Inducing German Semantic Verb Classes from Purely Syntactic Subcategorisation Information

Sabine Schulte im Walde
Institut für Maschinelle Sprachverarbeitung
Universität Stuttgart
Azenbergstraße 12, 70174 Stuttgart, Germany
schulte@ims.uni-stuttgart.de

Chris Brew
Department of Linguistics
The Ohio State University
Columbus, USA, OH 43210-1298
cbrew@ling.ohio-state.edu

Abstract

The paper describes the application of k-Means, a standard clustering technique, to the task of inducing semantic classes for German verbs. Using probability distributions over verb subcategorisation frames, we obtained an intuitively plausible clustering of 57 verbs into 14 classes. The automatic clustering was evaluated against independently motivated, hand-constructed semantic verb classes. A series of post-hoc cluster analyses explored the influence of specific frames and frame groups on the coherence of the verb classes, and supported the tight connection between the syntactic behaviour of the verbs and their lexical meaning components.

1 Introduction

A long-standing linguistic hypothesis asserts a tight connection between the meaning components of a verb and its syntactic behaviour: To a certain extent, the lexical meaning of a verb determines its behaviour, particularly with respect to the choice of its arguments. The theoretical foundation has been established in extensive work on semantic verb classes such as (Levin, 1993) for English and (Vázquez et al., 2000) for Spanish: each verb class contains verbs which are similar in their meaning and in their syntactic properties.

From a practical point of view, a verb classification supports Natural Language Processing tasks, since it provides a principled basis for filling gaps in available lexical knowledge. For example, the English verb classification has been used for applications such as machine translation (Dorr, 1997), word sense disambiguation (Dorr and Jones, 1996), and document classification (Klavans and Kan, 1998). Various attempts have been made to infer conveniently observable morpho-syntactic and semantic properties for English verb classes (Dorr and Jones, 1996; Lapata, 1999; Stevenson and Merlo, 1999; Schulte im Walde, 2000; McCarthy, 2001).

To our knowledge this is the first work to obtain German verb classes automatically. We used a robust statistical parser (Schmid, 2000) to acquire purely syntactic subcategorisation information for verbs. The information was provided in form of probability distributions over verb frames for each verb. There were two conditions: the first with relatively coarse syntactic verb subcategorisation frames, the second a more delicate classification subdividing the verb frames of the first condition using prepositional phrase information (case plus preposition). In both conditions verbs were clustered using k-Means, an iterative, unsupervised, hard clustering method with well-known properties, cf. (Kaufman and Rousseeuw, 1990). The goal of a series of cluster analyses was (i) to find good values for the parameters of the clustering process, and (ii) to explore the role of the syntactic frame descriptions in verb classification, to demonstrate the implicit induction of lexical meaning components from syntactic properties, and to suggest ways in which the syntactic information might further be refined. Our long term goal is to support the development of high-quality and large-scale lexical resources.

2 Syntactic Descriptors for Verb Frames

The syntactic subcategorisation frames for German verbs were obtained by unsupervised learning in a statistical grammar framework (Schulte im Walde et al., 2001): a German context-free grammar containing frame-predicting grammar rules and information about lexical heads was trained on 25 million words of a large German newspaper corpus. The lexicalised version of the probabilistic grammar served as source for syntactic descriptors for verb frames (Schulte im Walde, 2002b).

The verb frame types contain at most three arguments. Possible arguments in the frames are nominative (n), dative (d) and accusative (a) noun phrases, reflexive pronouns (r), prepositional phrases (p), expletive *es* (x), non-finite clauses (i), finite clauses (s-2 for verb second clauses, s-dass for *dass*-clauses, s-ob for *ob*-clauses, s-w for indirect *wh*-questions), and copula constructions (k). For example, subcategorising a direct (accusative case) object and a non-finite clause would be represented by `nai`. We defined a total of 38 subcategorisation frame types, according to the verb subcategorisation potential in the German grammar (Helbig and Buscha, 1998), with few further restrictions on argument combination.

We extracted verb-frame distributions from the trained lexicalised grammar. Table 1 shows an example distribution for the verb *glauben* 'to think/believe' (for probability values $>1\%$).

