Automatic Classification of German *an* Particle Verbs

Sylvia Springorum, Sabine Schulte im Walde, Antje Roßdeutscher

Institut für Maschinelle Sprachverarbeitung
Universität Stuttgart
{sylvia.springorum, schulte, antje}@ims.uni-stuttgart.de

Abstract

The current study works at the interface of theoretical and computational linguistics to explore the semantic properties of *an* particle verbs, i.e., German particle verbs with the particle *an*. Based on a thorough analysis of the particle verbs from a theoretical point of view, we identified empirical features and performed an automatic semantic classification. A focus of the study was on the mutual profit of theoretical and empirical perspectives with respect to salient semantic properties of the *an* particle verbs: (a) how can we transform the theoretical insights into empirical, corpus-based features, (b) to what extent can we replicate the theoretical classification by a machine learning approach, and (c) can the computational analysis in turn deepen our insights to the semantic properties of the particle verbs? The best classification result of 70% correct class assignments was reached through a GermaNet-based generalization of direct object nouns plus a prepositional phrase feature. These particle verb features in combination with a detailed analysis of the results at the same time confirmed and enlarged our knowledge about salient properties.

Keywords: particle verbs, semantic classification, distributional features

1. Introduction

German particle verbs are a challenge both to theoretical and computational linguistics, as both of their parts (i.e., the particles and the base verbs) may be highly ambiguous. For example, according to a recent study on the particle *an* (Springorum, 2009), 11 different readings have been identified, among them the "partitive" reading including *anschneiden (cut partially)*, *anknabbern (nibble)*, *anreißen (scribe)*, etc.; and the "cumulation" reading including *ansammeln (accumulate)*, *anhäufen (heap)*, *anwachsen (increase)*, etc. In addition, even a comparatively low-frequent base verb such as *schließen* has at least two different senses: *close* and *induce*. Consequently, the meaning of German particle verbs as a combination of ambiguous parts is difficult to determine automatically.

The current study provides a first step into the automatic disambiguation of German *an* particle verbs (PVs). We work at the interface of theoretical and computational linguistics to explore their semantic properties. Driven by a thorough analysis of the particle *an* from a theoretical point of view (Springorum, to appear), we identify empirical features of the *an* particle verbs, to perform an automatic semantic classification. A focus of the study is on the questions (a) how we can transform the theoretical insights into salient corpus-based features, (b) to what extent we can replicate the theoretical classification by a machine learning approach, and (c) whether the computational analysis will in turn deepen our insights to the semantic properties of the PVs.

2. Related Work

The majority of work on German particle verbs is devoted to theoretical investigations, such as Stiebels (1996), Lüdeling (2001), and Dehé et al. (2002). Similarly to the theoretical investigations by Springorum (to appear) who modelled the meanings of *an* particle verbs within Discourse Representation Theory (Kamp and Reyle, 1993), there have been investigations on *ab* particle verbs (Kliche, 2009), and *auf* particle verbs (Lechler and Roßdeutscher, 2009).

To our knowledge, so far only Aldinger (2004), Schulte im Walde (2004), Schulte im Walde (2005), Rehbein and van Genabith (2006), Hartmann (2008), and Kühner and Schulte im Walde (2010) have addressed German particle verbs from a corpus-based perspective, mostly with respect to their idiosyncratic behavior at the syntax-semantics interface, and to determine the compositionality.

For English, there has been more work on computational approaches. Most of them have been devoted to determine the compositionality of particle verbs. For example, Baldwin et al. (2003) defined a word-based model of Latent Semantic Analysis for English particle verbs and their parts, and measured the distributional similarity of the models to evaluate the resulting degree of compositionality against various WordNet-based gold standards. McCarthy et al. (2003) and McCarthy et al. (2007) exploited various measures on syntax-based distributional descriptions as well as selectional preferences, to predict the compositionality of English particle verbs. Bannard (2005) described a distributional approach that compared word-based cooccurrence within the British National Corpus for English particle verbs with those of the respective base verbs and particles. Cook and Stevenson (2006) addressed the compositionality and the meaning of English particle verbs by a distributional model encoding standard verb semantic features (especially subcategorization frame-based information) and PV-specific heuristics. In addition, we find approaches that address an automatic classification of English particle verbs, but as to our knowledge there is none that has a focus on particle verb meaning comparable to ours. In contrast, the classification approaches mostly concentrate on the automatic acquisition of the particle verbs (partly with lexicographic purposes in mind), thus relying on classification approaches to distinguish possible vs. impossible verb–particle combinations. Examples of this research direction are Blaheta and Johnson (2001), Villavicencio (2003), Baldwin (2005), Kim and Baldwin (2006), and Kummerfeld and Curran (2008).

