Exploring Features to Identify Semantic Nearest Neighbours: A Case Study on German Particle Verbs

Sabine Schulte im Walde
Computational Linguistics, Saarland University
Saarbrücken, Germany
schulte@coli.uni-sb.de

Abstract

This paper addresses the influence of specific factors in feature selection, in the context of empirical studies on lexical verb semantics. We identify the semantic nearest neighbours of German particle verbs, based on distributional similarity and standard similarity measures, with a focus on features at the syntax-semantics interface. Varying the gold standard explores the types of similarities between the particle verbs and their nearest neighbours. Finally, we apply a Latent Semantic Analysis to check the effect of dimensionality on the semantic choices.

1 Introduction

German particle verbs represent a challenge for statistical NLP: They show specific patterns of behaviour at the syntax-semantics interface, and the semantic relation to their base verbs (transparency vs. opaqueness) is largely non-deterministic. We are interested in automatically inducing semantic classes for German particle verbs to determine the semantically most similar verb groups and predict the compositionality. This paper presents a preliminary step on this path: A complex analysis such as classification requires the definition of multiple parameters, of which the choice of suitable distributional features is a crucial part and should be addressed on a simplified level. In this context, we present an exploration of features to describe German particle verbs. The simplified NLP task for applying the features is to identify the semantic nearest neighbours of the particle verbs, i.e. to identify the German verbs which are semantically most similar. We specifically address the influence of three factors in feature exploration that are important in the context of distributional similarity and have not yet been raised. Future work on classification will capitalise on our insights.

First Issue. We are interested in exploring *the importance of feature selection with respect to a considerable sub-class of verbs*, and choose German particle verbs for a case study. Earlier work concerned with the distributional similarity of verbs such as (McCarthy *et al.* 03; Weeds *et al.* 04) uses standard features (e.g. grammatical dependency relations) and concentrates on the influence of similarity measures. Approaches which address feature selection with respect to semantic classes of verbs such as (Joanis & Stevenson 03; Schulte im Walde 03) explore features for verbs in general to induce classes; so far, only (Merlo & Stevenson 01) address the issue of verb sub-classes, and identify semantic role features to distinguish intransitive verb classes.

Second issue. The evaluation of semantic similarity depends on the definition of a gold standard. However, available resources differ strongly in the types of semantic relations and the number of their instantiations. Previous work has ignored the influence of these evaluation parameters. We vary the gold standard (i) since it allows us to *assess the types of semantic relations* between the particle verbs and their nearest neighbours; and (ii) to get an intuition about the *influence of the gold standard size.*

Third issue. We apply a Latent Semantic Analysis (LSA) to our feature choice, to *explore whether a dimensionality reduction improves the results* by filtering the relevant information from the feature vectors, or makes the results worse by losing relevant information as provided by the feature vectors. LSA was designed to approach synonymy and polysemy of high-dimensional words (Deerwester *et al.* 90), and has been applied successfully to NLP semantic tasks such as measuring word similarity (Landauer & Dumais 97) and particle verb compositionality (Baldwin *et al.* 03). We investigate the difference of high- vs. low-dimensional vectors for our semantic task. Reaching an identical or better result with a reduced number of features would allow us to cut down on the time demands for complex NLP tasks.

2 German Particle Verbs

German particle verbs are productive compositions of a base verb and a prefix particle, whose part of speech varies between open-class nouns, adjectives, and verbs, and closed-class prepositions and adverbs. This work concentrates on prepositional particle verbs, such as *ab-holen*, *an-fangen*, *ein-führen*. Particle verb senses may be transparent (i.e. compositional) or opaque (i.e. non-compositional) with respect to their base verbs. For example, *ab-holen* 'fetch' is transparent with respect to its base verb *holen* 'fetch', *an-fangen* 'begin' is opaque with respect to *fangen* 'catch', and *ein-setzen* has both transparent (e.g. 'insert') and opaque (e.g. 'begin') verb senses with respect to *setzen* 'put/sit (down)'.

German particle verbs may change the syntactic behaviour of their base verbs: the particle can saturate or add an argument to the base verb's argument structure, cf. example (1) from (Lüdeling 01). Theoretical investigations (Stiebels 96) and corpus-based work (Aldinger 04) demonstrate that those changes are quite regular.

