

Can Human Verb Associations help identify Salient Features for Semantic Verb Classification?

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

Abstract

This paper investigates whether human associations to verbs as collected in a web experiment can help us to identify salient verb features for semantic verb classes. Assuming that the associations model aspects of verb meaning, we apply a clustering to the verbs, as based on the associations, and validate the resulting verb classes against standard approaches to semantic verb classes, i.e. GermaNet and FrameNet. Then, various clusterings of the same verbs are performed on the basis of standard corpus-based types, and evaluated against the association-based clustering as well as GermaNet and FrameNet classes. We hypothesise that the corpus-based clusterings are better if the instantiations of the feature types show more overlap with the verb associations, and that the associations therefore help to identify salient feature types.

1 Introduction

There are a variety of manual semantic verb classifications; major frameworks are the Levin classes (Levin, 1993), *WordNet* (Fellbaum, 1998), and *FrameNet* (Fontenelle, 2003). The different frameworks depend on different instantiations of semantic similarity, e.g. Levin relies on verb similarity referring to syntax-semantic alternation behaviour, *WordNet* uses synonymy, and *FrameNet* relies on situation-based agreement as defined in Fillmore's frame semantics (Fillmore, 1982). As an alternative to the resource-intensive manual classifications,

automatic methods such as classification and clustering are applied to induce verb classes from corpus data, e.g. (Merlo and Stevenson, 2001; Joanis and Stevenson, 2003; Korhonen et al., 2003; Stevenson and Joanis, 2003; Schulte im Walde, 2003; Ferrer, 2004). Depending on the types of verb classes to be induced, the automatic approaches vary their choice of verbs and classification/clustering algorithm. However, another central parameter for the automatic induction of semantic verb classes is the selection of verb features.

Since the target classification determines the similarity and dissimilarity of the verbs, the verb feature selection should model the similarity of interest. For example, Merlo and Stevenson (2001) classify 60 English verbs which alternate between an intransitive and a transitive usage, and assign them to three verb classes, according to the semantic role assignment in the frames; their verb features are chosen such that they model the syntactic frame alternation proportions and also heuristics for semantic role assignment. In larger-scale classifications such as (Korhonen et al., 2003; Stevenson and Joanis, 2003; Schulte im Walde, 2003), which model verb classes with similarity at the syntax-semantics interface, it is not clear which features are the most salient. The verb features need to relate to a behavioural component (modelling the syntax-semantics interplay), but the set of features which potentially influence the behaviour is large, ranging from structural syntactic descriptions and argument role fillers to adverbial adjuncts. In addition, it is not clear how fine-grained the features should be; for example, how much information is covered by low-level window co-occurrence vs. higher-order syntactic frame fillers?

In this paper, we investigate whether human associations to verbs can help us to identify salient verb features for semantic verb classes. We collected associations to German verbs in a web experiment, and hope that these associations represent a useful basis for a theory-independent semantic classification of the German verbs, assuming that the associations model a non-restricted set of salient verb meaning aspects. In a preparatory step, we perform an unsupervised clustering on the experiment verbs, as based on the verb associations. We validate the resulting verb classes (henceforth: *assoc-classes*) by demonstrating that they show considerable overlap with standard approaches to semantic verb classes, i.e. GermaNet and FrameNet. In the main body of this work, we compare the associations underlying the *assoc-classes* with standard corpus-based feature types: We check on how many of the associations we find among the corpus-based features, such as adverbs, direct object nouns, etc.; we hypothesise that the more associations are found as instantiations in a feature set, the better is a clustering as based on that feature type. We assess our hypothesis by applying various corpus-based feature types to the experiment verbs, and comparing the resulting classes (henceforth: *corpus-classes*) against the *assoc-classes*. On the basis of the comparison we intend to answer the question whether *the human associations help identify salient features to induce semantic verb classes*, i.e. do the corpus-based feature types which are identified on the basis of the associations outperform previous clustering results? By applying the feature choices to GermaNet and FrameNet, we address the question whether *the same types of features are salient for different types of semantic verb classes?*

In what follows, the paper presents the association data in Section 2 and the association-based classes in Section 3. In Section 4, we compare the associations with corpus-based feature types, and in Section 5 we apply the insights to induce semantic verb classes.

