

Clustering Verbs Semantically According to their Alternation Behaviour

Sabine Schulte im Walde

Institut für Maschinelle Sprachverarbeitung
Universität Stuttgart

Azenbergstraße 12, 70174 Stuttgart, Germany
schulte@ims.uni-stuttgart.de

Abstract

Verbs were clustered semantically on the basis of their alternation behaviour, as characterised by their syntactic subcategorisation frames extracted from maximum probability parses of a robust statistical parser, and completed by assigning WordNet classes as selectional preferences to the frame arguments. The clustering was achieved (a) iteratively by measuring the relative entropy between the verbs' probability distributions over the frame types, and (b) by utilising a latent class analysis based on the joint frequencies of verbs and frame types.

1 Introduction

This paper empirically investigates the proposition that verbs can be semantically classified according to their syntactic alternation behaviour concerning subcategorisation frames and their selectional preferences for the arguments within the frames. The idea is related to (Levin, 1993) who defined verb classes on the basis of verb alternation behaviour. For example, the semantic class of *Vehicle Names* contains verbs like *balloon*, *bicycle*, *canoe*, *skate*, *ski* which agree in the properties (1)-(4) below.

- (1) INTRANSITIVE USE, possibly followed by a path:
 - a. They skated.
 - b. They skated along the canal/over the bridge.
- (2) INDUCED ACTION ALTERNATION (some verbs):
 - a. He skated Penny around the rink.
(causing the action named by the verb;
typical causee is an animate volitional
entity)
 - b. Penny skated around the rink.
- (3) LOCATIVE PREPOSITION DROP ALTERNATION (some verbs):
 - a. They skated along the canals.
 - b. They skated the canals.

- (4) RESULTATIVE PHRASE:

Penny skated her skate blades blunt.

(an XP describing the state achieved by the referent of the noun phrase as a result of the action named by the verb)

Levin's work represents the basis for a range of recent investigations verifying (Dorr and Jones, 1996), evaluating (Stevenson and Merlo, 1999) or utilising (Lapata, 1999) the proposed classification as well as transferring it to other languages than English (Jones et al., 1994).

Generally, the definition of a verb's semantic class can be considered as part of its lexical entry, next to idiosyncratic information: the semantic class generalises as a type definition over a range of syntactic and semantic properties, to support Natural Language Processing in various areas like lexicography (Rappaport Hovav and Levin, 1998), word sense disambiguation (Dorr and Jones, 1996), or document classification (Klavans and Kan, 1998).

I attempted to automatically cluster verbs into semantic classes on the basis of the verbs' alternation behaviour. The input into the automatic induction process was characterised by the verbs' distribution over syntactic subcategorisation frames as extracted from maximum probability (Viterbi) parses of a robust statistical parser, and completed by assigning WordNet classes as selectional preferences to the frame arguments. The clustering was achieved (a) iteratively by measuring the relative entropy between the verbs' probability distributions over the frame types, and (b) by utilising a latent class analysis based on the joint frequencies of verbs and frame types. Using Levin's verb classification as evaluation basis, 61% of the verbs were classified correctly into semantic classes by method (a), and 54% by method (b).

Section 2 describes the three steps in the automatic acquisition of semantic verb classes; the evaluation takes place in section 3, and section 4 discusses the results.

2 Automatic Acquisition of Semantic Verb Classes

The first step was the induction of purely syntactic subcategorisation frames for verbs from the heterogeneous *British National Corpus (BNC)*. I used the robust statistical head-entity parser as described in (Carroll and Rooth, 1998) which utilises an English context-free grammar and a lexicalised probability model to produce parse forests, and extracted the maximum probability (Viterbi) parses, for a total of 5.5 million sentences. The trees were mapped to subcategorisation frame tokens consisting of a main verb and its arguments. Each syntactic category was accompanied by the lexical head, the prepositional phrase by the lexical prepositional head plus the head noun of the subordinated noun phrase. Proper names were accompanied by the identifier *pn*. The head information in the frames was lemmatised. For example, the sentence *Sammut handled the plaudits during the awards ceremony* would be represented by the frame token

handle subj*pn**sammut* obj*plaudit pp*during*ceremony.

To generalise over the verbs' usage of subcategorisation frames, I defined as 88 frame types the most frequent frames which appeared at least 2,000 times in total in the BNC sentence parses, disregarding the lexical head information. On the basis of the frame types I collected information about the joint frequencies of the verbs in the BNC and the subcategorisation frame types they appeared with. These frequency counts then represented the syntactic description of the verbs.