Frame	Prob
ns-dass	0.27945
ns-2	0.27358
np	0.09951
n	0.08811
na	0.08046
ni	0.05015
nd	0.03392
nad	0.02325
nds-2	0.01011

Table 1: Probability distribution for *glauben*

We also created a more delicate version of subcategorisation frames that discriminates between different kinds of pp-arguments. This was done by distributing the frequency mass of prepositional phrase frame types (np, nap, ndp, npr, xp) over the prepositional phrases, according to their frequencies in the corpus. Prepositional phrases are referred to by case and preposition, such as 'Dat.mit', 'Akk.für'. The resulting lexical subcategorisation for *reden* and the frame type `np` whose total joint probability is 0.35820, is displayed in Table 2 (for probability values $>1\%$).

Refined Frame		Prob
np:Akk.über	acc / 'about'	0.11981
np:Dat.von	dat / 'about'	0.11568
np:Dat.mit	dat / 'with'	0.06983
np:Dat.in	dat / 'in'	0.02031

Table 2: Refined `np` distribution for *reden*

The subcategorisation frame descriptions were formally evaluated by comparing the automatically generated verb frames against manual definitions in the German dictionary *Duden – Das Stilwörterbuch* (Dudenredaktion, 2001). The F-score was 65.30% with and 72.05% without prepositional phrase information: the automatically generated data is both easy to produce in large quantities and reliable enough to serve as proxy for human judgement (Schulte im Walde, 2002a).

3 German Semantic Verb Classes

Semantic verb classes have been defined for several languages, with dominant examples concerning English (Levin, 1993) and Spanish (Vázquez et al., 2000). The basic linguistic hypothesis underlying the construction of the semantic classes is that verbs in the same class share both meaning components and syntactic behaviour, since the meaning of a verb is supposed to influence its behaviour in the sentence, especially with regard to the choice of its arguments.

We hand-constructed a concise classification with 14 semantic verb classes for 57 German verbs before we initiated any clustering experiments. We have on hand a larger set of verbs and a more elaborate classification, but choose to work on the smaller set for the moment, since an important component of our research program is an informative post-hoc analysis which becomes infeasible with larger datasets. The semantic aspects and majority of verbs are closely related to Levin's English classes. They are consistent with the German verb classification in (Schumacher, 1986) as far as the relevant verbs appear in his less extensive semantic ‘fields’.

1. *Aspect*: anfangen, aufhören, beenden, beginnen, enden
2. *Propositional Attitude*: ahnen, denken, glauben, vermuten, wissen
3. *Transfer of Possession (Obtaining)*: bekommen, erhalten, erlangen, kriegen
4. *Transfer of Possession (Supply)*: bringen, liefern, schicken, vermitteln, zustellen
5. *Manner of Motion*: fahren, fliegen, rudern, segeln
6. *Emotion*: ärgern, freuen
7. *Announcement*: ankündigen, bekanntgeben, eröffnen, verkünden
8. *Description*: beschreiben, charakterisieren, darstellen, interpretieren
9. *Insistence*: beharren, bestehen, insistieren, pochen
10. *Position*: liegen, sitzen, stehen
11. *Support*: dienen, folgen, helfen, unterstützen
12. *Opening*: öffnen, schließen
13. *Consumption*: essen, konsumieren, lesen, saufen, trinken
14. *Weather*: blitzen, donnern, dämmern, nieseln, regnen, schneien

The class size is between 2 and 6, no verb appears in more than one class. For some verbs this is something of an oversimplification; for example, the verb *bestehen* is assigned to verbs of insistence, but it also has a salient sense more related to existence. Similarly, *schließen* is recorded under open/close, in spite of the fact it also has a meaning related to inference and the formation of conclusions. The classes include both high and low frequency verbs, because we wanted to make sure that our clustering technology was exercised in both data-rich and data-poor situations. The corpus frequencies range from 8 to 31,710.

Our target classification is based on semantic intuitions, not on our knowledge of the syntactic behaviour. As an extreme example, the semantic class *Support* contains the verb *unterstützen*, which syntactically requires a direct object, together with the three verbs *dienen, folgen, helfen* which dominantly subcategorise an indirect object. In what follows we will show that the semantic classification is largely recoverable from the patterns of verb-frame occurrence.