2 Verb Association Data

We obtained human associations to German verbs from native speakers in a web experiment (Schulte im Walde and Melinger, 2005). 330 verbs were selected for the experiment (henceforth: *experiment verbs*), from different semantic categories, and dif-

ferent corpus frequency bands. Participants were given 55 verbs each, and had 30 seconds per verb to type as many associations as they could. 299 native German speakers participated in the experiment, between 44 and 54 for each verb. In total, we collected 81,373 associations from 16,445 trials; each trial elicited an average of 5.16 responses with a range of 0-16.

All data sets were pre-processed in the following way: For each target verb, we quantified over all responses in the experiment. Table 1 lists the 10 most frequent response types for the verb *klagen* ‘complain, moan, sue’. The responses were not distinguished according to polysemic senses of the verbs.

<i>klagen</i> ‘complain, moan, sue’		
<i>Gericht</i>	‘court’	19
<i>jammern</i>	‘moan’	18
<i>weinen</i>	‘cry’	13
<i>Anwalt</i>	‘lawyer’	11
<i>Richter</i>	‘judge’	9
<i>Klage</i>	‘complaint’	7
<i>Leid</i>	‘suffering’	6
<i>Trauer</i>	‘mourning’	6
<i>Klagemauer</i>	‘Wailing Wall’	5
<i>laut</i>	‘noisy’	5

Table 1: Association frequencies for target verb.

In the clustering experiments to follow, the verb associations are considered as verb features. The underlying assumption is that verbs which are semantically similar tend to have similar associations, and are therefore assigned to common classes. Table 2 illustrates the overlap of associations for the polysemous *klagen* with a near-synonym of one of its senses, *jammern* ‘moan’. The table lists those associations which were given at least twice for each verb; the total overlap was 35 association types.

<i>klagen/jammern</i> ‘moan’		
<i>Frauen</i>	‘women’	2/3
<i>Leid</i>	‘suffering’	6/3
<i>Schmerz</i>	‘pain’	3/7
<i>Trauer</i>	‘mourning’	6/2
<i>bedauern</i>	‘regret’	2/2
<i>beklagen</i>	‘bemoan’	4/3
<i>heulen</i>	‘cry’	2/3
<i>nervig</i>	‘annoying’	2/2
<i>nölen</i>	‘moan’	2/3
<i>traurig</i>	‘sad’	2/5
<i>weinen</i>	‘cry’	13/9

Table 2: Association overlap for target verbs.

3 Association-based Verb Classes

We performed a standard clustering on the 330 experiment target verbs: The verbs and their features were taken as input to agglomerative (bottom-up) hierarchical clustering. As similarity measure in the clustering procedure (i.e. to determine the distance/similarity for two verbs), we used the *skew divergence*, a smoothed variant of the *Kullback-Leibler divergence* (Lee, 2001). The goal of these experiments was not to explore the optimal feature combination; thus, we rely on previous experiments and parameter settings, cf. Schulte im Walde (2003).

Our claim is that the hierarchical verb classes and their underlying features (i.e. the verb associations) represent a useful basis for a theory-independent semantic classification of the German verbs. To support this claim, we validated the assoc-classes against standard approaches to semantic verb classes, i.e. *GermaNet* as the German WordNet (Kunze, 2000), and the German counterpart of FrameNet in the *Salsa* project (Erk et al., 2003). Details of the validation can be found in (Schulte im Walde, 2006); the main issues are as follows.