3. Gold Standard Classification

The basic idea of this work is to combine the strengths of theoretical and empirical perspectives on particle verbs to address the meaning classes of German particles and the respective particle verbs. The theoretical perspective contributes a gold standard for the classification experiments to come, which is based on an elaborate case study on the particle *an* (Springorum, to appear). In addition, we collected human judgments on the classification of particle verbs, to create a second gold standard that is more independent of our own intuitions and criteria. The two gold standards are described below.

3.1. Theory-based Gold Standard

The verb particle *an* has about eleven different readings, according to the detailed analysis by Springorum (2009) and Springorum (to appear). For the current study, we chose four of the readings as semantic classes of *an* particle verbs, each containing 10 verbs. The four readings were chosen such that the classes have a disjunct semantics and in addition provide a sufficient number of verbs (except for the partitive class) with several hundred corpus occurrences. Besides this, we aimed for a subset of the original classes where the target classification contains no ambiguity. Table 1 shows the gold standard classification and the corpus frequencies of the verbs. We abbreviate 'partially' as 'part.'.

(i) Topological verbs describe a contact situation that typically occurs between a direct object of the *an* particle verb and an implicit background, cf. Example (1). *an* describes a contact situation between the dog (via the leash) and an unmentioned background. In Example (2), *an* describes a contact situation between the child (subject) and the cat (direct object).

(1) *Maria kettet den Hund an.*
Maria chains the dog.

(2) *Das Kind fasst die Katze an.*
The child touches the cat.

(3) *Der Postbote klebt die Briefmarke an.*
The mailman glues the stamp on.

Figure 1 shows a semantic representation for Example (3) in the framework of Discourse Representation Theory.[1] The meaning of the particle *an* contributes a presupposition with a prestate s_0, meaning that there is no contact between y (the stamp) and v (something which has the function of a background). As a result of the glueing event e' required by the verb, the state changes to s, meaning that there is a contact relation between the stamp and the implicit background.

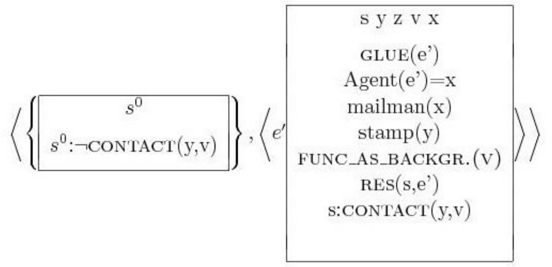

Figure 1: DRS for sentence (3).

(ii) Directional verbs: In most cases, the verb event points from the subject to the direct object of the *an* particle verb, cf. Example (4). This reading has sub readings which typically express an additional communication attempt as in Example (5), where the subject is part of a communication action directed towards the object of the verb.

(4) *Maria starrt ihre Mutter an.*
Maria stares at her mother.

(5) *Maria lächelt ihre Mutter an.*
Maria smiles at her mother.

(iii) Event initiation verbs describe an event initiation where the *an* particle contributes a change from a non-progressive state to a progressive state, cf. Example (6). The whistling, together with the *an*, is responsible for the starting of the game event. In Example (7), the event initiated by *an* is the heating of the oven.

(6) *Der Schiedsrichter pfeift das Spiel an.*
The referee starts the game by whistling.

(7) *Der Großvater heizt den Ofen an.*
The grandfather heats up the oven.

(iv) Partitive verbs: The verb event is performed only on parts of the direct object, like in Example (8), where the sawing event is only accomplished on a part of the plank, or in Example (9) where the mouse performs a *nibble*-event only on a part of the cheese.

(8) *Der Dachdecker sägt das Brett an.*
The roofer partially saws the plank.