(1) Sie *lächelt*.
'She smiles.'
*Sie *lächelt* [$_{NP_{acc}}$ ihre Mutter].
'Sie smiles her mother.'
Sie *lächelt* [$_{NP_{acc}}$ ihre Mutter] *an*.
'Sie smiles her mother at.'

Even though German particle verbs constitute a significant part of the verb lexicon, recent work is mostly devoted to theoretical investigations. To my knowledge, so far only (Aldinger 04) and (Schulte im Walde 04) have addressed German particle verbs from a corpus-based perspective: (Aldinger 04) defines alternation patterns for subcategorisation frames of particle and base verbs; (Schulte im Walde 04) describes the automatic identification and quantitative analysis of German particle verbs. This work relies on the data by (Schulte im Walde 04) and explores features at the syntax-semantics interface to identify the semantically most similar verbs of German particle verbs, a preliminary step towards determining transparency/opaqueness.

Syntax-Semantics Interface Previous work on empirical verb semantics has shown that distributional similarity which models verb behaviour (mainly with reference to subcategorisation, partly including selectional preferences) is a useful indicator of semantic classes, e.g. (Merlo & Stevenson 01; Joanis & Stevenson 03; Korhonen *et al.* 03; Schulte im Walde 03). The underlying hypothesis is that to a certain extent, the lexical meaning of a verb determines its behaviour, particularly with respect to the choice of its arguments, cf. (Levin 93). To check on the behaviour-meaning relationship for the specific case of particle verbs, we use the following distributions to describe verbs.

(1) *syntax* – syntactic frame types
(2) *syntax-pp* – syntactic frame types + PPs
(3) *pref:frame-noun* – selectional preferences; nouns with reference to frame type and slot
(4) *pref:noun* – selectional preferences; nouns without reference to frame type and slot

With descriptions (1) and (2) we follow previous work and assume syntactic frames and prepositional phrases as useful indicators of verb behaviour to induce semantic similarity. Descriptions (3) and (4) take a step away and refer to specific definitions of selectional preferences.

Quantitative Verb Descriptions The quantitative data are from a statistical grammar (Schulte im Walde 03), whose parameters were estimated in an unsupervised training, using 35 million words of a German newspaper corpus. The subcategorisation information was evaluated against dictionary entries, to ensure reliability.

(1) Subcategorisation Frames: The verbs are described by probability distributions over 38 frame types. Possible arguments in the frames are nominative (n), dative (d) and accusative (a) noun phrases, reflexive pronouns (r), prepositional phrases (p), expletive *es* (x), non-finite clauses (i), finite clauses (s), copula constructions (k). For example, the frame type 'nai' indicates the subcategorisation of the obligatory nominative NP (the subject of the clause), an accusative NP (the direct object) and a non-finite clause.

(2) Subcategorisation Frames + PPs: In addition to the syntactic frame information, the frame types distinguish prepositional phrase types by distributing the probability mass of pp-frames over prepositional phrases, according to their corpus frequencies. We consider the 30 most frequent PPs, referred to by case and preposition such as 'Dat.mit', 'Akk.für'. For example, the refined frame type 'nap:Dat.mit' indicates a nominative and an accusative NP, plus a PP with the prepositional head *mit*, requiring dative case.

(3/4) Selectional Preferences: The grammar provides selectional preference information on a fine-grained level: it specifies argument realisations by their lexical heads, with reference to a specific verb-frame-slot combination. For example, the most frequent nominal heads subcategorised in the transitive frame 'na' by the verb *einsetzen* 'insert, start' are for the nominative slot *Polizei* 'police', *Regierung* 'government', *Wehr* 'army', *Bahn* 'railway services', and for the accusative slot *Gas* 'gas', *Mittel* 'means', *Kommission* 'committee', *Waffe* 'weapon'. Our distributions restrict the selectional preferences to frames which are 'relevant' for particle verbs: particle verbs do not show the same diversity of frame usage as non-prefixed verbs but rather focus on intransitive and transitive variants, including adjuncts, cf. (Aldinger 04; Schulte im Walde 04). We construct an *intransitive frame set* where we consider the nominative NPs in the frame types 'n' and 'np', and a *transitive frame set* where we consider the accusative NPs in the frame types 'na', 'nap', 'nad', 'nai', 'nas'. The frame sets therefore include the original frame types 'n' (intransitive) and 'na' (transitive), plus frame types which are their potential extensions, i.e. which add an argument/adjunct to the frame. The distributions *pref:frame-noun* and *pref:noun* refer to the probabilities of nouns in these frame types; the former distribution does encode the reference of the nouns to the specific frame and slot, the latter does not, i.e. frequencies of identical nouns in different frame types and positions are merged and then transfered to probabilities. The underlying assumption for this rather crude simplification refers to the observation that the selectional preferences of particle verbs overlap with those of semantically similar verbs, but not necessarily in identical frames (Schulte im Walde 04). Finally, we define frequency cut-offs, to investigate the influence of the number and frequency range of nouns. The cut-offs are induced from the statistical grammar, referring to the total frequencies of the nouns in the training corpus.