We did not directly compare the assoc-classes against the GermaNet/FrameNet classes, since not all of our 330 experiment verbs were covered by the two resources. Instead, we replicated the above cluster experiment for a reduced number of verbs: We extracted those classes from the resources which contain association verbs; light verbs, non-association verbs, other classes as well as singletons were disregarded. This left us with 33 classes from GermaNet, and 38 classes from FrameNet. These remaining classifications are polysemous: The 33 GermaNet classes contain 71 verb senses which distribute over 56 verbs, and the 38 FrameNet classes contain 145 verb senses which distribute over 91 verbs. Based on the 56/91 verbs in the two gold standard resources, we performed two cluster analyses, one for the GermaNet verbs, and one for the FrameNet verbs. As for the complete set of experiment verbs, we performed a hierarchical clustering on the respective subsets of the experiment verbs, with their associations as verb features. The actual validation procedure then used the reduced classifications: The resulting analyses were evaluated against the resource classes on each level in

the hierarchies, i.e. from 56/91 classes to 1 class. As evaluation measure, we used a pair-wise measure which calculates precision, recall and a harmonic f-score as follows: Each verb pair in the cluster analysis was compared to the verb pairs in the gold standard classes, and evaluated as true or false positive (Hatzivassiloglou and McKeown, 1993).

The association-based clusters show overlap with the lexical resource classes of an f-score of 62.69% (for 32 verb classes) when comparing to GermaNet, and 34.68% (for 10 verb classes) when comparing to FrameNet. The corresponding upper bounds are 82.35% for GermaNet and 60.31% for FrameNet.¹ The comparison therefore demonstrates considerable overlap between association-based classes and existing semantic classes. The different results for the two resources are due to their semantic background (i.e. capturing synonymy vs. situation-based agreement), the numbers of verbs, and the degrees of ambiguity (an average of 1.6 senses per verb in FrameNet, as compared to 1.3 senses in GermaNet).

The purpose of the validation against semantic resources was to demonstrate that a clustering as based on the verb associations and a standard clustering setting compares well with existing semantic classes. We take the positive validation results as justification to use the assoc-classes as source for cluster information: The clustering defines the verbs in a common association-based class, and the features which are relevant for the respective class. For example, the 100-class analysis contains a class with the verbs *bedauern* ‘regret’, *heulen* ‘cry’, *jammern* ‘moan’, *klagen* ‘complain, moan, sue’, *verzweifeln* ‘become desperate’, and *weinen* ‘cry’, with the most distinctive features *Trauer* ‘mourning’, *weinen* ‘cry’, *traurig* ‘sad’, *Tränen* ‘tears’, *jammern* ‘moan’, *Angst* ‘fear’, *Mitleid* ‘pity’, *Schmerz* ‘pain’.

4 Exploring Semantic Class Features

Our claim is that the features underlying the association-based classes help us guide the feature selection process in future clustering experiments, because we know which semantic classes are based

¹The upper bounds are below 100%, because the hierarchical clustering assigns a verb to only one cluster, but the lexical resources contain polysemy. We created a hard version of the lexical resource classes where we randomly chose one sense of each polysemous verb, to calculate the upper bounds.

on which associations/features. We rely on the assoc-classes in the 100-class analysis of the hierarchical clustering² and features which exist for at least two verbs in a common class (and therefore hint to a minimum of verb similarity), and compare the associations underlying the assoc-classes with standard corpus-based feature types: We check on how many of the associations we find among the corpus-based features, such as adverbs, direct object nouns, etc. There are various possibilities to determine corpus-based features that potentially cover the associations; we decided in favour of feature types which have been suggested in related work:

a) Grammar-based relations: Previous work on distributional similarity has focused either on a specific word-word relation (such as Pereira et al. (1993) and Rooth et al. (1999) referring to a direct object noun for describing verbs), or used any syntactic relationship detected by a chunker or a parser (such as Lin (1998) and McCarthy et al. (2003)). We used a statistical grammar (Schulte im Walde, 2003) to filter all verb-noun pairs where the nouns represent nominal heads in NPs or PPs in syntactic relation to the verb (subject, object, adverbial function, etc.), and to filter all verb-adverb pairs where the adverbs modify the verbs.

b) Co-occurrence window: In previous work (Schulte im Walde and Melinger, 2005), we showed that only 28% of all noun associates were identified by the above statistical grammar as subcategorised nouns, but 69% were captured by a 20-word co-occurrence window in a 200-million word newspaper corpus. This finding suggests to use a co-occurrence window as alternative source for verb features, as compared to specific syntactic relations. We therefore determined the co-occurring words for all experiment verbs in a 20-word window (i.e. 20 words preceding and following the verb), irrespective of the part-of-speech of the co-occurring words.