(9) *Die Maus knabbert den Käse an.*
The mouse nibbles the cheese partially.

3.1.1. Judgment-based Gold Standard

We collected human judgments on the semantic classes of the *an* particle verbs as follows. Eight linguists, who did not have any previous knowledge about the classification, were given the 40 verbs and asked to classify them into four classes with 10 verbs each. As a starting point, one of the verbs of each class that we considered as central to the meaning of the respective class was provided: *anketten (chain)* as topological verb, *antreiben (activate)* as event initiation verb, *anschreien (scream at)* as directional verb, and *anknabbern (nibble partially)* as partitive verb.
The proportion of agreement between the manual classifi-

[1] In Springorum (2009), such DRS representations are introduced for each of the classes.

TOPOLOGICAL		EVENT INITIATION		DIRECTIONAL		PARTITIVE	
Verb	Freq	Verb	Freq	Verb	Freq	Verb	Freq
anbauen (install)	6642	anblicken (gaze at)	423	anheizen (heat up)	1916	anbohren (drill part.)	176
anbinden (tie up)	2806	angucken (look at)	2415	ankurbeln (boost)	2482	anbraten (roast part.)	2982
anfassen (touch)	3623	anlächeln (smile at)	390	anpfeifen (whistle)	532	anbrechen (broach)	1828
anketten (chain)	492	anpeilen (locate)	846	anregen (instigate)	21194	anknabbern (nibble)	216
anlehnen (lean on)	3066	anreden (address)	1209	anrichten (wreak)	10445	anreißen (scribe)	1033
anmalen (paint)	586	anschreiben (write to)	2819	anspornen (cheer on)	1346	anrösten (toast part.)	405
anschließen (affiliate)	28056	anschreien (scream at)	557	anstiften (incite)	818	ansägen (saw part.)	58
anschnallen (belt on)	600	anstarren (stare at)	1034	anstimmen (intone)	1259	anschneiden (cut part.)	1164
ansiedeln (settle)	14935	anstreben (aspire)	21203	antreiben (activate)	4914	ansengen (scorch)	47
anstreichen (brush)	624	anvisieren (aim for)	1137	anzetteln (plot)	1217	anzahlen (deposit)	49

Table 1: Gold standard classification.

cations was calculated as $p_o = \frac{1}{N} \sum_{i=1}^{k} n_{ii}$ with N the total number of participant decisions, k the number of classes, and n_{ii} the correct classifications. The resulting proportion of agreement $p_o = 0.79$ serves as upper bound for the automatic classification.

4. Classification, Data and Tools

4.1. Empirical Features

Our classification targets (i.e., the *an* particle verbs) are modeled by empirical features that potentially disambiguate the readings. The choice of the empirical features is a key issue regarding our research questions (a) how we could transform the theoretical insights into empirical, corpus-based features, and (b) to what extent we could replicate the theoretical classification by a machine learning approach. Two kinds of empirical features at the syntax-semantics interface have been identified on the basis of the theoretical PV analysis in Springorum (to appear):
(i) German particles and prepositions are historically closely related to each other and provide similarities in their semantics. Therefore, there are also correlations between the particle readings and the prepositional heads of subcategorized prepositional phrases (PPs).[2] Example (10) exemplifies this assumption. Here, the PP with the preposition *an*, i.e., *an dem Fahrradständer*, makes the implicit background of Example (1) explicit.

(10) *Maria kettet den Hund an dem Fahrradständer an.*
Maria chains the dog at the bicycle rack.

(ii) Direct objects subcategorized by the particle verbs are a second feature group we identified as salient, based on the theoretical analysis. For example, concerning directional verbs we expected that the communication attempt enforces persons as direct objects, while topological verbs are more likely to subcategorize physical objects because of their contact semantics.
To reduce the data sparseness concerning the specific semantic direct object heads, in some experiments we performed a semantic generalization of the nominal heads. These generalizations where defined using the hypernymy relation from GermaNet version 5.2 (Kunze, 2000). For the generalization, we used a Java script by Sebastian Padó which takes a list of nouns as input and returns the corresponding hypernyms to each of these nouns as output. The tool allows to choose the level of abstraction of the hypernyms. To provide an example of the generalization, concerning directional verbs with an additional communication attempt, the direct object is very likely to be animated and with consciousness. In contrast, with event initiation verbs the direct object tends to be an event.
(iii) The baseline is defined by using subjects as classification features, which are expected to provide less support for the classification, as the diversity of subject types is typically less strong than the diversity of direct objects and prepositional objects. For example, many of our *an* particle verbs occur with agentive subjects, across the classes.