The next step was to refine the subcategorisation frame types by a preferential ordering on conceptual classes for the argument slots in the frames. The basis I could use for the selectional preferences was provided by the lexical heads in the frame tokens. For example, the nouns appearing in the direct object slot of the transitive frame for the verb *drink* included *coffee*, *milk*, *beer*, indicating a conceptual class like *beverage* for this argument slot.

I followed (Resnik, 1993)/(Resnik, 1997) who defined *selectional preference* as the amount of information a verb provides about its semantic argument classes. He utilised the WordNet taxonomy (Beckwith et al., 1991) for a probabilistic model capturing the co-occurrence behaviour of verbs and conceptual classes, where the conceptual classes were identified by WordNet synsets, sets of synonymous nouns within a semantic hierarchy. Referring to the above example, the three nouns *coffee*, *milk*, *beer* are in three different synsets—since they are not synonyms—, but are all subordinated to the synset *{beverage, drink, potable}*. The goal in this example would therefore be to determine the relevant synset as the most selectionally preferred synset for the direct object slot of the verb *drink*.

Redefined for my usage, the selectional preference of a verb *v* for a certain semantic class *c* within a subcategorisation frame slot *s* was determined by the association *ass* between verb and semantic class:

$$ass(v_s, c_s) =_{def} p(c_s|v_s) \log \frac{p(c_s|v_s)}{p(c_s)} \quad (5)$$

with the probabilities estimated by maximum likelihood:

$$p(c_s|v_s) = \frac{f(v_s, c_s)}{f(v_s)} \quad (6)$$

$$p(c_s) = \frac{f(c_s)}{\sum_{c' \in class} f(c')} = \frac{f(c_s)}{f(s)} \quad (7)$$

and the following interpretation:

1. $f(v_s, c_s)$: number of times a semantic class appeared in a frame slot of a verb's frame type
2. $f(v_s)$: frequency of a verb regarding a specific frame type, i.e. the joint frequency of verb and frame type
3. $f(c_s)$: number of times a semantic class appeared in a frame slot of a frame type disregarding the verb
4. $\sum_{c' \in class} f(c')$ equals $f(s)$, the frequency of the argument slot within a certain frame type, since summing over all possible classes within a subcategorisation frame slot equals the number of times the slot appeared
5. $f(s)$: number of times the frame type appeared, since the frequency of a frame type equals the frequency of that frame with a certain slot marked

The frequencies of a semantic class concerning an argument slot of a frame type (dependent or independent of a verb) were calculated by an approach slightly different to Resnik's, originally proposed by (Ribas, 1994)/(Ribas, 1995). For each noun appearing in a certain argument position its frequency was divided by the number of senses the noun was assigned by the WordNet hierarchy,¹ to take account of the uncertainty about the sense of the noun. The fraction was allocated to each conceptual class in the hierarchy to which the noun belonged and accumulated upwards until a top node was reached. The result was a numerical distribution over the WordNet classes:

$$f(c_s) = \sum_{noun \in c_s} \frac{f(noun)}{|senses(noun)|} \quad (8)$$

¹For example, when considering the noun *coffee* isolated from its context, we do not know whether we are talking about the beverage *coffee*, the plant *coffee* or a *coffee* bean. Therefore, a third of the frequency of the noun was assigned to each of the three classes.

I restricted the possible conceptual classes within the frames' argument slots to 23 WordNet nodes,² to facilitate generalisation and comparison of the verbs' selectional preference behaviour.

On the basis of the information about subcategorisation frame types and their arguments' conceptual classes I clustered 153 verbs from Levin's classification. I chose (i) some polysemous verbs to investigate how this phenomenon could be handled by the clustering algorithms, and (ii) high and low frequent verbs to see the influence of frequency on the algorithms: the 153 verbs had 226 verb senses which belonged to 30 different semantic classes. Four of the verbs were low-frequency verbs with a total corpus frequency below 100.

To cluster the verbs I applied two different algorithms, and each algorithm clustered the verbs both (A) according to only the syntactic information about the subcategorisation frames, and (B) according to the information about the subcategorisation frames including their selectional preferences.

- *Iterative clustering based on a definition by (Hughes, 1994):*

In the beginning, each verb represented a singleton cluster. Iteratively, the distances between the clusters were measured and the closest clusters merged together.