Relying on the verb information extracted for a) and b), we checked for each verb-association pair whether it occurred among the grammar or window pairs. Table 3 illustrates which proportions of the associations we found in the two resource types. For the grammar-based relations, we checked argu-

ment NPs and PPs (as separate sets and together), and in addition we checked verb-noun pairs in the most common specific NP functions: *n* refers to the (nominative) intransitive subject, *na* to the transitive subject, and *n_a* to the transitive (accusative) object. For the windows, *all* checks on co-occurrence of verbs and associations in the whole 200-million word corpus. *cut* also checks the whole corpus, but disregards the most and least frequent co-occurring words: verb-word pairs were only considered if the co-occurrence frequency of the word over all verbs was above 100 (disregarding low frequency pairs) and below 200,000 (disregarding high frequency pairs). Using the cut-offs, we can distinguish the relevance of high- and low-frequency features. Finally, *ADJ*, *ADV*, *N*, *V* perform co-occurrence checks for the whole corpus, but breaks down the *all* results with respect to the association part-of-speech.

As one would have expected, most of the associations (66%) were found in the 20-word co-occurrence window, because the window is neither restricted to a certain part-of-speech, nor to a certain grammar relation; in addition, the window is potentially larger than a sentence. Applying the frequency cut-offs reduces the overlap of association types and co-occurring words to 58%. Specifying the window results for the part-of-speech types illustrates that the nouns play the most important role in describing verb meaning (39% of the verb association types in the assoc-classes were found among the nouns in the corpus windows, 16% among the verbs, 9% among the adjectives, and 2% among the adverbs).³

The proportions of the nouns with a specific grammar relationship to the verbs show that we find more associations among direct objects than intransitive/transitive subjects. This insight confirms the assumption in previous work where only direct object nouns were used as salient features in distributional verb similarity, such as Pereira et al. (1993). However, the proportions are all below 10%. Considering all NPs and/or PPs, we find that the proportions increase for the NPs, and that the NPs play a more important role than the PPs. This insight confirms work on distributional similarity where not only direct object nouns, but all functional nouns

²The exact number of classes or the verb-per-class ratio are not relevant for investigating the use of associations.

³Caveat: These numbers correlate with the part-of-speech types of all associate responses: 62% of the responses were nouns, 25% verbs, 11% adjectives, and 2% adverbs.

Features	grammar relations						
	<u>n</u>	<u>na</u>	<u>na</u>	NP	PP	NP&PP	ADV
Cov. (%)	3.82	4.32	6.93	12.23	5.36	14.08	3.63

Features	co-occurrence: window-20					
	all	cut	ADJ	ADV	N	V
Cov. (%)	66.15	57.79	9.13	1.72	39.27	15.51

Table 3: Coverage of verb association features by grammar/window resources.

were considered as verb features, such as Lin (1998) and McCarthy et al. (2003). Of the adverb associations, we find only a small proportion among the parsed adverbs. All in all, the proportions of association types among the nouns/adverbs with a syntactic relationship to the verbs are rather low. Comparing the NP/PP proportions with the window noun proportions shows that salient verb features are not restricted to certain syntactic relationships, but also appear in a less restricted context window.

5 Inducing Verb Classes with Corpus-based Features

In the final step, we applied the corpus-based feature types to clusterings. The goal of this step was to determine whether the feature exploration helped to identify salient verb features, and whether we can outperform previous clustering results. The clustering experiments were as follows: The 330 experiment verbs were instantiated by the feature types we explored in Section 4. As for the assoc-classes, we then performed an agglomerative hierarchical clustering. We cut the hierarchy at a level of 100 clusters, and evaluated the clustering against the 100-class analysis of the original assoc-classes. We expect that feature types with a stronger overlap with the association types result in a better clustering result. The assumption is that the associations are salient feature for verb clustering, and the better we model the associations with grammar-based or window-based features, the better the clustering.

For checking the clusterings with respect to the semantic class type, we also applied the corpus-based features to GermaNet and FrameNet classes.