4.2. Corpus Data

The empirical features for the classification experiments were derived from the German SdeWaC Corpus (Faaß et al., 2010), a German web corpus with about 880 million words. The *SdeWaC* itself is a cleaned version of the *deWaC*, the German web corpus provided by the WaCky community (Baroni et al., 2009).
The features were derived after the corpus was preprocessed by the Tree Tagger (Schmid, 1994) and by the dependency parser *FSPar* (Schiehlen, 2003). *FSPar* explicitly provides ambiguous analyses and does not distinguish between arguments and adjuncts. We decided to use only those syntactic categories that were identified unambiguously within the parses. Concerning PPs, only those which were unambiguously parsed as arguments were taken into account. Concerning subjects and direct objects, only those where the case (and, thus, the function) was identified unambiguously were taken into account.

4.3. Classification Tool

The classification was carried out using the *WEKA* tool (Hall et al., 2009; Witten et al., 2011) with the J48 decision tree algorithm with pruned trees. We used a stratified 10-fold cross-validation. Using decision trees allowed us to retrace the classification which is important for our motivation to obtain insight into the feature selection.

[2] The heads refer to the preposition itself plus case information: (acc) for accusative and (dat) for dative case.

Experiment	Feature	Accuracy		Top.	Ev.I.	Dir.	Par.
Baseline	Subject	13	32.50%	0	3	1	9
Judgments			79.06%				
Exp. 1	PPs	25	62.50%	6	5	5	9
Exp. 2	Objects	11	27.50%	0	0	2	9
Exp. 3	Object Classes	27	67.50%	1	8	8	10
Exp. 4	*an*+Object Classes	28	70.00%	4	7	7	10

Table 2: Classification results.

5. Classification Experiments

The classification task was carried out on the 40 verbs listed in Section 3. The vectors to describe the particle verbs were built from proportions on the basis of corpus frequencies greater than 1. Table 2 shows the best results of the experiments in comparison to the baseline and the human judgments. Experiment 1 used 25 PPs as empirical features, Experiment 2 used 651 direct objects, Experiment 3 used 252 semantic class generalizations of direct objects from GermaNet, and Experiment 4 used the most successful PP type (with prepositional head *an*) plus the semantic classes of direct objects, a total of 253 features. The accuracy lists the amount of the correctly classified verbs, and the columns with the labels Top., Ev.I., Dir. and Par. show the correct decisions with respect to the classes.

We can see that the subjects –as expected– do not provide a reasonable group of features for classifying the *an* particle verbs, and therefore serve as an appropriate baseline. Concerning the four experiments, only the results from Experiment 2 (direct objects) are below the baseline. That this is not due to the actual semantic contribution of the direct objects but rather due to data sparseness can be concluded from the fact that using the generalization of the objects in Experiment 3 is the second successful experiment. Even more, the semantic class generalization in Experiment 3 is also a crucial part of Experiment 4, which adds the most salient PP type *an* to reach the best result, where 70% of the particle verbs are classified correctly. This proportion of correct assignments is only 9% below the upper bound.