4 Clustering Methodology

Clustering is a standard procedure in multivariate data analysis. It is designed to uncover an inherent natural structure of the data objects, and the equivalence classes induced by the clusters provide a means for generalising over these objects. In our case, clustering is realised on verbs: the data objects are represented by verbs, and the data features for describing the objects are realised by a probability distribution over syntactic verb frame descriptions.

Clustering is applicable to a variety of areas in Natural Language Processing, e.g. by utilising class type descriptions such as in machine translation (Dorr, 1997), word sense disambiguation (Dorr and Jones, 1996), and document classification (Klavans and Kan, 1998), or by applying clusters for smoothing such as in machine translation (Prescher et al., 2000), or probabilistic grammars (Riezler et al., 2000).

We performed clustering by the k-Means algorithm as proposed by (Forgy, 1965), which is an unsupervised hard clustering method assigning n data objects to exactly k clusters. Initial verb clusters are iteratively re-organised by assigning each verb to its closest cluster (centroid) and re-calculating cluster centroids until no further changes take place.

One parameter of the clustering process is the distance measure used. Standard choices include the cosine, Euclidean distance, Manhattan metric, and variants of the Kullback-Leibler (KL) divergence. We concentrated on two variants of KL in Equation (1): information radius, cf. Equation (2), and skew divergence, recently shown as an effective measure for distributional similarity (Lee, 2001), cf. Equation (3).

$$d(v_1, v_2) = D(p \,||\, q) = \sum_i p_i \, log \, \frac{p_i}{q_i} \quad (1)$$

$$d(v_1, v_2) = D(p \,||\, \frac{p+q}{2}) + D(q \,||\, \frac{p+q}{2}) \quad (2)$$

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

Measures (2) and (3) can tolerate zero values in the probability distribution, because they work with a weighted average of the two distributions compared. For the skew-divergence, we set the weight w to 0.9, as was done by Lee.

Furthermore, because the k-Means algorithm is sensitive to its starting clusters, we explored the option of initialising the cluster centres based on other clustering algorithms. We performed agglomerative hierarchical clustering on the verbs which first assigns each verb to its own cluster and then iteratively determines the two closest clusters and merges them, until the specified number of clusters is left. We tried several amalgamation methods: single-linkage, complete-linkage, average verb distance, distance between cluster centroids, and Ward's method.

The clustering was performed as follows: the 57 verbs were associated with probability distributions over frame types[1] (in condition 1 there were 38 frame types, while in the more delicate condition 2 there were 171, with a concomitant increase in data sparseness), and assigned to starting clusters (randomly or by hierarchical clustering). The k-Means algorithm was then allowed to run for as many iterations as it takes to reach a fixed point, and the resulting clusters were interpreted and evaluated against the manual classes.

Related work on English verb classification or clustering utilised supervised learning by decision trees (Stevenson and Merlo, 1999), or a method related to hierarchical clustering (Schulte im Walde, 2000).

5 Clustering Evaluation

The task of evaluating the result of a cluster analysis against the known gold standard of hand-constructed verb classes requires us to assess the similarity between two sets of equivalence relations. As noted by (Strehl et al., 2000), it is useful to have an evaluation measure that does not depend on the choice of similarity measure or on the original dimensionality of the input data, since that allows meaningful comparison of results for which these parameters vary. This is similar to the perspective of (Vilain et al., 1995), who present, in the context of the MUC co-reference evaluation scheme, a model-theoretic measure of the similarity between equivalence classes.

Strehl et al. consider a clustering C that partitions n objects $(x_o \ldots x_n)$ into k clusters; the clusters C_i of C are the sets for which $C_i = \{x_j \mid C[x_j] = i\}$. We call the cluster result A and the desired gold-standard B. For measuring the quality of an individual cluster, the cluster *purity* of each cluster A_i is defined by its largest AB_{ik}, the number of members A_i that are projected into the same class B_k. The measure is biased towards small clusters, with the extreme case of singleton clusters, which is an undesired property for our (linguistic) needs.