- **GermaNet:** We randomly extracted 100 verb classes from all GermaNet synsets, and created a hard classification for these classes, by randomly deleting additional senses of a verb so

as to leave only one sense for each verb. This selection made the GermaNet classes comparable to the assoc-classes in size and polysemy. The 100 classes contain 233 verbs. Again, we performed an agglomerative hierarchical clustering on the verbs (as modelled by the different feature types). We cut the hierarchy at a level of 100 clusters, which corresponds to the number of GermaNet classes, and evaluated against the GermaNet classes.

- **FrameNet:** In a pre-release version from May 2005, there were 484 verbs in 214 German FrameNet classes. We disregarded the high-frequency verbs *gehen*, *geben*, *sehen*, *kommen*, *bringen* which were assigned to classes mostly on the basis of multi-word expressions they are part of. In addition, we disregarded two large classes which contained mostly support verbs, and we disregarded singletons. Finally, we created a hard classification of the classes, by randomly deleting additional senses of a verb so as to leave only one sense for each verb. The classification then contained 77 classes with 406 verbs. Again, we performed an agglomerative hierarchical clustering on the verbs (as modelled by the different feature types). We cut the hierarchy at a level of 77 clusters, which corresponds to the number of FrameNet classes, and evaluated against the FrameNet classes.

For the evaluation of the clustering results, we calculated the *accuracy* of the clusters, a cluster similarity measure that has been applied before, cf. (Stevenson and Joanis, 2003; Korhonen et al., 2003).⁴ Accuracy is determined in two steps:

⁴Note that we can use *accuracy* for the evaluation because we have a fixed cut in the hierarchy as based on the gold standard, as opposed to the evaluation in Section 3 where we explored the optimal cut level.

	frames		grammar relations						
	f-pp	f-pp-pref	<u>n</u>	<u>na</u>	<u>na</u>	NP	PP	NP&PP	ADV
Assoc	37.50	37.80	35.90	37.18	39.25	39.14	37.97	41.28	38.53
GN	46.98	49.14	58.01	53.37	51.90	53.10	54.21	51.77	51.82
FN	33.50	32.76	29.46	30.13	32.74	34.16	28.72	33.91	35.24

	co-occurrence: window-20					
	all	cut	ADJ	ADV	N	V
Assoc	39.33	39.45	37.31	36.89	39.33	38.84
GN	51.53	52.42	50.88	47.79	52.86	49.12
FN	missing	32.84	31.08	31.00	34.24	31.75

Table 4: Accuracy for induced verb classes.

1. For each class in the cluster analysis, the gold standard class with the largest intersection of verbs is determined. The number of verbs in the intersection ranges from one verb only (in case all clustered verbs are in different classes in the gold standard) to the total number of verbs in a cluster (in case all clustered verbs are in the same gold standard class).
2. Accuracy is calculated as the proportion of the verbs in the clusters covered by the same gold standard classes, divided by the total number of verbs in the clusters. The upper bound of the accuracy measure is 1.

Table 4 shows the accuracy results for the three types of classifications (assoc-classes, GermaNet, FrameNet), and the grammar-based and window-based features. We added frame-based features, as to compare with earlier work: The frame-based features provide a feature description over 183 syntactic frame types including PP type specification (f-pp), and the same information plus coarse selectional preferences for selected frame slots, as obtained from GermaNet top-level synsets (f-pp-pref), cf. (Schulte im Walde, 2003). The following questions are addressed with respect to the result table.

1. Do the results of the clusterings with respect to the underlying feature types correspond to the overlap of associations and feature types, cf. Table 3?
2. Do the corpus-based feature types which were identified on the basis of the associations outperform previous clustering results?
3. Do the results generalise over the semantic class type?

First of all, there is no correlation between the overlap of associations and feature types on the one hand and the clustering results as based on the feature types on the other hand (Pearson’s correlation, $p > .1$), neither for the assoc-classes or the GermaNet or FrameNet classes. The human associations therefore did not contribute to identify salient feature types, as we had hoped. In some specific cases, we find corresponding patterns; for example, the clustering results for the intransitive and transitive subject and the transitive object correspond to the overlap values for the assoc-classes and FrameNet: $\underline{n} < \underline{na} < \underline{na}$. Interestingly, the GermaNet clusterings behave in the opposite direction.