3 Gold Standard Resources

A gold standard in our nearest neighbour classification is applied to two tasks: (1) as source for nearest neighbour candidates, i.e. to define a set of verbs among which the nearest neighbours are chosen, and (2) to evaluate the chosen neighbours on the existence and the type of semantic relation with respect to the particle verbs. Varying the gold standard allows us to assess different types of semantic relations between the particle verbs and their nearest neighbours, and to explore the experiment setup with respect to the size of the gold standard.

GermaNet (GN) (Kunze 00) is the German version of *WordNet* (Fellbaum 98), a lexical semantic taxonomy which organises nouns, verbs, adjectives and adverbs into classes of synonyms, and connects the classes by paradigmatic relations such as antonymy, hypernymy, meronymy, etc. We extracted all particle verbs from GermaNet, a total of 1,856 verbs; for 605 of them GN provides synonyms, for 113 antonyms, and for 1,138 hypernyms. As candidate verbs we extracted all verbs related to any of the particle verbs, a total of 2,338. For comparing different sizes of verb sets, we created a reduced set of particle and candidate verbs (GN-red), by randomly extracting 25 particle verbs each with antonymy, synonymy, and direct and indirect hypernymy relations. We obtained 95 particle and 613 candidate verbs.

Dictionary (DIC): We use one out of numerous monolingual print dictionaries defining synonyms and antonyms (Bulitta & Bulitta 03), and manually copied all synonyms and antonyms for particle verbs which also appeared with a minimum frequency of 500 in the grammar model. This provides us with a total of 63 particle verbs (referring to 18 different base verbs) and 1,645 candidate verbs.

Human Associations (Assoc): In a set of two online web experiments (Melinger & Schulte im Walde 05), we obtained human associations on particle verbs. In the experiments, we asked German native speakers to list spontaneous associations. Each participant provided associations for 50/55 verbs, the total number of verbs in the experiments was 330/100. In the first experiment, 36 particle verbs were included in the 330 verbs; in the second experiment, 76 out of 100 verbs were particle verbs. Each verb was given associations by 46–54 (exp1) and 32–34 (exp2) participants. We use all associated verbs from the experiment as candidates.

Table 1 shows for each gold standard resource the number of particle verbs (pv), the number of candidate verbs (cand), the average number of candidate verbs with a semantic relation to a par-

	pv	cand	avg rel	baseline
GN	1,856	2,338	10	0.43%
GN-red	95	613	12	1.93%
DIC	63	1,645	47	2.84%
Assoc1	36	623	25	4.01%
Assoc2	76	1,040	19	1.84%

Table 1: Verbs and baseline

ticle verb (avg rel), and the average number of related verbs in relation to the number of candidate verbs. The last column represents the baseline for the experiments, since it is the chance of 'guessing' a related verb. Note that the baselines are very low because of the large number of candidate verbs.[1]

4 Semantic Nearest Neighbours

The experiments explore the semantic nearest neighbours of the German particle verbs in the following way. The particle verbs and their candidates are instantiated by probability distributions based on the feature descriptions, and for each particle verb the nearest neighbour is determined. Semantic similarity is calculated by the distance measure *skew divergence*, cf. Equation (3), a variant of the Kullback-Leibler (KL) divergence, cf. Equation (2). The skew divergence measures the distance between the particle verbs v_1 and the candidate verbs v_2 and determines the closest verb. It has been shown an effective measure for distributional similarity (Lee 01). As compared to KL, it tolerates zero values in the distributions, because it smoothes the distances by a weighted average of the two distributions compared. The weight w is set to 0.9.

