

Modelling Regular Subcategorization Changes in German Particle Verbs

Stefan Bott Sabine Schulte im Walde

Institut für Maschinelle Sprachverarbeitung

Universität Stuttgart

Pfaffenwaldring 5b, 70569 Stuttgart, Germany

{stefan.bott, schulte}@ims.uni-stuttgart.de

Abstract

German particle verbs are a type of multi word expression which is often compositional with respect to a base verb. If they are compositional they tend to express the same types of semantic arguments, but they do not necessarily express them in the same syntactic subcategorization frame: some arguments may be expressed by differing syntactic subcategorization slots and other arguments may be only implicit in either the base or the particle verb. In this paper we present a method which predicts syntactic slot correspondences between syntactic slots of base and particle verb pairs. We can show that this method can predict subcategorization slot correspondences with a fair degree of success.

1 Introduction

In German, particle verbs (PVs) are a very frequent and productive type of multi word expression. Particle verbs, such as *anstarren* (*to stare at*) in (1-a), are built from a base verb (BV) and a particle. Similar to other multi word expressions, German PVs may show a varying degree of compositionality with respect to the BV and to the particle. But German PVs also have another particularity: if they are compositional, the mapping from semantic arguments to syntactic subcategorization frames may be different between the PV and its corresponding BV.

- (1) a. Die Katze starrt (den Vogel | die Wohnungstür) an.
The cat-N-nom stares (the bird-N-acc | the apartment_door-N-acc) at-PRT.
The cat stares at the (bird | apartment door).
- b. Die Katze starrt auf den Vogel.
The cat-N-nom stares at-P the bird-acc.
- c. Die Katze starrt zur Wohnungstür.
The cat-N-nom stares at-P the apartment_door-dat.

The events expressed with the PV *anstarren* in (1-a) can also be expressed with the BV *starren* in (1-b) and (1-c). But while the argument *Vogel* or *Wohnungstür* is expressed as an accusative object in (1-a) it is expressed as a PP in both (1-b) and (1-c), headed by the preposition *auf* and *zu*, respectively.

Related to this phenomenon, the change in the typical subcategorization frame from the BV to the PV can also lead to an incorporation or an addition of syntactic complements (Stiebels, 1996; Lüdeling, 2001), as illustrated by (2). The BV *bellen* (*to bark*) is strictly intransitive, while the corresponding PV *anbellen* (*to bark at*) is transitive and takes an obligatory accusative object which expresses the person or entity being barked at. This is a case of argument extensions in the PV with respect to its BV. The PV *anschrauben* (*to screw onto*) displays incorporation: it can nearly never select an argument which expresses the location onto which something is screwed, while its BV *schrauben* (*to screw*) requires the expression of the location with a PP.

- (2) a. Der Hund bellt.
The dog-N-nom barks.
- b. Der Hund bellt den Postboten an.
The dog-N-nom barks the postman-N-acc at-PRT.
- c. Der Mechaniker schraubt die Abdeckung auf die Öffnung.
The mechanic-N-nom screws the cover on the opening-N-acc.
- d. Der Mechaniker schraubt die Abdeckung an.
The mechanic-N-nom screws the cover on-PRT.
- (3) a. Der Metzger bringt seiner Frau Blumen.
The butcher brings his wife flowers.
The butcher brings his wife flowers.
- b. Der Metzger bringt das Lämmchen um.
The butcher brings the little lamb PRT.
The butcher assassinates the little lamb.

Finally, if the meaning of the PV is not compositional with respect to the BV, there are no semantic correspondences between subcategorization slots of the PV and the BV. The problem of non-compositionality is illustrated by (3) which uses the PV *umbringen* (to assassinate), which has a totally different meaning from its BV *bringen* (to bring). A successful mapping between the subcategorization slots of both can thus be expected to have a direct relation to the assessment of PV compositionality.

