

The research presented in this paper was funded by the DFG Grant SCHU-2580/2-1 (Maximilian Köper), and the DFG Heisenberg Fellowship SCHU-2580/1-1 (Sabine Schulte im Walde).

8. References

- Boban Arsenijevic, Timothy Baldwin, and Beata Trawinski, editors. 2006. *Proceedings of the 3rd ACL-SIGSEM Workshop on Prepositions*, Trento, Italy.
- Timothy Baldwin, Valia Kordoni, and Aline Villavicencio, editors. 2009. *Computational Linguistics, Volume 35, Number 2, June 2009 - Special Issue on Prepositions*, volume 35. MIT Press.
- Timothy Baldwin. 2006. Distributional Similarity and Preposition Semantics. In Saint-Dizier (Saint-Dizier, 2006b), chapter 1, pages 197–210.
- Marco Baroni, Silvia Bernardini, Adriano Ferraresi, and Eros Zanchetta. 2009. The WaCky Wide Web: A Collection of Very Large Linguistically Processed Web-Crawled Corpora. *Language Resources and Evaluation*, 43(3):209–226.
- Bernd Bohnet. 2010. Top Accuracy and Fast Dependency Parsing is not a Contradiction. In *Proceedings of the 23rd International Conference on Computational Linguistics*, pages 89–97, Beijing, China.
- Gemma Boleda, Sebastian Padó, and Jason Utt. 2012a. Regular Polysemy: A Distributional Model. In *Proceedings of the 1st Joint Conference on Lexical and Computational Semantics*, pages 151–160, Montréal, Canada.
- Gemma Boleda, Sabine Schulte im Walde, and Toni Badia. 2012b. Modelling Regular Polysemy: A Study on the Semantic Classification of Catalan Adjectives. *Computational Linguistics*, 38(3):575–616.
- Fintan Costello, John Kelleher, and Martin Volk, editors. 2007. *Proceedings of the 4th ACL-SIGSEM Workshop on Prepositions*, Prague, Czech Republic.
- Bonnie J. Dorr. 1997. Large-Scale Dictionary Construction for Foreign Language Tutoring and Interlingual Machine Translation. *Machine Translation*, 12(4):271–322.
- Katrin Erk and Sebastian Padó. 2010. Exemplar-based Models for Word Meaning in Context. In *Proceedings of the ACL Conference Short Papers*, pages 92–97, Uppsala, Sweden.