For the representation of the verbs, each verb v was assigned a distribution over the different types of subcategorisation frames t , according to the maximum likelihood estimate of (A) the verb appearing with the frame type:

$$p(t|v) = \frac{f(v, t)}{f(v)} \quad (9)$$

with $f(v, t)$ the joint frequency of verb and frame type, and $f(v)$ the frequency of the verb, and (B) the verb appearing with the frame type and a selectionally preferred class combination C for the argument positions s in t :

$$p(t, C|v) =_{def} p(t|v) * p(C|v, t) \quad (10)$$

with $p(t|v)$ defined as in equation (9), and

$$p(C|v, t) =_{def} \frac{\prod_{s \in t} ass(v_s, c_s)}{\sum_{c'_s \in class} \prod_{s \in t} ass(v_s, c'_s)} \quad (11)$$

which intuitively estimates the probability of a certain class combination by comparing its association value with the sum over all possible class combinations, concerning the respective verb and frame.

²I chose the 11 top level nodes of the 11 WordNet hierarchies as conceptual classes. The top level node **Entity** seemed too general as conceptual class, so it was replaced by its 13 subordinated synsets.

Starting out with each verb representing a singleton cluster, I iteratively determined the two closest clusters by applying the information-theoretic measure *relative entropy*³ (Kullback and Leibler, 1951) to compare the distributions. The nearest clusters were merged into one cluster, and their distributions were merged by calculating a weighted average. Based on test runs I defined heuristics about how many clustering iterations were performed. In addition, I limited the maximum number of verbs within one cluster to four elements because otherwise the verbs showed the tendency to cluster together in a few large clusters only; so after the overall clustering process was finished, each cluster with more than four members initialised a further clustering pass on itself.

- *Unsupervised latent class analysis as described in (Rooth, 1998), based on the expectation-maximisation algorithm:*

The algorithm identified categorical types among indirectly observed multinomial distributions by applying the EM-algorithm (Dempster et al., 1977) to maximise the joint probability of (A) the verb and frame type: $p(v, t)$, and (B) the verb and frame type considering the selectional preferences: $p(v, t, C)$.

Input to the algorithm were absolute frequencies of the verbs appearing with the subcategorisation frames. Test runs showed that 80 clusters modelled the semantic verb classes best. To be able to compare the analysis with the iterative clustering approach, I also limited the number of verbs within a cluster to four – considering that generally all verbs appear within each cluster when using this approach, the verbs with the highest probabilities were chosen.

For version (A) the frequencies were provided by the joint frequencies of verbs and frame types, for version (B) I used the association values of the verbs with the frame types considering selectional preferences, as described by equation (10).

The unsupervised algorithm then classified joint events of verbs and subcategorisation frames with 200 iterations of the EM-algorithm into 80 clusters τ , based on the iteratively estimated values

$$p(v, t) = \sum_{\tau} p(\tau, v, t) = \sum_{\tau} p(\tau) p(v|\tau) p(t|\tau) \quad (12)$$

³Concerning the two typical problems one has with this measure, (i) zero frequencies were smoothed by adding 0.5 to all frequencies, and (ii) since the measure is not symmetric, the respective smaller value was used as distance.

Information	Clusters		Verbs			
	Total	Correct	Total	Correct	Recall	Precision
SFs	31	20	90	55	36%	61%
SFs + Prefs	30	14	81	31	20%	38%

Figure 1: Evaluation based on Iterative Clustering

Information	Clusters		Verbs(Senses)			
	Total	Correct	Total	Correct	Recall	Precision
SFs	80	36	107(159)	58(90)	38(40)%	54(57)%
SFs + Prefs	80	22	153(226)	47(56)	31(25)%	31(25)%

Figure 2: Evaluation based on Latent Classes

$$p(v, t, C) = \sum_{\tau} p(\tau, v, t, C) = \sum_{\tau} p(\tau)p(v|\tau)p(t, C|\tau) \quad (13)$$

for versions (A) and (B), respectively.

3 Evaluation

The evaluation of the resulting clusters was based on Levin’s classification. Figures 1 and 2 present the success of the two clustering algorithms, considering the two different informational versions (A) and (B). They contain the total number of clusters the algorithms had formed (clusters containing between two and four verbs in the iterative algorithm, and the fixed number of 80 clusters in the latent class analysis), the proportion of correct clusters (non-singleton clusters which were subsets of a Levin class, for example the cluster containing the verbs *need*, *like*, *want*, *desire* is a subset of the Levin class *Desire*), and the number of verbs within those clusters. In figure 2 the number of verbs in brackets refers to the respective number of their senses, since a verb could be clustered several times according to its senses. For example, the verb *want* could be member of the classes *Desire* and *Declaration*.