The problem we address here can be called the *syntactic transfer problem*: the subcategorization frame of a BV can be mapped onto a subcategorization frame of the PV, where semantic arguments are not necessarily realized with the same syntactic positions in both of the verbs. A good approximation to this problem is potentially very useful in computational lexicography and other NLP tasks, such as machine translation and information extraction. We also expect it to be helpful to assess other aspects of German particle verbs, such as the prediction of compositionality levels.

In order to tackle the problem of argument slot matching we use a vector space model to represent distributional semantics. We expect that high distributional similarity between two given subcategorization slots taken from a verb pair signals a correspondence of these slots in a pair of subcategorization frames. On the contrary, we expect that low distributional similarity signals that no such correspondence can be established. Further on, if for a given subcategorization slot, either from a BV or a PV, no matching slot can be found in the complementary PV/BV automatically, this typically corresponds to a case of argument incorporation or argument extension.

In short, in this paper we make the following contributions: We present a method of automatically mapping syntactic subcategorization slots of BVs and PVs which is based on distributional semantics and we show that this method can outperform a random baseline with a high level of success.

The rest of this paper is organized as follows: In section 2 we present related work. Section 3 describes our experimental setup, including the method of correspondence prediction, the elicitation of human judgements and the evaluation. Section 4 presents the results which are then discussed in section 5. Section 6 concludes the paper with some final remarks and outlook on future work.

2 Related Work

Particle verbs have been studied from the theoretical perspective and, to a more limited extent, from the aspect of the computational identifiability, predictability of the degree of semantic compositionality (the transparency of their meaning with respect to the meaning of the base verb and the particle) and the semantic classifiability of PVs.

For English, there is work on the automatic extraction of PVs from corpora (Baldwin and Villavicencio, 2002; Baldwin, 2005; Villavicencio, 2005) and the determination of compositionality (McCarthy et al., 2003; Baldwin et al., 2003; Bannard, 2005). To the best of our knowledge Aldinger (2004) is the first work that studies German PVs from a corpus based perspective, with an emphasis on the syntactic behavior and syntactic change. Schulte im Walde (2004; 2005; 2006) presents several preliminary distri-

butional studies to explore salient features at the syntax-semantics interface that determine the semantic nearest neighbours of German PVs. Relying on the insights of those studies, Schulte im Walde (2006) and Hartmann (2008) present preliminary experiments on modelling the subcategorization transfer of German PVs with respect to their BVs, in order to strengthen PV-BV distributional similarity. The main goal for them is to use transfer information in order to predict the degree of semantic compositionality of PVs. Kühner and Schulte im Walde (2010) use unsupervised clustering to determine the degree of compositionality of German PVs, via common PV-BV cluster membership. They are, again, mainly interested in the assessment of compositionality, which is done on the basis of lexical information. They use syntactic information, but only as a filter and for lexical heads as cooccurrence features in order to limit the selected argument slots to certain syntactic functions. They compare different feature configurations and conclude that the best results can be obtained with information stemming from direct objects and PP-objects. The incorporation of syntactic information in the form of dependency arc labels (concatenated with the head nouns) does not yield satisfactory results, putting the syntactic transfer problem in evidence, the problem which we address here. They conclude that an incorporation of syntactic transfer information between BVs and PVs could possibly improve the results. In Bott and Schulte im Walde (2014a) we present a method to assess PV compositionality without recurring to any syntactic features, but we assume that the results of this method could be improved if additional syntactic transfer information was incorporated.

Based on a theoretical study (Springorum, 2011) which explains particle meanings in terms of Discourse Representation Theory (Kamp and Reyle, 1993), Springorum et al. (2012) show that four classes of PVs with the particle *an* can be classified automatically. They take a supervised approach using decision trees. The use of decision trees also allows them to manually inspect and analyze the decisions made by the classifier. As predictive features they use the head nouns of objects, generalized classes of these nouns and PP types. In Bott and Schulte im Walde (2014b) we present an experiment to classify semantic classes of PVs, based on subcategorization information stemming from both the BV and the BV of each BV-PV pair. In this work we use the same gold standard we use here. This experiment is also related to the one presented here in that we assume that the syntactic transfer patterns are quite stable within semantic classes.