To capture the quality of a whole clustering, Strehl et al. combine the mutual information between A and B (based on the shared verb membership AB_{ij}) with a scaling factor corresponding to the numbers of verbs in the respective clusters, $|A|$ and $|B|$.

$$MI(A,B) = \frac{1}{n} \sum_i \sum_j AB_{ij} \frac{log(\frac{AB_{ij}\ .\ n}{\sum_i AB_{ij}\ \sum_j AB_{ij}})}{log(|A|.|B|)} \tag{4}$$

This manipulation is designed to remove the bias towards small clusters:[2] using the 57 verbs from our study we generated 50 random clusters for each cluster size between 1 and 57, and evaluated the results against the gold standard, returning the best result for each replication. We found that even using the scaling factor the measure favours smaller clusters. But this bias is strongest at the extremes of the range, and does not appear to impact too heavily on our results.

Unfortunately none of Strehl et al's measures have all the properties which we intuitively require from a measure of linguistic cluster quality. For example, if we restrict attention to the case in which all verbs in an inferred cluster are drawn from the same actual class, we would like it to be the case that the evaluation measure is a monotonically increasing function of the size of the inferred cluster. We therefore introduced an additional, more suitable measure for the evaluation of individual clusters, based on the representation of equivalence classes as sets of pairs. It turns out that pairwise precision and recall have some of the counter-intuitive properties that we objected to in Strehl et al's measures, so we adjust pairwise precision with a scaling factor based on the size

[1] We also tried various transformations and variations of the probabilities, such as frequencies and binarisation, but none proved as effective as the probabilities.

[2] In the absence of the penalty, mutual information would attain its maximum (which is the entropy of B) not only when A is correct but also when A contains only singleton clusters.

of the hypothesised cluster.

$$APP(A_i) = \frac{\text{number of correct pairs in } A_i}{\text{number of verbs in } A_i + 1} \quad (5)$$

We call this measure APP, for *adjusted pairwise precision*. As with any other measure of individual cluster quality we can associate a quality value with a clustering C which assigns each of the items x_j to a cluster $C[x_j]$ by taking a weighted average over the qualities of the individual clusters.

$$APP(C) = \frac{1}{|C|} \sum_j \frac{APP(C_j)}{|C_j|} \quad (6)$$

Figures 1 and 2 summarise the two evaluation measures for overall cluster quality, showing the variation with the KL-based distance measures and with different strategies for seeding the initial cluster centres in the k-Means algorithm. Figure 1 displays quality scores referring to the coarse condition 1 subcategorisation frame types, Figure 2 refers to the clustering results obtained by verb descriptions based on the more delicate condition 2 subcategorisation frame types including PP information. Baseline values are 0.017 (APP) and 0.229 (MI), calculated as average on the evaluation of 10 random clusters. Optimum values, as calculated on the manual classification, are 0.291 (APP) and 0.493 (MI). The evaluation function is extremely non-linear, which leads to a severe loss of quality with the first few clustering mistakes, but does not penalise later mistakes to the same extent.

From the methodological point of view, the clustering evaluation gave interesting insights into k-Means' behaviour on the syntactic frame data. The more delicate verb-frame classification, i.e. the refinement of the syntactic verb frame descriptions by prepositional phrase specification, improved the clustering results. This does not go without saying: there was potential for a sparse data problem, since even frequent verbs can only be expected to inhabit a few frames. For example, the verb *anfangen* with a corpus frequency of 2,554 has zero counts for 138 of the 171 frames. Whether the improvement really matters in an application task is left to further research.

We found that randomised starting clusters usually give better results than initialisation from a hierarchical clustering. Hierarchies imposing a strong structure on the clustering (such as single-linkage: the output clusterings contain few very large and many singleton clusters) are hardly improved by k-Means. Their evaluation results are noticeably below those for random clusters. But initialisation using Ward's method, which produces tighter clusters and a narrower range of cluster sizes does outperform random cluster initialisation. Presumably the issue is that the other hierarchical clustering methods place k-Means in a local minimum from which it cannot escape, and that uniformly shaped cluster initialisation gives k-Means a better chance of avoiding local minima, even with a high degree of perturbation.

6 Linguistic Investigation

The clustering setup, proceeding and results provide a basis for a linguistic investigation concerning the German verbs, their syntactic properties and semantic classification.