$$d(v_1, v_2) = D(p \,||\, q) = \sum_i p_i \, log \, \frac{p_i}{q_i} \qquad (2)$$

$$d(v_1, v_2) = D(p \,||\, w * q + (1 - w) * p) \qquad (3)$$

A nearest neighbour is correct if it bears a semantic relation to the particle verb, according to the gold standard. The success of the experiments is measured by precision, the number of correct neighbours in relation to the total number of guesses, i.e. the number of particle verbs in the gold standard. Table 2 presents precision results for the different kinds of distributions. The numbers of features are given in italics. The *pref* distributions refer to the *intransitive frame set* and the *transitive frame set*, and to noun cut-offs of 10, 100, 500 and 1,000. Considering higher cut-offs than 1,000 resulted in lower precision results than in the presented table. The best number per gold standard is printed in bold.

The precision results might appear quite low at first sight; but relating them to the respective baselines (between 0.43% and 4.01%) demonstrates the success of the higher table scores. The syntactic behaviour by itself (distribution: *syntax*) is not much help for identifying semantic nearest neighbours; additional prepositional information improves the results (distribution: *syntax-pp*) only slightly. This insight is especially interesting because it is specific for particle verbs; comparable experiments on non-prefixed verbs demonstrated that *syntax-pp* information is a very useful hint for semantic verb similarity, sometimes even better than selectional preference information, cf. (Joanis & Stevenson 03; Schulte im Walde 03). For the particle verbs, the most successful distributions are clearly the nominal preferences (distributions: *pref:frame-noun* and *pref:noun*), with only slight differences between the cut-offs. Interestingly, the differences between *pref:frame-noun* (with reference to the frame) and *pref:noun* (without reference to the frame) are also minimal.

For *DIC* and *Assoc1*, the differences between the *syntax* and the *pref* variants are significant,[2] while the differences within those groups are not. For the other resources, none of the differences are significant. We conclude that the relevant information in the distributions are the nouns; the references to the argument structure (and, therefore, the functions of the nouns) are of minor importance. Triggered by the observation that the nouns play such a major role in the verb descriptions, we performed a follow-up experiment where we created verb distributions that used all nouns in the window of the respective verbs, disregarding the noun function completely. We used windows of 5, 20 and 50 words to the left and the right of the verbs, and noun frequency cut-offs as before, 10, 100, 500 and 1,000. None of the window distributions reached the results as based on the *pref* distributions; summarising, the relation of the nouns to the verbs is of minor importance (as we said above), but yet it plays a role that only

[1] We realise that our baseline is generous, but it is sufficient, since the baseline is not crucial for our exploration.

[2] All significance tests have been performed with $\chi^2, df = 1, \alpha = 0.05$.

	syntax	syntax-pp	pref:frame-noun				pref:noun			
			10	100	500	1000	10	100	500	1000
	38	*183*	*81,710*	*51,092*	*22,314*	*13,570*	*14,371*	*5,989*	*2,072*	*1,170*
GN	2.13	3.11	8.11	8.77	7.87	7.38	**9.67**	9.59	9.26	8.11
GN-red	6.32	9.47	17.89	17.89	15.79	12.63	**20.00**	**20.00**	**20.00**	15.79
DIC	6.35	12.70	33.33	34.92	**36.51**	34.92	31.75	31.75	33.33	31.75
Assoc1	16.76	22.22	50.00	50.00	50.00	47.22	50.00	52.78	**55.56**	50.00
Assoc2	9.21	11.84	**21.05**	**21.05**	**21.05**	19.74	18.42	15.79	15.79	19.74

Table 2: Precision of *skew* nearest neighbours

nouns with specific functions are included in the distributions. In addition, varying the frequency cut-offs for nouns illustrates that using very high or very low cut-offs (referring to using most vs. only high-frequent nouns) tends to be less successful than keeping to a medium range.

The results with *syntax* and *syntax-pp* show that the syntax-semantics mapping hypothesis does not apply to particle verbs as it does to verbs in general, and we provide the following explanation. *Transparent particle verbs* are semantically similar to their base verbs, but nevertheless do not necessarily agree with them in their syntactic behaviour. (Recall that German particle verbs may change the syntactic behaviour of their base verbs, cf. Section 2.) And since we know that semantically similar non-prefixed verbs show agreement in their behaviour to a large extent, we assume that the frame mismatch transfers from the base verbs to other verbs in their respective semantic class. This means that a syntactic description of transparent particle verbs and semantically similar verbs is not expected to show strong overlap. As a follow-up step on this insight, future work will implement Aldinger's alternation patterns for subcategorisation frames of particle verbs and their base verbs, and investigate whether the syntactic features are more helpful when they include the regular mappings of typical frames. For *opaque particle verbs*, we cannot make strong statements. Since they compositionally represent idioms, we assume that they undergo the syntax-semantic relationship, i.e. that they behave similarly as semantically similar verbs. For both particle verb categories, there is general agreement in the selectional preferences of particle verbs and verbs in the same semantic class, as the *pref* results illustrate.