3 Experimental Setup

In order to test our hypothesis we selected a set of 32 PVs listed in Fleischer and Barz (2012), including 14 PVs with the particle *an* and 18 with the particle *auf*.¹ We concentrated on two particles here in order to have a small and controlled test bed which allows us to study the syntactic transfers. We selected verbs which we considered to be highly compositional in order to be able to study the correspondence of subcategorization slots. The set contained verbs which have argument slots which are typically realized as different syntactic subcategorizations. The set also contained PVs which show argument incorporation or the introduction of an additional syntactic complement with respect to their BV. We excluded verbs which we could clearly perceive as being polysemous. This set of verbs was processed automatically and presented to human raters, as described below. The test set can be seen in table 1. This test set was already used in Bott and Schulte im Walde (2014b), where it was used as a gold standard for the automatic classification of semantic classes of particle verbs, based on syntactic transfer patterns. The subcategorization patterns listed here are the ones we *expected* to find, so the second and the third row together represent the expected syntactic transfer pattern. The values given in these two columns are a lexicographic presentation of the transfer patterns we expected to find. The task of the system was defined as to find matches between slots from both verbs automatically. The verbs were grouped together in classes which are both semantically similar and also expected to have a similar syntactic behaviour. The labels in the column for the *semantic class* are taken from Fleischer and Barz (2012), but broken down into more detailed classes, such as verbs of *trying*, *gaze* or *sound*. The latter label extensions were

¹Fleischer and Barz list more than 100 PVs for both *an* and *auf*, but they embed this listing in a descriptive text. Some of the verbs listed are very rare or highly ambiguous. Since particle verbs in German are a highly productive paradigm and give rise to many neologisms, compiling a complete list of PVs is nearly impossible.

added by us. In the present work we are not interested in the semantic classes as such, but we assume that the transfer patterns are similar in each semantic class.

3.1 Automatic Classification

Since we wanted to test the predictability of syntactic slot correspondences, we first had to identify the typical elements of the subcategorization frames for both BVs and PVs. In order to do so, we extracted all observable subcategorization patterns from a parsed corpus. Then we selected the 5 most frequent subcategorization patterns for each verb (either BV or PV). These patterns were then broken down into their individual elements. The simple transitive pattern, for example, contained a subject and an accusative object. Since some subordinate structures miss overt subjects and in German all verbs have a subject slot, we always included the subject in the representation of all verbs. The rationale behind this method, which is based on the frequency of subcategorization patterns rather than the frequency of slots, was that we were not interested in subcategorization slots *per se*, but in subcategorization patterns as a typical representation structure in computational lexicography.

Then we built a vector space model for all possible combinations of BV-complements and PV-complements of each BV-PV pair. The dimensions of the vector were instantiated by the head nouns of the syntactic relation in question. The extension in each dimension is equal to the frequency of the head noun in the relevant position. For this experiment no term weighting was applied. Table 2 shows the strongest dimensions for the vectors corresponding to the PP-argument headed by the verbs *heften* (*to attach*) and *anheften* (*to attach to*). The two verbs can be used in quite similar contexts with very similar arguments. Accordingly, the two vectors are similar to each other. Although the two vectors correspond to PP slots headed by the preposition *an*, it can be seen that there is a syntactic transfer from accusative to dative case. Both vectors include head nouns expressing typical places to which things can be attached to, such as a *pin board* (*Pinnwand*), a *wall* (*Wand*) or a *board* (*Brett*). The verb *heften* is frequently found in the idiom *sich an jemandes Ferse heften* (*to attach oneself to someone's heels*, which means *to follow someone closely*), while this idiom cannot be formed with the PV *anheften*. For this reason the dimension for *Ferse* is very strong. This example, especially the vector for *anheften* also shows that the features are often sparsely represented, which presents a problem for our approach.