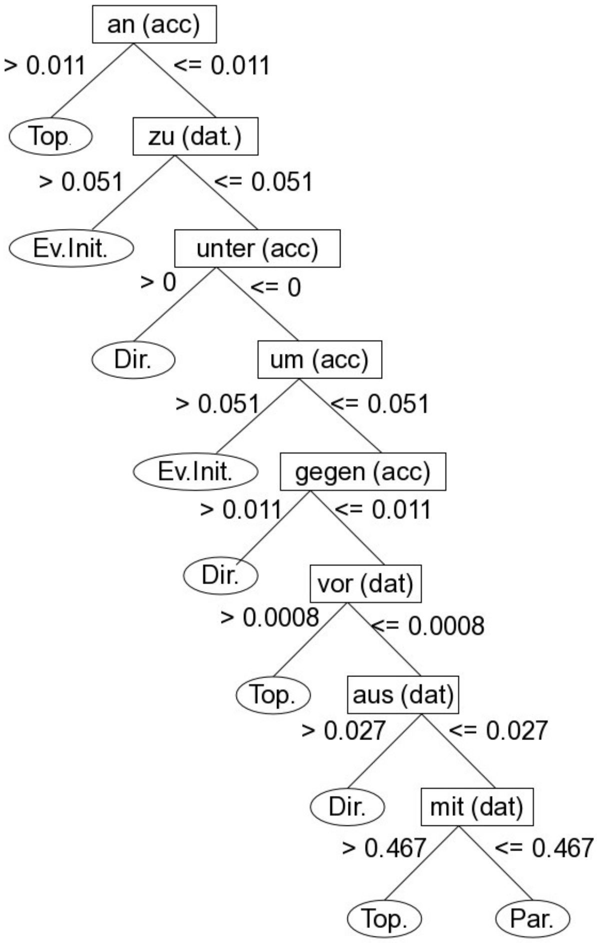

Figure 2: Decision tree of Experiment 1 (PPs).

6. Discussion

In the following, we will look into the classification results in more detail, by analyzing the underlying decision trees and their rules, and by inspecting the actual class assignments of the particle verbs. We concentrate on Experiment 1 (because we are especially interested in the contribution of PPs to classify the *an* readings), and on Experiment 4 (because this was the best experiment).

The decision tree of Experiment 1 with PPs as attributes is presented in Figure 2. Each branch stands for a decision rule, and the leaves represent the classes. Table 3 shows the corresponding class assignments of the particle verbs within a confusion matrix. The rightmost column declares the gold standard of the verbs, and the columns with the letters A-D are the decisions of the classifier. The diagonal shows the correctly classified verbs, in bold font.

The first rule in the decision tree of Experiment 1 checks if there is a PP with an accusative *an* and a proportion greater than 0.011. If this applies then the verb is classified as topological. If the proportion of this feature is smaller or equal than 0.011, the tree is proceeded to the second attribute rule with the dative PP headed by *zu*, etc. The tree demonstrates that the prepositions *an(acc)*, *vor(dat)* and *mit(dat)* are indicators for the topological reading; *zu(dat)* and *um(acc)* for the event initiation reading; and *unter(acc)*, *gegen(acc)* and *aus(dat)* for the directional reading. If none of the 8 selected PP types in the tree are relevant to a verb, it was classified as partitive.

The two top features (*an(acc)* and *zu(dat)*) are the most meaningful in this experiment, as they are responsible for most assignments to the classes Topological and Event Initiation. This corresponds closely to the theoretical con-

A	B	C	D	
anbauen **anbinden** **anfassen** **anketten** **anlehnen** **anschließen**		anmalen	anschnallen ansiedeln anstreichen	A=TOP.
anheizen ankurbeln	**anregen** **anspornen** **anstiften** **anstimmen** **antreiben**	anrichten	anpfeifen anzetteln	B=EV.I.
anstreben		**anblicken** **anpeilen** **anreden** **anschreiben** **anschreien**	angucken anlächeln anstarren anvisieren	C=DIR.
		anreißen	**anbohren** **anbraten** **anbrechen** **anknabbern** **anrösten** **anschneiden** **ansengen** **ansägen** **anzahlen**	D=PAR.

Table 3: Class assignments of Experiment 1.

siderations underlying the feature selection. More specifically, the PP-*an(acc)* indicator is able to assign 6 out of the 10 topological verbs to the correct class. Three of the remaining verbs (*ansiedeln (settle), anstreichen (brush [on]), anschnallen (belt on)*) were analyzed as partitive verbs because they do not have any distinctive PPs to allow the correct class assignments, and *anmalen (paint [on])* was analyzed as directional verb because it occurred with the PP type *gegen(acc)* in the corpus data. The PP-*zu(dat)* indicator was able to assign 5 out of the 10 event initialization verbs to the correct class. This fits into our assumption that a *zu* PP in this *an* class expresses some kind of intention or plan, i.e., it provides an explanation or justification for the initiation of the event, cf. Example (11). The remaining 5 verbs did not appear with *zu(dat)* in the corpus data, even though that PP type was considered the most salient feature for the class. *anpfeifen (start sth. by whistling)* and *anzetteln (plot)* have sparse data vectors, so they were classified as partitive verbs (i.e., the class with no distinctive PPs).