Comparing the grammar-based relations with each other shows that for the assoc-classes using all NPs is better than restricting the NPs to (subject) functions, and using both NPs and PPs is best; similarly for the FrameNet classes where using all NPs is the second best results (but adverbs). Differently, for the GermaNet classes the specific function of intransitive subjects outperforms the more general feature types, and the PPs are still better than the NPs. We conclude that not only there is no correlation between the association overlap and feature types, but in addition the most successful feature types vary strongly with respect to the gold standard. None of the differences within the feature groups ($\underline{n}/\underline{na}/\underline{na}$ and NP/PP/NP&PP) are significant ($\chi^2, df = 1, \alpha = 0.05$). The adverbial features are surprisingly successful in all three clusterings, in some cases outperforming the noun-based features.

Comparing the grammar-based clustering results with previous results, the grammar-based features outperform the frame-based features in all clusterings for the GermaNet verbs. For the FrameNet

verbs and the experiment verbs, they outperform the frame-based features only in specific cases. The adverbial features outperform the frame-based features in any clustering. However, none of the differences between the frame-based clusterings and the grammar-based clusterings are significant ($\chi^2, df = 1, \alpha = 0.05$).

For all gold standards, the best window-based clustering results are below the best grammar-based results. Especially the *all* results demonstrate once more the missing correlation between association/feature overlap and clustering results. However, it is interesting that the clusterings based on window co-occurrence are not significantly worse (and in some cases even better) than the clusterings based on selected grammar-based functions. This means that a careful choice and extraction of specific relationships for verb features does not have a significant impact on semantic classes.

Comparing the window-based features against each other shows that even though we discovered a much larger proportion of association types in an unrestricted window *all* than elsewhere, the results in the clusterings do not differ accordingly. Applying the frequency cut-offs has almost no impact on the clustering results, which means that it does no harm to leave away the rather unpredictable features. Somehow expected but nevertheless impressive is the fact that only considering nouns as co-occurring words is as successful as considering all words independent of the part-of-speech.

Finally, the overall accuracy values are much better for the GermaNet clusterings than for the experiment-based and the FrameNet clusterings. The differences are all significant ($\chi^2, df = 1, \alpha = 0.05$). The reason for these large differences could be either (a) that the clustering task was easier for the GermaNet verbs, or (b) that the differences are caused by the underlying semantics. We argue against case (a) since we deliberately chose the same number of classes (100) as for the association-based gold standard; however, the verbs-per-class ratio for GermaNet vs. the assoc-classes and the FrameNet classes is different (2.33 vs. 3.30/5.27) and we cannot be sure about this influence. In addition, the average verb frequencies in the GermaNet classes (calculated in a 35 million word newspaper corpus) are clearly below those in the other two classifica-

tions (1,040 as compared to 2,465 and 1,876), and there are more low-frequency verbs (98 out of 233 verbs (42%) have a corpus frequency below 50, as compared to 41 out of 330 (12%) and 54 out of 406 (13%)). In the case of (b), the difference in the semantic class types is modelling synonyms with GermaNet as opposed to situation-based agreement in FrameNet. The association-based class semantics is similar to FrameNet, because the associations are unrestricted in their semantic relation to the experiment verb (Schulte im Walde and Melinger, 2005).

6 Summary

The questions we posed in the beginning of this paper were (i) whether human associations help identify salient features to induce semantic verb classes, and (ii) whether the same types of features are salient for different types of semantic verb classes. An association-based clustering with 100 classes served as source for identifying a set of potentially salient verb features, and a comparison with standard corpus-based features determined proportions of feature overlap. Applying the standard feature choices to verbs underlying three gold standard verb classifications showed that (a) in our experiments there is no correlation between the overlap of associations and feature types and the respective clustering results. The associations therefore did not help in the specific choice of corpus-based features, as we had hoped. However, the assumption that window-based features do contribute to semantic verb classes – this assumption came out of an analysis of the associations – was confirmed: simple window-based features were not significantly worse (and in some cases even better) than selected grammar-based functions. This finding is interesting because window-based features have often been considered too simple for semantic similarity, as opposed to syntax-based features. (b) Several of the grammar-based nominal and adverbial features and also the window-based features outperformed feature sets in previous work, where frame-based features (plus prepositional phrases and coarse selectional preference information) were used. Surprisingly well did adverbs: they only represent a small number of verb features, but obviously this small selection can outperform frame-based features and even some nomi-

nal features. (c) The clustering results were significantly better for the GermaNet clusterings than for the experiment-based and the FrameNet clusterings, so the chosen feature sets might be more appropriate for the synonymy-based than the situation-based classifications.