As a similarity measure we used the cosine distance between two vectors. A variable threshold was applied on the cosine distance to, which serves to separate corresponding subcategorization slots from non-corresponding ones. This is especially important for the detection of argument incorporation or argument extension (cf. example (2)). If, for example, for a given BV slot no PV slot can be found with a cosine value above the threshold, we interpret this as a case of argument extension. On the other hand, a slot from a PV which cannot be match to a slot of its BV is taken to signal argument incorporation. Among the vectors compared to each target subcategorization slot only the one with the highest cosine value was considered as a possible correspondence. Finally, since we want to capture both argument incorporation and argument extension, we computed correspondences for both BVs and PVs separately. Even if this means that most slot pairs are computed twice, this allowed zero-correspondences for slots from both verbs. It theoretically also allows for one-to-many and many-to-one matches, even if we did not exploit them here. We excluded closed class dependencies of verbs, such as negations. We also excluded clausal complements, because they could not be properly represented by our vector extraction method. To get an idea of the lower bound of the outcome values, we used a select-1 baseline. This baseline was obtained by calculating the expected precision and recall for the case that for each subcategorization slot a matching slot from the corresponding other verb is assigned randomly.

As training data we used a lemmatized and tagged version of the SDeWaC corpus (Faaß and Eckart, 2013), a corpus of nearly 885 million words. The corpus was processed with the Mate dependency parser (Bohnet, 2010). The output of this parser represents the syntactic complements of the verbs as labelled arcs. In the case of nominal objects the nominal heads could be directly read of the dependent nodes and the syntactic relation of the arc labels. In the case of PP-complements we read the nominal heads of the nominal node which depends on the preposition which in turn depends on the verb. For the extraction of features we could rely on the database compiled by (Scheible et al., 2013).

Particle	Typical frames for the BV	Typical frames for the PV	Semantic Class	Verbs in Class
an	NPnom +NPacc +PP-an	NPnom +NPacc +PP-an	locative/ relational tying	an binden to tie at an ketten to chain at
	NPnom +PP-zu/in/ nach/auf	NPnom +NPacc	locative/ relational gaze	an blicken to glance at an gucken to look at an starren to stare at
	NPnom +NPacc +PP-mit	NPnom +NPacc +PP-mit	ingressive consump- tion	an brechen start to break an reißen start to tear an schneiden start to cut
	NPnom	NPnom +NPacc	locative/ relational sound	an brüllen to roar at an fauchen to hiss at an meckern to bleat at
	NPnom +NPacc +PP-an	NPnom +NPacc	locative/ relational fixation	an heften to stick at an kleben to glue at an schrauben to screw at
auf	NPnom	NPnom	locative blaze- bubble	auf brodeln to bubble up auf flammen to light up auf lodern to blaze up auf spudeln to bubble up
	NPnom +PP-zu/in/ nach/auf	NPnom	locative gaze	auf blicken to glance up auf schauen to look up auf sehen to look up
	NPnom +NPacc	NPnom +NPacc	locative/ dimensional instigate	auf hetzen to instigate auf scheuchen to rouse
	NPnom +NPacc +PP-auf	NPnom +NPacc	locative/ relational fixation	auf heften to staple on auf kleben to glue on auf pressen to press on
	NPnom	NPnom	ingressive sound	auf brüllen suddenly roar auf heulen suddenly howl auf klingen suddenly sound auf kreischen suddenly scream auf schluchzen suddenly sob auf stöhnen suddenly moan

Table 1: The gold standard classes for the experiments, with subcategorization patterns.