(11) *Der Film regt die Leute zum Denken an.*
The movie makes the audience think.

Concerning the directional verbs, the class assignments were caused by the PP types *gegen(acc), unter(acc)* and *aus(dat)*. Only *gegen(acc)* had been considered a salient feature for the class in the theoretical analyses. The underlying corpus data confirmed our assumption that *gegen(acc)* is a directed opposition (e.g., subcategorizing war, experience and home sickness). In addition, we found that *aus(dat)* described a source for *anblicken, angucken (look at), anstarren (stare at)*, subcategorizing mainly *Auge (eye)*. As mentioned before, the partitive class does not have any PP features for disambiguation. Thus, we consider it as a remainder class, containing all verbs which do not have distinctive PP types and thus cannot be assigned to any of the other classes.

The decision tree of Experiment 4 with direct object classes and *an(acc)* as attributes is presented in Figure 3. Table 4 shows the corresponding class assignments of the particle verbs.

The decision tree in Figure 3 is noticeably small, thus selecting only a small number of the most indicative features. The top of the tree is dominated by two semantic classes, so the direct object information (as generalized class information) is obviously important. In addition, we also find the only PP type we added to the feature set within the small tree, *an(acc)*, so the decision to add this feature has also been justified. In the following, we go into more details concerning the choice of features.

At the top of the tree, the semantic class Event is identified as an effective feature for the Event Initiation class, assigning 7 out of the 10 verbs to the correct class (cf. Table 4). This corresponds to the theoretical observation that the semantics of the verbs operates on events which are often introduced through a direct object. The event initiation verbs *anspornen (cheer on), anstiften (incite)* and *anstimmen (intone)* are wrongly classified as directional verbs because they usually take *Higher life form* as an object (cf.

A	B	C	D	
anbauen **anketten** **anlehnen** **anschließen**	ansiedeln	anbinden anfassen anmalen	anschnallen anstreichen	A=Top.
	anheizen **ankurbeln** **anpfeifen** **anregen** **anrichten** **antreiben** **anzetteln**	anspornen anstiften anstimmen		B=Ev.I.
anstreben	angucken anvisieren	**anblicken** **anlächeln** **anpeilen** **anreden** **anschreiben** **anschreien** **anstarren**		C=Dir.
			anbohren **anbraten** **anbrechen** **anknabbern** **anreißen** **anrösten** **anschneiden** **ansengen** **ansägen** **anzahlen**	D=Par.

Table 4: Class assignments of Experiment 4.

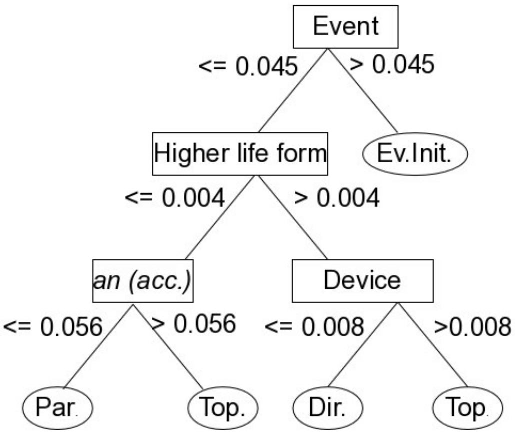

Figure 3: Decision tree of Experiment 4 (*an*-PP and direct object classes).

Example (12)), which is however used as a main feature for the Directional class. The event in these cases is expressed by a PP with *zu(dat)*, like in Example (11), where the initiated event is the thinking action of the audience. Even though the classification of the three event initiation verbs was wrong, the analysis of the decision tree made us realize that the definition in the gold standard concerning the verb *anspornen (incite)* was not sufficient, as there is also a directional component in the semantics of *anspornen (incite)*, cf. Example (12).

(12) *Der Chef spornt seine Mitarbeiter zu Höchstleistungen an.*
The boss incites his employees to work more efficiently.