Comparing the results with respect to the gold standard resources, we observe strong differences; for *Assoc1* we obtain significantly better results than for all other resources except *DIC*. *GN* is significantly worse than most other resources. The differences illustrate the difficulty of the task; it is easier to 'guess' a correct nearest neighbour for *DIC* and *Assoc1* than for the other resources, especially *GN*, cf. Table 1. This has to do with the size of the resources and also with their 'generosity' of providing related verbs. Furthermore, the semantic nearest neighbours allow us to investigate the kinds of semantic relations which are detected. In the GermaNet results, the hypernyms dominate the relations: the neighbours in the best results include 72/68% hypernyms, 23/21% synonyms, and 2/0% antonyms; in some cases the neighbours are defined in GermaNet as both synonyms and hypernyms (e.g. *anfeuern* 'shout encouragement'–*animieren* 'animate' where *animieren* can be a synonym or a hypernym). The fact that the hypernyms dominate the results is not surprising, because they represent 44% of the current GN relations (as compared to 10% synonyms and 1% antonyms), but the proportion is even stronger than in GN. This means that our distributional similarity corresponds rather to the GermaNet hypernym than the GermaNet synonym/antonym definitions. In the dictionary results, we encounter more balanced proportions: 43% synonyms vs. 48% antonyms, plus 2 cases defining a synonymous and antonymous relation at the same time. Still, as compared to 51% and 49% of all encoded relations representing synonyms/antonyms, the proportion of antonyms in our results is slightly stronger than for synonyms. Finally, the human associations demonstrate a more variable picture of semantic verb relations: we find a large number of synonyms or near-synonyms such as *abhalten–veranstalten* 'arrange, organise', *zunehmen–ansteigen* 'increase'; antonyms such as *aufhören–anfangen* 'stop' vs. 'begin', *einpacken–auspacken* 'pack' vs. 'unpack'; but only a few hypernyms such as *aufbrechen–öffnen* 'break open' vs. 'open', *einschärfen–mitteilen* 'inculcate' and 'inform'. In addition, we find verb pairs with backward presupposition, such as *abstürzen–fliegen* 'crash (with respect to a plane)' and 'fly', causal relations such as *einbrocken–auslöffeln* 'get into/out of trouble', *einstürzen–renovieren* 'collapse' and 'renovate', and verbs referring to temporally related script-based events, such as *einschenken–trinken* 'pour' and 'drink', and *umbringen–sterben* 'kill' and 'die'. The examples show that semantic similarity as based on our distributional similarity refers to a variety of semantic relations, which are not covered by the standard manual resources. Future work will address the question of which kinds of features/distributions are associated with which kinds of relations.

5 Latent Semantic Analysis

In a final step of feature exploration, we apply a Latent Semantic Analysis (LSA) to the feature distributions, and then identify the nearest neighbours on basis of the LSA matrix. LSA is a technique for dimensionality reduction which was introduced by (Deerwester *et al.* 90) to address the synonymy and polysemy of high-dimensional word vectors. It performs a Singular-Value Decomposition on high-dimensional vectors: The original *object* $\times$ *feature* matrix $M_{o \times f}$ is represented as the product of three matrixes $O_{o \times k} * S_{k \times k} * F_{k \times f}$, with the diagonal of S as the linearly independent singular vectors. Choosing k considerably smaller than the original number of dimensions f, the matrix O represents a dimensionality reduction of M, approximating a least squares best fit to M. The optimal number of dimensions varies, depending on the task.