anheften-MO-an-dat	count	heften-MO-an-acc	count
Oberfläche	3	Ferse	154
Gerichtstafel	3	Brust	48
Stelle	2	Revers	43
Schluss	2	Kreuz	32
Unterlage	1	Wand	30
Kirchentüre	1	Spur	12
Brett	1	Tafel	11
Pinnwand	1	Fahne	11
Körper	1	Tür	11
Wand	1	Pinnwand	9
Bauchdecke	1	Kleid	6
Baum	1	Brett	6
Schleimhautzelle	1	Mastbaum	6
Himmel	1	Körper	5
Spur	1	ihn	5
Sphäre	1	Kleidung	5
Wand	1	Oberfläche	5
Spur	1	Stelle	4
Engstelle	1	Baum	4
Pflanze	1	Jacke	4
Protein	1	Mantel	4
Unterseite	1	Teil	3
Zweig	1	Krebszelle	3
Pin-Wand	1	schwarz	3

Table 2: The strongest dimensions for two sample vectors representing subcategorization slots of the verbs *heften* and *anheften*.

3.2 Human rating elicitation

We asked human raters to rate the same examples which the system classified automatically. Each of the pairs of subcategorization slots described in section 3.1 was rated individually. The pairs were always presented in the order <BV-subcategorization-slot,PV-subcategorization-slot> and in visual blocks corresponding to BV subcategorization slots. So the raters could see the possible PV subcategorization slots in direct comparison. The order of blocks was randomized. The raters were asked to judge every pair and rate whether or not they could correspond to a single semantic argument. They were invited to invent example sentences, but because of the length of the annotation session they were not asked to write them down. They were told that, as a criterion for semantic correspondence, each of the verbs in a pair should be usable to describe at least one event or situation they could think of. One annotation example, which did not stem from the set to be rated, was given.

Four human raters were asked to rate examples. All annotators were experts with either a linguistic or NLP background. They were all German native speakers and none of them was otherwise involved in the work presented in this paper. Because of the large size of the data set to be annotated we had to distribute the set over two annotation forms and each annotation form was annotated by two raters. Before the annotation started, one of the authors carried out the same annotation in order to estimate the time needed for each annotation and the level of success which could be expected from the system. Also this annotation was done blindly, without knowledge of the system output, but with a precise knowledge of the task.

The annotation turned out to be much more difficult than we had originally expected. The annotators described the annotation as being hard to perform. This was also reflected by inter annotator agreement; we could only observe a fair agreement, with a Fleiss' Kappa score of 0.31. The agreement between the annotator ratings and the rating by the author was somewhat higher with a Fleiss' Kappa score of 0.44. Some annotators gave detailed feedback, once they had completed the annotation.

4 Results

Table 3 shows the results we obtained. The columns show precision, recall and the harmonic F-score obtained by comparing the system output to the human ratings. We used a precision/recall schema because the task can be seen as the system selecting the most likely slot correspondences from a set of all possible correspondences. So a true positive is obtained if the system selects the same slot that a human rater would select. False positives correspond to a slot selected by the system, which was not chosen by the annotator and a false negative instances are those which are marked by an annotator and not chosen by the system.² Since there was more than one annotators and the annotations differed, we took the sum of true and falls positives and false negatives from all annotators and calculated the scores over this sum. The last column shows the harmonic F-score values we obtained with the annotations produced by one of the authors. The lines represent those threshold values for which the highest precision or F-score could be obtained. The last line represents the baseline. Since a variable threshold was applied there is a trade-off between precision and recall. This is represented in figure 1, which displays the same information as table 3, but in a graphical way.

As expected, the precision improves with higher thresholds, but this comes at the cost of a lower recall. The F-score stays relatively constant. The baseline is quite low, especially the recall. This can be explained because the human raters were free to assign zero-correspondences (i.e. argument incorporations or argument extensions, as exemplified by the examples in (2)) or more than one correspondence per target slot.

5 Discussion

We could observe that the system can predict the correspondences between syntactic subcategorization slots to a fair degree of success and that our method can clearly outperform the baseline. Our hypotheses

²Precision was calculated as $\frac{\{true\ positives\}}{\{true\ positives\} + \{false\ positives\}}$ and recall as $\frac{\{true\ positives\}}{\{true\ positives\} + \{false\ negatives\}}$. The F-score was calculated as $(precision + recall)/2$.