The following clustering result is an intuitively plausible semantic verb classification, accompanied by the cluster quality scores APP, and class labels illustrating the majority vote of the verbs in the cluster.[3] The cluster analysis was obtained by running k-Means on a random cluster initialisation, with information radius as distance measure; the verb description contained condition 2 subcategorisation frame types with PP information.

a) **ahnen, vermuten, wissen** (0.75) *Propositional Attitude*
b) **denken, glauben** (0.33) *Propositional Attitude*
c) **anfangen, aufhören, beginnen**, beharren, **enden**, insistieren, rudern (0.88) *Aspect*
d) **liegen, sitzen, stehen** (0.75) *Position*
e) **dienen, folgen, helfen** (0.75) *Support*
f) **nieseln, regnen, schneien** (0.75) *Weather*
g) **dämmern** (0.00) *Weather*
h) **blitzen, donnern**, segeln (0.25) *Weather*
i) **bestehen**, ***fahren, fliegen***, **pochen** (0.4) *Insisting* or *Manner of Motion*
j) **freuen, ärgern** (0.33) *Emotion*
k) **essen, konsumieren, saufen, trinken**, verkünden (1.00) *Consumption*
l) **bringen**, eröffnen, lesen, **liefern, schicken**, schließen, **vermitteln**, öffnen (0.78) *Supply*

[3] Verbs that are part of the majority are shown in **bold face**, others in plain text. Where there is no clear majority, both class labels are given.

distance	evaluation	k-Means cluster centre initialisation					
		random	hierarchical				
			single	complete	average	centroid	ward
irad	APP	0.125	0.043	0.087	0.079	0.073	0.101
	MI	0.328	0.226	0.277	0.262	0.250	0.304
skew	APP	0.111	0.043	0.091	0.067	0.062	0.102
	MI	0.315	0.226	0.281	0.256	0.252	0.349

Figure 1: Cluster quality variation based on condition 1 verb descriptions

distance	evaluation	k-Means cluster centre initialisation					
		random	hierarchical				
			single	complete	average	centroid	ward
irad	APP	0.144	0.107	0.123	0.118	0.081	0.151
	MI	0.357	0.229	0.319	0.298	0.265	0.332
skew	APP	0.114	0.104	0.126	0.118	0.081	0.159
	MI	0.320	0.289	0.330	0.298	0.265	0.372

Figure 2: Cluster quality variation based on condition 2 verb descriptions

m) ankündigen, beenden, bekanntgeben, ***bekommen***, **beschreiben, charakterisieren, darstellen**, ***erhalten, erlangen***, **interpretieren**, ***kriegen***, unterstützen (1.00) *Description* and *Obtaining*

n) **zustellen** (0.00) *Supply*

We compared the clustering to the gold standard and examined the underlying verb frame distributions. We undertook a series of post-hoc cluster analyses to explore the influence of specific frames and frame groups on the formation of verb classes, such as: what is the difference in the clustering result (on the same starting clusters) if we deleted all frame types containing an expletive *es* (frame types including `x`)? Space limitations allow us only a few insights.

- Clusters (a) and (b) are pure sub-classes of the semantic verb class *Propositional Attitude*. The verbs agree in their syntactic subcategorisation of a direct object (`na`) and finite clauses (`ns-2, ns-dass`); *denken* and *glauben* are assigned to a different cluster, because they also appear as intransitives, subcategorise the prepositional phrase `Akk.an`, and show especially strong probabilities for `ns-2`. Deleting `na` or frames containing `s` from the verb description destroys the coherent clusters.
- Cluster (c) contains two sub-classes from *Aspect* and *Insistence*, polluted by the verb *rudern* 'to row'. All *Aspect* verbs show a 50% preference for an intransitive usage, and a minor 20% preference for the subcategorisation of non-finite clauses. By mistake, the infrequent verb *rudern* (corpus frequency 49) shows a similar preference for `ni` in its frame distribution and therefore appears within the same cluster as the *Aspect* verbs. The frame confusion has been caused by parsing mistakes for the infrequent verb; `ni` is not among the frames possibly subcategorised by *rudern*.
  Even though the verbs *beharren* and *insistieren* have characteristic frames `np:Dat.auf` and `ns-2`, they share an affinity for `n` with the aspect verbs. When eliminating `n` from the feature description of the verbs, the cluster is reduced to those verbs using `ni`.
- Cluster (d) is correct: *Position*. The syntactic usage of the three verbs is rather individual with strong probabilities for `n, np:Dat.auf and np:Dat.in`. Even the elimination of any of the three frame features does not cause a separation of the verbs in the clustering.
- Cluster (j) represents the semantic class *Emotion* which, in German, has a highly characteristic signature in its strong association with reflexive frames; the cluster evaporates if we remove the distinctions made in the `r` feature group.
- *zustellen* in cluster (n) represents a singleton

because of its extraordinarily strong preference (>50%) for the ditransitive usage. Eliminating the frame from the verb description assigns *zustellen* to the same cluster as the other verbs of *Transfer of Possession (Supply)*.