The goal of applying LSA to our data is twofold: (i) to explore whether a dimensionality reduction improves the results by using relevant information from the feature vectors, or makes the results worse by losing relevant information provided by the vectors; (ii) reaching an identical or better result with a reduced number of features cuts down on time demands for NLP tasks. As basis for LSA, we use the most successful verb-feature combination from our experiments, with *Assoc1* as gold standard and *pref:noun* and cut-off 500 as feature set. The verb-noun matrix has $623 \times 2,072$ dimensions. As matrix values we use (a) the original verb-noun co-occurrence frequencies f_{vn}, (b) the frequencies transformed to their logarithm: $f_{vn}^{log} = log(f_{vn} + 1)$, and (c) weighted by their *idf (inverse document frequency)* value: $f_{vn}^{idf} = f_{vn} * log(N/n)$, with N the total number

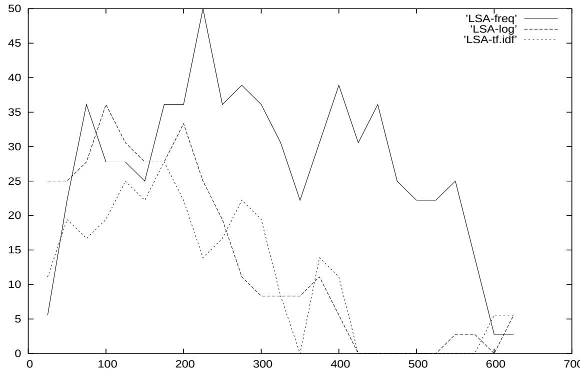

Figure 1: Precision for varying dimensions

of features, and n the number of features a verb co-occurs with. The transformations (b) and (c) are common matrix transformations in LSA, cf. (Deerwester *et al.* 90; Manning & Schütze 99). LSA is applied to the three matrixes, and the feature dimensions are systematically reduced to $k = 25, 50, ..., 2050$. Since the lower-dimensional vectors are not probability distributions, we cannot apply the *skew divergence*; we use the *cosine* of the vectors' angle, another standard measure. For comparison reasons, our previous experiments were repeated with the *cosine*; the precision for *Assoc1/pref:noun500* is 38.89%, non-significantly worse than the *skew divergence* result (55.56%).

Figure 1 shows the precision results of identifying the semantic nearest neighbours with the LSA matrices (a) LSA-freq, (b) LSA-log, and (c) LSA-tf.idf. LSA does improve the results on semantic neighbourhood, but only when performed on the original frequencies, and only with specific dimensionality (225 dimensions). That LSA is most successful on the original frequencies is surprising, since previous work emphasised the importance of feature weighting for LSA, e.g. (Landauer & Dumais 97). The improvement is non-significant. In addition, even the best results with the *cosine* measure for reduced dimensionality are still below the results as obtained with the *skew divergence* for the original probability vectors.

Summarising, in our task of identifying semantic nearest neighbours on the basis of specific verb-noun relations, the task precision suffers from reducing the matrix information by LSA. Only when using the original frequencies and with certain dimensionality, the task-relevant information is preserved. However, for the purpose of time-saving experiments, a single specific reduction is sufficient. In conclusion, it is advisable to apply LSA (and invest the time to find the optimal dimensions) only in cases where succeeding experiments profit from the reduced number of features.

6 Summary

In this paper, we addressed the influence of three factors in feature exploration that are important in the context of distributional semantic similarity. In a case study on German particle verbs the task was to determine their semantic nearest neighbours. First, we showed that the effect of features at the syntax-semantics interface differs for particle verbs as compared to the standard case of non-prefixed verbs. In accordance with theoretical observations, the relevant information in the distributions are the nouns; the references to the argument structure (and, therefore, the functions of the nouns) are of minor importance. Our results illustrate the importance of feature selection with respect to a specific set of data and the task. Second, we varied the gold standard in the evaluation of the nearest neighbours, to check the dependencies on the various types of similarities and the number of correct solutions. We demonstrated that the precision is related to the number of correct choices, which shows how much the size of the gold standard influences the success. Our best result was a precision rate of 55.56%, as compared to a baseline of 4.01%. This result was obtained on a gold standard of human associations in web experiments. It outperforms precision values for gold standard resources encoding only synonymy, antonymy and hypernymy, and illustrates that semantic similarity as based on our distributional similarity refers to a variety of semantic relations, such as temporal and causal relations, which are not covered by the standard manual resources. Finally, a dimensionality reduction by LSA reduced the features to an optimised number of dimensions. In contrast to previous work, we demonstrated that LSA on the original frequency distribution is more appropriate for our data and task than using the weighted versions. But only specific lower-dimensional representations outperform the high-dimensional representations, so it is advisable to apply LSA only in cases where succeeding experiments profit from the reduced number of features. In future work we will investigate which of our insights transfer from the case study to the general case of German verbs.