Threshold	Precision	Recall	F-score	Author F-score
0.15	0.48	0.38	0.43	0.68
0.6	0.69	0.21	0.45	0.63
0.85	0.75	0.14	0.44	0.59
baseline	0.38	0.23	0.31	0.31

Table 3: Results of the evaluation in precision, recall and harmonic F-score. The last column represents the pilot annotation carried out by one of the authors.

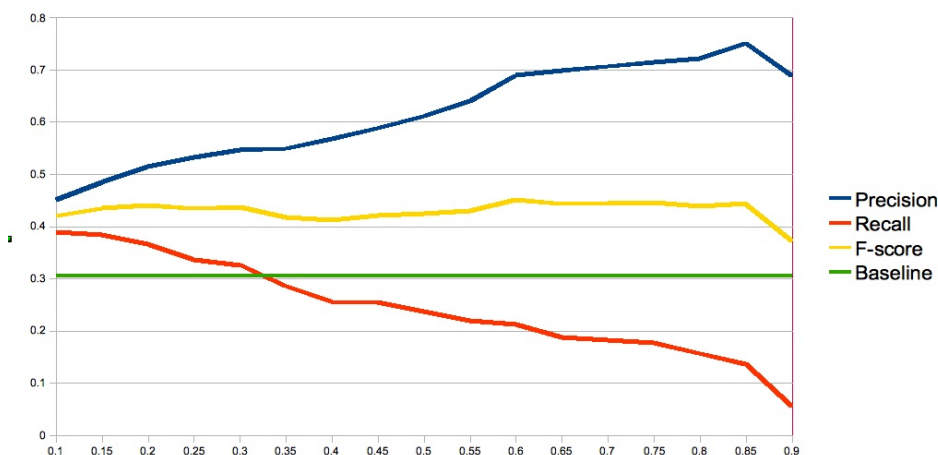


Figure 1: Trade-off between precision and recall. The F-Score remains relatively stable.

that correspondence between subcategorization slots can be predicted to a large degree by distributional semantic similarity can thus be confirmed. On the other hand, the success was not as high as we initially expected. It is surprising that the precision and recall values obtained with the annotations of the human raters are much lower than the values obtained in the initial annotation produced by the author. The author annotation has to be seen as overly optimistic, since it was done with a deeper understanding of the computational task which was to be carried out by the system. Still, this annotation was done blindly. So the big difference we observed is surprising. As already mentioned, the annotators all reported that they found the annotation task difficult to carry out and we attribute the low agreement to this difficulty. The fact that the agreement among different raters was also only fair ($\kappa = 0.31$) hints in the same direction. It must be said that some annotators found the annotation task more difficult than others. Two of the raters reported less annotation difficulty than the remaining. These two annotators were also the ones with most annotation experience and they were both familiar with the topic of particle verbs from a theoretical perspective. When the system output was compared to the ratings of best annotator, a maximum F-score of 0.55 could be achieved, which is still lower than the values obtained in comparison to the author annotation, but much higher than the average of all annotations.

Since some of the annotators gave detailed comments after the annotation was completed, we could detect some problems, which made the annotation difficult, but also extends to the automatic matching. For example, some base verbs have a resultative reading which do not express an *agent* and match the *patient* with the nominal subject position. One such verb is *kleben* (to stick/glue) as exemplified in (4). Accordingly among the strongest dimensions of the vector that represents the subject slot of *kleben*, many nouns appear, which are typical things that stick, such as *band aids* (*Pflaster*), *dough* (*Teig*) and *blood* (*Blut*). The closest vector to the vector for the accusative object vector of *ankleben* was also the accusative object vector of *kleben* (cosine=0.64), but the subject vector was still relatively strong (cosine=0.19).