Recall that we used two different sets of syntactic frames, the second of which makes more delicate distinctions in the area of prepositional phrases. As pointed out in Section 5, refining the syntactic verb information by PPs was helpful for the semantic clustering. But, contrary to our original intuitions, the detailed prepositional phrase information is less useful in the clustering of verbs with obligatory PP arguments than in the clustering of verbs where the PPs are optional; we performed a first test on the role of PP information: eliminating all PP information from the verb descriptions (not only the delicate PP information in condition 2, but also PP argument information in the coarse condition 1 frames) produced obvious deficiencies in most of the semantic classes, among them *Weather* and *Support*, whose verbs do not require PPs as arguments. A second test confirmed the finding: we augmented our coarse-grained verb frame repertoire with a much reduced set of PPs, those commonly assumed as argument PPs. This provides some but not all of the PP information in condition 2. The clustering result is deficient mainly in its classification of the verbs of *Propositional Attitude, Support, Opening*, and few of these subcategorise for PPs.

Clusters such as (k) to (l) suggest directions in which it might be desirable to subdivide the verb frames, for example by adding a limited amount of information about selectional preferences. Previous work has shown that sparse data issues preclude across the board incorporation of selectional information (Schulte im Walde, 2000), but a rough distinction such as physical object vs. abstraction on the direct object slot could, for example, help to split *verkünden* from the other verbs in cluster (k).

The linguistic investigation gives some insight into the reasons for the success of our (rather simple) clustering technique. We successfully exploited the connection between the syntactic behaviour of a verb and its meaning components. The clustering result shows a good match to the manually defined semantic verb classes, and in many cases it is clear which of and how the frames are influential in the creation of which clusters. We showed that we acquired implicit components of meaning through a syntactic extraction from a corpus, since the semantic verb classes are strongly related to the patterns in the syntactic descriptors. Everything in this study suggests that the move to larger datasets is an appropriate next move.

7 Conclusion

The paper presented the application of k-Means to the task of inducing semantic classes for German verbs. Based on purely syntactic probability distributions over verb subcategorisation frames, we obtained an intuitively plausible clustering of 57 verbs into 14 classes. The automatic clustering was evaluated against hand-constructed semantic verb classes. A series of post-hoc cluster analyses explored the influence of specific frames and frame groups on the coherence of the verb classes, and supported the tight connection between the syntactic behaviour of the verbs and their meaning components.

Future work will concern the extension of the clustering experiments to a larger number of verbs, both for the scientific purpose of refining our understanding of the semantic and syntactic status of verb classes and for the more applied goal of creating a large, reliable and high quality lexical resource for German. For this task, we will need to further refine our verb classes, further develop the repertoire of syntactic frames which we use, perhaps improve the statistical grammar from which the frames were extracted and find techniques which allow us to selectively include such information about selectional preferences as is warranted by the availability of training data and the capabilities of clustering technology.

References

Bonnie J. Dorr and Doug Jones. 1996. Role of Word Sense Disambiguation in Lexical Acquisition: Predicting Semantics from Syntactic Cues. In *Proceedings of the 16th International Conference on Computational Linguistics*, Copenhagen, Denmark.

Bonnie Dorr. 1997. Large-Scale Dictionary Construction for Foreign Language Tutoring and Interlingual Machine Translation. *Machine Translation*, 12(4):271–322.

Dudenredaktion, editor. 2001. *DUDEN – Das Stilwörterbuch*. Number 2 in 'Duden in zwölf Bänden'. Dudenverlag, Mannheim, 8th edition.

Edward W. Forgy. 1965. Cluster Analysis of Multivariate Data: Efficiency vs. Interpretability of Classifications. *Biometrics*, 21:768–780.