References

(Aldinger 04) Nadine Aldinger. Towards a Dynamic Lexicon: Predicting the Syntactic Argument Structure of Complex Verbs. In *Proceedings of the 4th International Conference on Language Resources and Evaluation*, 2004.

(Baldwin *et al.* 03) Timothy Baldwin, Colin Bannard, Takaaki Tanaka, and Dominic Widdows. An Empirical Model of Multiword Expression Decomposability. In *Proceedings of the ACL-2003 Workshop on Multiword Expressions*, 2003.

(Bulitta & Bulitta 03) Erich Bulitta and Hildegard Bulitta. *Wörterbuch der Synonyme und Antonyme.* Fischer Taschenbuch Verlag, 2003.

(Deerwester *et al.* 90) Scott Deerwester, Susan T. Dumais, George W. Furnas, Thomas K. Landauer, and Richard Harshman. Indexing by Latent Semantic Analysis. *Journal of the American Society of Information Science*, 1990.

(Fellbaum 98) Christiane Fellbaum, editor. *WordNet – An Electronic Lexical Database.* MIT Press, 1998.

(Joanis & Stevenson 03) Eric Joanis and Suzanne Stevenson. A General Feature Space for Automatic Verb Classification. In *Proceedings of the 10th Conference of the European Chapter of the Association for Computational Linguistics*, 2003.

(Korhonen *et al.* 03) Anna Korhonen, Yuval Krymolowski, and Zvika Marx. Clustering Polysemic Subcategorization Frame Distributions Semantically. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, 2003.

(Kunze 00) Claudia Kunze. Extension and Use of GermaNet, a Lexical-Semantic Database. In *Proceedings of the 2nd International Conference on Language Resources and Evaluation*, 2000.

(Landauer & Dumais 97) Thomas K. Landauer and Susan T. Dumais. A Solution to Plato's Problem: the Latent Semantic Analysis Theory of Acquisition, Induction and Representation of Knowledge. *Psychological Review*, 1997.

(Lee 01) Lillian Lee. On the Effectiveness of the Skew Divergence for Statistical Language Analysis. *Artificial Intelligence and Statistics*, 2001.

(Levin 93) Beth Levin. *English Verb Classes and Alternations.* The University of Chicago Press, 1993.

(Lüdeling 01) Anke Lüdeling. *On German Particle Verbs and Similar Constructions in German.* CSLI Publications, 2001.

(Manning & Schütze 99) Christopher D. Manning and Hinrich Schütze. *Foundations of Statistical Natural Language Processing.* MIT Press, 1999.

(McCarthy *et al.* 03) Diana McCarthy, Bill Keller, and John Carroll. Detecting a Continuum of Compositionality in Phrasal Verbs. In *Proceedings of the ACL-SIGLEX Workshop on Multiword Expressions*, 2003.

(Melinger & Schulte im Walde 05) Alissa Melinger and Sabine Schulte im Walde. Evaluating the Relationships Instantiated by Semantic Associates of Verbs. In *Proceedings of the 27th Annual Conference of the Cognitive Science Society*, 2005.

(Merlo & Stevenson 01) Paola Merlo and Suzanne Stevenson. Automatic Verb Classification Based on Statistical Distributions of Argument Structure. *Computational Linguistics*, 2001.

(Schulte im Walde 03) Sabine Schulte im Walde. *Experiments on the Automatic Induction of German Semantic Verb Classes.* Unpublished PhD thesis, Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart, 2003.

(Schulte im Walde 04) Sabine Schulte im Walde. Identification, Quantitative Description, and Preliminary Distributional Analysis of German Particle Verbs. In *Proceedings of the COLING Workshop on Enhancing and Using Electronic Dictionaries*, 2004.

(Stiebels 96) Barbara Stiebels. *Lexikalische Argumente und Adjunkte. Zum semantischen Beitrag von verbalen Präfixen und Partikeln.* Akademie Verlag, 1996.

(Weeds *et al.* 04) Julie Weeds, David Weir, and Diana McCarthy. Characterising Measures of Lexical Distributional Similarity. In *Proceedings of the 20th International Conference of Computational Linguistics*, 2004.