- (4) a. Gerda klebt den Zettel an die Tür.
Gerda sticks the Note on the door.
- b. Der Zettel klebt an der Tür.
The Note sticks-to the door.

The particle verb *ankleben* can be used to describe the same state of affairs as in (4-a), but not as in (4-b). This is evidently a problem which is hard to solve with our approach because the correspondence of slots from BV and PV interferes with a slot correspondence among different uses of the BV.³

Finally, we found that many of the feature vectors were sparsely instantiated. This can be seen, for example, in the vector that represents the dative PP modifier headed by *an* of the verb *anheften* shown in table 2. The sparsity problem could be remedied by reducing the number of dimensions with the application of some kind of abstraction over the head nouns. For example the concepts of *Tür* (*door*) and *Kirchentür* (*church door*) are strongly related and could be represented in one dimension of the feature vector. The same holds for the concepts of *Pinnwand* (*pin board*), *Wand* (*wall*) and *Tafel* (*blackboard*) and other groups of concepts. With a certain level of abstraction over such concepts, the distance between vectors would also be reduced in case they are sparse. This abstraction is, however, not a trivial problem in itself. The application of lexical ontologies like WordNet (as used by e.g. Springorum et al. (2012)), for example, has the danger of reducing the semantics of head nouns to level of abstraction which is too high, since WordNet has only few top-level categories and few levels of conceptual inheritance.

6 Conclusion and Outlook

We started the work described in this paper out of an interest to approach the syntactic transfer problem of German particle verbs from a computational perspective. We wanted to know in how far the subcategorization slots of a particle verb can be associated with subcategorization slots of a base verbs from which it is derived. The information we used for this matching is based on distributional semantics. We could show that can be done with a good degree of success. From the elicitation of human judgements we learned that the task is also not an easy one for human raters. This also sheds some light on the difficulty of the problem as a computational task.

The work we present here is relevant for computational lexicography. Firstly it can help relate lexical entries of such closely related lexical items as particle verbs and the base verbs they incorporate. The findings we made here may be also applicable to other types of multi word expressions.

In future work we would like to remedy the problem sparse vector representation with the use of abstraction over the head-nouns which will reduce the dimensionality of the feature vector. We also plan to see in how far an automatic clustering of particle verbs into semantic groups can strengthen the prediction of slot correspondences under the assumption that semantically similar verbs tend to undergo the same syntactic transfer. Finally, the problem of syntactic transfer between two elements is also related to the predictability of the degree of compositionality between BV-PV pairs. We are especially interested in this last problem and in future work we plan to investigate in which way subcategorization slot matching can be used as a predictor for compositionality levels.

Acknowledgements

This work was funded by the DFG Research Project "Distributional Approaches to Semantic Relatedness" (Stefan Bott, Sabine Schulte im Walde), and the DFG Heisenberg Fellowship SCHU-2580/1-1 (Sabine Schulte im Walde). We would also like to thank the participants of the human rating experiment.

References

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

³This problem is similar to the prediction of argument realizations in diathesis alternations, such as pairs found in pairs of sentences like "The boy rolled the ball down the hill" vs "the ball rolled down the hill".