Gerhard Helbig and Joachim Buscha. 1998. *Deutsche Grammatik*. Langenscheidt – Verlag Enzyklopädie, 18th edition.

Leonard Kaufman and Peter J. Rousseeuw. 1990. *Finding Groups in Data – An Introduction to Cluster Analysis*. Probability and Mathematical Statistics. John Wiley and Sons, Inc.

Judith L. Klavans and Min-Yen Kan. 1998. The Role of Verbs in Document Analysis. In *Proceedings of the 17th International Conference on Computational Linguistics*, Montreal, Canada.

Maria Lapata. 1999. Acquiring Lexical Generalizations from Corpora: A Case Study for Diathesis Alternations. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*, pages 397–404.

Lillian Lee. 2001. On the Effectiveness of the Skew Divergence for Statistical Language Analysis. *Artificial Intelligence and Statistics*, pages 65–72.

Beth Levin. 1993. *English Verb Classes and Alternations*. The University of Chicago Press, Chicago, 1st edition.

Diana McCarthy. 2001. *Lexical Acquisition at the Syntax-Semantics Interface: Diathesis Alternations, Subcategorization Frames and Selectional Preferences*. Ph.D. thesis, University of Sussex.

Detlef Prescher, Stefan Riezler, and Mats Rooth. 2000. Using a Probabilistic Class-Based Lexicon for Lexical Ambiguity Resolution. In *Proceedings of the 18th International Conference on Computational Linguistics*, Saarbrücken.

Stefan Riezler, Detlef Prescher, Jonas Kuhn, and Mark Johnson. 2000. Lexicalized Stochastic Modeling of Constraint-Based Grammars using Log-Linear Measures and EM Training. In *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics*, Hong Kong.

Helmut Schmid. 2000. Lopar: Design and Implementation. Arbeitspapiere des Sonderforschungsbereichs 340 *Linguistic Theory and the Foundations of Computational Linguistics* 149, Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.

Sabine Schulte im Walde, Helmut Schmid, Mats Rooth, Stefan Riezler, and Detlef Prescher. 2001. Statistical Grammar Models and Lexicon Acquisition. In Christian Rohrer, Antje Rossdeutscher, and Hans Kamp, editors, *Linguistic Form and its Computation*. CSLI Publications, Stanford, CA.

Sabine Schulte im Walde. 2000. Clustering Verbs Semantically According to their Alternation Behaviour. In *Proceedings of the 18th International Conference on Computational Linguistics*, pages 747–753, Saarbrücken, Germany.

Sabine Schulte im Walde. 2002a. Evaluating Verb Subcategorisation Frames learned by a German Statistical Grammar against Manual Definitions in the *Duden* Dictionary. In *Proceedings of the 10th EURALEX International Congress*, Copenhagen, Denmark. To appear.

Sabine Schulte im Walde. 2002b. A Subcategorisation Lexicon for German Verbs induced from a Lexicalised PCFG. In *Proceedings of the 3rd Conference on Language Resources and Evaluation*, Las Palmas de Gran Canaria, Spain. To appear.

Helmut Schumacher. 1986. *Verben in Feldern*. de Gruyter, Berlin.

Suzanne Stevenson and Paola Merlo. 1999. Automatic Verb Classification Using Distributions of Grammatical Features. In *Proceedings of the 9th Conference of the European Chapter of the Association for Computational Linguistics*, pages 45–52.

Alexander Strehl, Joydeep Ghosh, and Raymond Mooney. 2000. Impact of Similarity Measures on Web-page Clustering. In *Proceedings of the 17th National Conference on Artificial Intelligence (AAAI 2000): Workshop of Artificial Intelligence for Web Search*, Austin, Texas.

Marc Vilain, John Burger, John Aberdeen, Dennis Connolly, and Lynette Hirschman. 1995. A Model-Theoretic Coreference Scoring Scheme. In *Proceedings of the 6th Message Understanding Conference*, pages 45–52, San Francisco.

Gloria Vázquez, Ana Fernández, Irene Castellón, and M. Antonia Martí. 2000. *Clasificación Verbal: Alternancias de Diátesis*. Number 3 in Quaderns de Sintagma. Universitat de Lleida.