Acknowledgements Thanks to Christoph Clodo and Marty Mayberry for their system administrative help when running the cluster analyses.

References

- Katrin Erk, Andrea Kowalski, Sebastian Padó, and Manfred Pinkal. 2003. Towards a Resource for Lexical Semantics: A Large German Corpus with Extensive Semantic Annotation. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, Sapporo, Japan.
- Christiane Fellbaum, editor. 1998. *WordNet – An Electronic Lexical Database*. Language, Speech, and Communication. MIT Press, Cambridge, MA.
- Eva Esteve Ferrer. 2004. Towards a Semantic Classification of Spanish Verbs based on Subcategorisation Information. In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics*, Barcelona, Spain.
- Charles J. Fillmore. 1982. Frame Semantics. *Linguistics in the Morning Calm*, pages 111–137.
- Thierry Fontenelle, editor. 2003. *FrameNet and Frame Semantics*, volume 16(3) of *International Journal of Lexicography*. Oxford University Press. Special issue devoted to FrameNet.
- Vasileios Hatzivassiloglou and Kathleen R. McKeown. 1993. Towards the Automatic Identification of Adjectival Scales: Clustering Adjectives According to Meaning. In *Proceedings of the 31st Annual Meeting of the Association for Computational Linguistics*, pages 172–182, Columbus, Ohio.
- Eric Joanis and Suzanne Stevenson. 2003. A General Feature Space for Automatic Verb Classification. In *Proceedings of the 10th Conference of the European Chapter of the Association for Computational Linguistics*, Budapest, Hungary.
- Anna Korhonen, Yuval Krymolowski, and Zvika Marx. 2003. Clustering Polysemic Subcategorization Frame Distributions Semantically. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, pages 64–71, Sapporo, Japan.
- 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.
- Lillian Lee. 2001. On the Effectiveness of the Skew Divergence for Statistical Language Analysis. *Artificial Intelligence and Statistics*, pages 65–72.
- Dekang Lin. 1998. Automatic Retrieval and Clustering of Similar Words. In *Proceedings of the 17th International Conference on Computational Linguistics*, Montreal, Canada.
- 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.
- Paola Merlo and Suzanne Stevenson. 2001. Automatic Verb Classification Based on Statistical Distributions of Argument Structure. *Computational Linguistics*, 27(3):373–408.
- Fernando Pereira, Naftali Tishby, and Lillian Lee. 1993. Distributional Clustering of English Words. In *Proceedings of the 31st Annual Meeting of the Association for Computational Linguistics*, pages 183–190.
- Mats Rooth, Stefan Riezler, Detlef Prescher, Glenn Carroll, and Franz Beil. 1999. Inducing a Semantically Annotated Lexicon via EM-Based Clustering. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*, Maryland, MD.
- Sabine Schulte im Walde and Alissa Melinger. 2005. Identifying Semantic Relations and Functional Properties of Human Verb Associations. In *Proceedings of the joint Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 612–619, Vancouver, Canada.
- Sabine Schulte im Walde. 2003. *Experiments on the Automatic Induction of German Semantic Verb Classes*. Ph.D. thesis, Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart. Published as AIMS Report 9(2).
- Sabine Schulte im Walde. 2006. Human Verb Associations as the Basis for Gold Standard Verb Classes: Validation against GermaNet and FrameNet. In *Proceedings of the 5th Conference on Language Resources and Evaluation*, Genoa, Italy.
- Suzanne Stevenson and Eric Joanis. 2003. Semi-supervised Verb Class Discovery Using Noisy Features. In *Proceedings of the 7th Conference on Natural Language Learning*, pages 71–78, Edmonton, Canada.