- Timothy Baldwin and Aline Villavicencio. 2002. Extracting the Unextractable: A Case Study on Verb Particles. In *Proceedings of the Sixth Conference on Computational Natural Language Learning*, pages 98–104, Taipei, Taiwan.
- Timothy Baldwin, Colin Bannard, Takaaki Tanaka, and Dominic Widdows. 2003. An Empirical Model of Multiword Expression Decomposability. In *Proceedings of the ACL-2003 Workshop on Multiword Expressions: Analysis, Acquisition and Treatment*, pages 89–96, Sapporo, Japan.
- Timothy Baldwin. 2005. Deep Lexical Acquisition of Verb–Particle Constructions. *Computer Speech and Language*, 19:398–414.
- Collin Bannard. 2005. Learning about the Meaning of Verb–Particle Constructions from Corpora. *Computer Speech and Language*, 19:467–478.
- 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.
- Stefan Bott and Sabine Schulte im Walde. 2014a. Optimizing a Distributional Semantic Model for the Prediction of German Particle Verb Compositionality. In *Proceedings of the 9th International Conference on Language Resources and Evaluation*, pages 509–516, Reykjavik, Iceland.
- Stefan Bott and Sabine Schulte im Walde. 2014b. Syntactic Transfer Patterns of German Particle Verbs and their Impact on Lexical Semantics. In *Proceedings of the Third Joint Conference on Lexical and Computational Semantics*, Dublin, Ireland.
- 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*, Darmstadt, Germany.
- Wolfgang Fleischer and Irmhild Barz. 2012. *Wortbildung der deutschen Gegenwartssprache*. Walter de Gruyter, 4th edition.
- Silvana Hartmann. 2008. Einfluss syntaktischer und semantischer Subkategorisierung auf die Kompositionalität von Partikelverben. Studienarbeit. Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart. Supervision: Sabine Schulte im Walde and Hans Kamp.
- Hans Kamp and Uwe Reyle. 1993. *From discourse to logic: Introduction to modeltheoretic semantics of natural language, formal logic and discourse representation theory*. Number 42. Springer.
- Natalie Kühner and Sabine Schulte im Walde. 2010. Determining the Degree of Compositionality of German Particle Verbs by Clustering Approaches. In *Proceedings of the 10th Conference on Natural Language Processing*, pages 47–56, Saarbrücken, Germany.
- Anke Lüdeling. 2001. *On German Particle Verbs and Similar Constructions in German*. Dissertations in Linguistics. CSLI Publications, Stanford, CA.
- Diana McCarthy, Bill Keller, and John Carroll. 2003. Detecting a Continuum of Compositionality in Phrasal Verbs. In *Proceedings of the ACL-SIGLEX Workshop on Multiword Expressions: Analysis, Acquisition and Treatment*, Sapporo, Japan.
- Silke Scheible, Sabine Schulte im Walde, Marion Weller, and Max Kisselew. 2013. A Compact but Linguistically Detailed Database for German Verb Subcategorisation relying on Dependency Parses from a Web Corpus: Tool, Guidelines and Resource. In *Proceedings of the 8th Web as Corpus Workshop*, pages 63–72, Lancaster, UK.
- Sabine Schulte im Walde. 2004. Identification, Quantitative Description, and Preliminary Distributional Analysis of German Particle Verbs. In *Proceedings of the COLING Workshop on Enhancing and Using Electronic Dictionaries*, pages 85–88, Geneva, Switzerland.
- Sabine Schulte im Walde. 2005. Exploring Features to Identify Semantic Nearest Neighbours: A Case Study on German Particle Verbs. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing*, pages 608–614, Borovets, Bulgaria.
- Sabine Schulte im Walde. 2006. The Syntax-Semantics Interface of German Particle Verbs. Panel discussion at the 3rd ACL-SIGSEM Workshop on Prepositions at the 11th Conference of the European Chapter of the Association for Computational Linguistics.

- Sylvia Springorum, Sabine Schulte im Walde, and Antje Roßdeutscher. 2012. Automatic Classification of German *an* Particle Verbs. In *Proceedings of the 8th International Conference on Language Resources and Evaluation*, pages 73–80, Istanbul, Turkey.
- Sylvia Springorum. 2011. DRT-based Analysis of the German Verb Particle "an". *Leuvense Bijdragen*, 97:80–105.
- Barbara Stiebels. 1996. *Lexikalische Argumente und Adjunkte. Zum semantischen Beitrag von verbalen Präfixen und Partikeln*. Akademie Verlag, Berlin.
- Aline Villavicencio. 2005. The Availability of Verb-Particle Constructions in Lexical Resources: How much is enough? *Computer Speech & Language*, 19(4):415–432.