- Katrin Erk. 2009. Representing Words in Regions in Vector Space. In *Proceedings of the 13th Conference on Computational Natural Language Learning*, pages 57–65, Boulder, Colorado.
- Katrin Erk. 2012. Vector Space Models of Word Meaning and Phrase Meaning: A Survey. *Language and Linguistics Compass*, 6(10):635–653.
- Stefan Evert. 2005. *The Statistics of Word Co-Occurrences: Word Pairs and Collocations*. Ph.D. thesis, Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.
- Gertrud Faaß and Kerstin Eckart. 2013. SdeWaC – a Corpus of Parsable Sentences from the Web. In *Proceedings of the International Conference of the German Society for Computational Linguistics and Language Technology*, pages 61–68, Darmstadt, Germany.
- Gerhard Helbig and Joachim Buscha. 1998. *Deutsche Grammatik*. Langenscheidt – Verlag Enzyklopädie, 18th edition.
- Dirk Hovy, Stephen Tratz, and Eduard Hovy. 2010. What’s in a Preposition? Dimensions of Sense Disambiguation for an Interesting Word Class. In *Proceedings of the 23rd International Conference on Computational Linguistics*, pages 454–462, Beijing, China.
- Leonard Kaufman and Peter J. Rousseeuw. 1990. *Finding Groups in Data – An Introduction to Cluster Analysis*. Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York.
- Valia Kordoni and Aline Villavicencio, editors. 2005. *Proceedings of the 2nd ACL-SIGSEM Workshop on The Linguistic Dimensions of Prepositions and their Use in Computational Linguistics Formalisms and Applications*, Colchester, England.
- Kenneth C. Litkowski and Orin Hargraves. 2005. The Preposition Project. In *Proceedings of the 2nd ACL-SIGSEM Workshop on The Linguistic Dimensions of Prepositions and their Use in Computational Linguistics Formalisms and Applications*, pages 171–179, Colchester, England.
- Kenneth C. Litkowski and Orin Hargraves. 2007. SemEval-2007 Task 06: Word-Sense Disambiguation of Prepositions. In *Proceedings of the 4th International Workshop on Semantic Evaluations*, pages 24–29, Prague, Czech Republic.
- Kevin Lund and Curt Burgess. 1996. Producing High-Dimensional Semantic Spaces from Lexical Co-Occurrence. *Behavior Research Methods, Instruments, and Computers*, 28(2):203–208.
- Antje Müller, Olaf Hülscher, Claudia Roch, Katja Keßelmeier, Tobias Stadtfeld, Jan Strunk, and Tibor Kiss. 2010. An Annotation Schema for Preposition Senses in German. In *Proceedings of the ACL Linguistic Annotation Workshop*, Uppsala, Sweden.
- Antje Müller, Claudia Roch, Tobias Stadtfeld, and Tibor Kiss. 2011. Annotating Spatial Interpretations of German Prepositions. In Lisa O’Conner, editor, *Proceedings of the 5th IEEE International Conference on Semantic Computing*, pages 459–466, Stanford, CA.
- Antje Müller, Claudia Roch, Tobias Stadtfeld, and Tibor Kiss. 2012. The Annotation of Preposition Senses in German. In Britta Stolterfoht and Sam Featherston, editors, *Empirical Approaches to Linguistic Theory: Studies in Meaning and Structure*, pages 63–82. Mouton de Gruyter.
- Antje Müller. 2013. *Spatiale Bedeutungen deutscher Präpositionen. Bedeutungsdifferenzierung und Annotation*. Ph.D. thesis, Ruhr-Universität Bochum. Bochumer Linguistische Arbeitsberichte.
- Joseph Reisinger and Raymond J. Mooney. 2010. Multi-Prototype Vector-Space Models of Word Meaning. In *Proceedings of the 11th Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 109–117, Los Angeles, CA.
- Andrew Rosenberg and Julia Hirschberg. 2007. V-Measure: A Conditional Entropy-based External Cluster Evaluation Measure. In *Proceedings of the joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 410–420, Prague, Czech Republic.
- Patrick Saint-Dizier, editor. 2003. *Proceedings of the ACL-SIGSEM Workshop on the Linguistic Dimensions of Prepositions and their Use in Computational Linguistics Formalisms and Applications*, Toulouse, France.
- Patrick Saint-Dizier. 2005. An Overview of PrepNet: Abstract Notions, Frame, and Inferential Patterns. In *Proceedings of the 2nd ACL-SIGSEM Workshop on The Linguistic Dimensions of Prepositions and their Use in Computational Linguistics Formalisms and Applications*, Colchester, England.
- Patrick Saint-Dizier. 2006a. PrepNet: A Multilingual Lexical Description of Prepositions. In *Proceedings of the 5th Conference on Language Resources and Evaluation*, pages 1021–1026, Genoa, Italy.
- Patrick Saint-Dizier, editor. 2006b. *Syntax and Semantics of Prepositions*. Kluwer Academic Publishers, Dordrecht, The Netherlands.
- Hinrich Schütze. 1992. Dimensions of Meaning. In *Proceedings of Supercomputing*, pages 787–796.
- Hinrich Schütze. 1998. Automatic Word Sense Discrimination. *Computational Linguistics*, 24(1):97–123. Special Issue on Word Sense Disambiguation.
- Sylvia Springorum, Sabine Schulte im Walde, and Jason Utt. 2013. Detecting Polysemy in Hard and Soft Cluster Analyses of German Preposition Vector Spaces. In *Proceedings of the 6th International Joint Conference on Natural Language Processing*, pages 632–640, Nagoya, Japan.
- Stephen Tratz and Dirk Hovy. 2009. Disambiguation of Preposition Sense using Linguistically Motivated Features. In *Proceedings of the NAACL-HLT Student Research Workshop and Doctoral Consortium*, pages 96–100, Boulder, CO.
- Peter D. Turney and Patrick Pantel. 2010. From Frequency to Meaning: Vector Space Models of Semantics. *Journal of Artificial Intelligence Research*, 37:141–188.
- Laurens J.P. van der Maaten and Geoffrey E. Hinton. 2008. Visualizing High-Dimensional Data Using t-SNE. *Journal of Machine Learning Research*, 9:2579–2605.