If we look at the assignments for the directional verbs, we see that the decisions coming from the features with *Higher life form* and *Device* as objects work well, with 7 of out 10 correctly classified verbs. All correctly classified verbs describe a communication between individuals. The verbs *anstarren (stare at)* and *anblicken (gaze at)* are not directly conceptualized as a communication act but their meaning is close to this idea. To gaze at an inanimate thing seems to be odd. In contrast, the verb *anstarren (stare at)* can come along with an inanimate object, but this usage seems to be restricted to few situations like watching TV or starring at a wall. The verbs *anstreben (aspire), angucken (look at)* and *anvisieren (aim for)* do not primarily subcategorize higher life forms. *angucken (look at)* for example often occurs with a game as direct object, which is generalized as an event. In this case, a more suitable translation of the verb is *watch*. Many objects of *anstreben (aspire)* and *anvisieren (aim for)* can be summarized as aims such as an apprenticeship or success, so the semantic analysis must be amplified with an intensional component.

The topological verbs *anbauen (install)*, *anschließen (affiliate)*, *anketten (chain)* and *anlehnen (lean against)* do usually not occur with events. Therefore, they all get to the node *Higher life form*, where only *anschließen (affiliate)* holds this feature and thus is tested on whether the device indicator is able to identify it as a topological verb. The classification of *anlehnen (lean gainst), anbauen (install)* and *anketten (chain)* took place because of the prepositional feature *an(acc)*. The verb *ansiedeln (settle)* was classified wrongly as event initiation verb because of the first decision rule *Event*, and *anmalen (paint), anbinden (tie up)* and *anfassen (touch)* were classified wrongly as directional because they potentially have a *Higher life form* as direct object. The verbs *anstreichen (brush)* and *anschnallen (belt on)* do not have any of the features relevant for the classification and therefore ended up in the Partitive class. Again, the Partitive class does not have any features for disambiguation, and we consider it as a remainder class. All partitive verbs are actually assigned to this class because they do not provide any features for another allocation.

In sum, the feature Event is an indicator for the event initiation reading, the feature *Higher life form* is a feature for the directional reading, and the PP with *an(acc)* is an indicator for a topological reading even though only three verbs can be recognized by the preposition *an*, but there are nearly no other verbs which come along with this preposition *an(acc)*. So it is a feature for the topological reading but not sufficient to disambiguate all verbs belonging to this class. The feature *Device* is only dedicated to one verb, but it shows a tendency which also emerged in experiments not presented here: It is very likely that topological verbs subcategorize direct objects which are artifacts. This makes sense because the semantics of the verbs belonging to this class contains a contact relation which is usually given between physical objects.

7. Conclusion

We presented a study on the automatic classification of German *an* particle verbs, with an explicit interest concerning the interface of theoretical and computational perspectives on the semantic properties of the particles and the particle verbs. The best classification result was reached by a GermaNet-based generalization of direct object nouns in combination with the most successful prepositional phrase feature. This combination of features largely corresponds to the linguistic intuitions based on our former linguistic studies. Thus, we succeeded in transforming our theoretical insights about the semantic particle classes into empirical, corpus-based features, and were able to replicate the semantic classification by a machine learning approach to an extent of 70%, which is only 9% below the upper bound of human judgments.

We also succeeded in the goal to deepen our insights of the semantic properties of the particle verbs through the computational analysis. For example, the verbs *anspornen (cheer on)* and *anstiften (incite)* (cf. Examples (12) and (13)) were characterized as Event Initiation, but we found that this classification is not sufficient because they also have a directional component (i.e., communication) and therefore occur with an object with consciousness. Hence, the Event Initiation refers to a communication with a person to make her act. Therefore the semantics of the verbs should take into account a combination of both meanings.

(13) *Er stiftet den Bruder zu Unfug an.*
He incites the brother to rag.

We also learned from the empirical experiments that *anvisieren (aim for)* and *anstreben (aspire)* should be treated as a subclass of the directional reading which additionally encodes intensionality with respect to future plans. As already mentioned in the previous chapter, these verbs subcategorize direct objects which are events. Example (14) can be paraphrased by Example (15), which demonstrates that the event descriptions here are plan descriptions.

(14) *eine Lehre anvisieren*
to aim for an apprenticeship

(15) *anvisieren eine Lehre zu machen*
to aim making an apprenticeship

In sum, the explicit interface between theory and computation provided more insights into the semantic properties of *an* particles and particle verbs than only one perspective would have given.

8. Acknowledgements

Sylvia Springorum and Antje Roßdeutscher have been supported by the DFG Collaborative Research Center 732. Sabine Schulte im Walde is funded by the DFG Heisenberg Fellowship SCHU-2580/1-1.

9. 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.

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.

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.

Don Blaheta and Mark Johnson. 2001. Unsupervised learning of multi-word verbs. In *ACL Workshop on Collocation*, pages 54–60.

Paul Cook and Suzanne Stevenson. 2006. Classifying Particle Semantics in English Verb-Particle Constructions. In *Proceedings of the ACL/COLING Workshop on Multiword Expressions: Identifying and Exploiting Underlying Properties*, Sydney, Australia.

Nicole Dehé, Ray Jackendoff, Andrew McIntyre, and Silke Urban, editors. 2002. *Verb-Particle Explorations*. Number 1 in Interface Explorations. Mouton de Gruyter, Berlin.

Gertrud Faaß, Ulrich Heid, and Helmut Schmid. 2010. Design and application of a gold standard for morphological analysis: Smor as an example of morphological evaluation. In Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, Mike Rosner, and Daniel Tapias, editors, *LREC*. European Language Resources Association.

Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. 2009. The weka data mining software: An update. *SIGKDD Explorations*, 11(1).

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*. Kluwer. Dordrecht.

Su Nam Kim and Timothy Baldwin. 2006. Interpreting Semantic Relations in Noun Compounds via Verb Semantics. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistics*, pages 491–498, Sydney, Australia.

Fritz Kliche. 2009. Zur Semantik der Partikelverben auf *ab*. Eine Studie im Rahmen der Diskursepräentationstheorie. Master's thesis, Universität Tübingen.

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.

Jonathan K. Kummerfeld and James R. Curran. 2008. Classification of verb particle constructions with the google web1t corpus. In *Proceedings of the Australasian Language Technology Association Workshop 2008*, volume 6, pages 55–63, Hobart, Australia, December.

Claudia Kunze. 2000. Extension and Use of GermaNet, a Lexical-Semantic Database. In *Proceedings of the 2nd International Conference on Language Resources and Evaluation*, pages 999–1002, Athens, Greece.

Andrea Lechler and Antje Roßdeutscher. 2009. German Particle Verbs with *auf*. Reconstructing their Composition in a DRT-based Framework. *Linguistische Berichte*, 220.

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.

Diana McCarthy, Sriram Venkatapathy, and Aravind K. Joshi. 2007. Detecting Compositionality of Verb-Object Combinations using Selectional Preferences. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 369–379.

Ines Rehbein and Josef van Genabith. 2006. German Particle Verbs and Pleonastic Prepositions. In *Proceedings of the 3rd ACL-SIGSEM Workshop on Prepositions*, pages 57–64, Trento, Italy.

Michael Schiehlen. 2003. A Cascaded Finite-State Parser for German. In *Proceedings of the 10th Conference of the European Chapter of the Association for Computational Linguistics*, pages 163–166, Budapest, Hungary.

Helmut Schmid. 1994. Probabilistic Part-of-Speech Tagging using Decision Trees. In *Proceedings of the 1st International Conference on New Methods in Language Processing*.

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.

Sylvia Springorum. 2009. Zur Semantik der Partikelverben mit *an*. Eine Studie zur Konstruktion ihrer Bedeutung im Rahmen der Diskursrepräsentationstheorie. Studienarbeit. Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.

Sylvia Springorum. to appear. Drt-based analysis of the german verb particle an. *Leuvense Bijdragen 97*.

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

Aline Villavicencio. 2003. Verb-Particle Constructions and Lexical Resources. In *Proceedings of the ACL-2003 Workshop on Multiword Expressions: Analysis, Acquisition and Treatment*, pages 57–64, Sapporo, Japan.

Ian H. Witten, Eibe Frank, and Mark A. Hall. 2011. *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, Burlington, MA, 3 edition.