Recall was defined by the percentage of verbs (verb senses) within the correct clusters compared to the total number of verbs (verb senses) to be clustered:

$$rec = \frac{\frac{|verbs_{correct\ clusters}|}{153}}{\left(\frac{|verb\ senses_{correct\ clusters}|}{226}\right)}$$

and precision was defined by the percentage of verbs (verb senses) appearing in the correct clusters compared to the number of verbs (verb senses) appearing in any cluster:

$$prec = \frac{\frac{|verbs_{correct\ clusters}|}{|verbs_{all\ clusters}|}}{\left(\frac{|verb\ senses_{correct\ clusters}|}{|verb\ senses_{all\ clusters}|}\right)}$$

Concerning precision, the assignment of verbs into semantic classes was most successful when using the

iterative distance clustering method; 61% of all verbs were clustered into correct classes. Clustering the verbs into latent classes was with 54% less successful. With both clustering methods the results became worse when adding information about the selectional preferences for the arguments in the subcategorisation frames.

A baseline experiment was performed in order to determine how hard the task of verb clustering was: each verb was randomly assigned another verb as "closest neighbour", which resulted in only 5% of the verbs being paired with a verb from the same Levin class. Performing the same experiment by assigning the closest neighbour on the basis of measuring the relative entropy between two verbs’ distributions over subcategorisation frames resulted in 61% of the verbs pointing to a verb from the same Levin class.

4 Discussion⁴

The classifications of both clustering approaches illustrate the close relationship between alternation behaviour and semantic classes. For example, the common preferences of verbs (see the five most probable frames) in the iteratively created *Desire* class were towards a subject followed by an infinitival phrase (**subj:to**). Alternatively a transitive **subj:obj** frame was used, partly followed by an additional infinitival phrase indicated by **to**:⁵

⁴For a more detailed discussion see the original work (Schulte im Walde, 1998).

⁵Note that the (wrongly chosen) intransitive frame is listed as well. This is due to underlying sentences containing an NP ellipsis, parsing mistakes and frame extraction.

Verb	Frame	Probability
need	subj:to	0.38
	subj:obj	0.32
	subj	0.10
	subj:obj:to	0.05
	subj:obj:pp.for	0.02
like	subj:to	0.34
	subj:obj	0.34
	subj	0.14
	subj:obj:adv	0.04
	subj:obj:obj	0.03
want	subj:to	0.53
	subj:obj	0.15
	subj	0.11
	subj:obj:to	0.10
	subj:to:adv	0.02
desire	subj:obj	0.25
	subj	0.24
	subj:to	0.20
	subj:obj:to	0.07
	subj:sent	0.02

Adding information about the selectional preferences of the verbs' arguments helps to get a deeper idea about their lexical semantics. For example, *Manner of Motion* verbs preferably appeared with a subject only, sometimes with a following adverb. The subject was an inanimate object, for *move* it might also be a part (such as a body part like *finger*) or a group. *roll* and *fly* alternatively used the transitive frame type `subj:obj`, preferably with a living entity as subject, followed by an inanimate object:

Verb	Frame	Probability
roll	subj(PhysObject)	0.24
	subj(PhysObject):adv	0.10
	subj(Agent):obj(PhysObject)	0.07
	subj(LifeForm):obj(PhysObject)	0.07
	subj(Agent):obj(Part)	0.05
fly	subj(PhysObject)	0.34
	subj(PhysObject):adv	0.12
	subj(LifeForm):obj(PhysObject)	0.07
	subj(LifeForm):pp.to(LifeForm)	0.05
	subj(LifeForm):pp.to(Agent)	0.04
move	subj(PhysObject)	0.20
	subj(PhysObject):adv	0.11
	subj(Part)	0.09
	subj(Group):adv	0.04
	subj(Part):adv	0.04

Parallel examples created by the latent class analysis present the clusters with the most probable verbs and frames, according to cluster membership (first column). The dot indicates whether the verb-frame combination was seen in the data, the number next to the verb frame gives the probability of the verb-frame combination.

Some verbs of *Telling* were clustered mainly according to their similar transitive use combined with an infinitival phrase:

Cluster		0.75	0.09	0.05	0.02
		subj:obj:to	subj	subj:obj	subj:pp.on
0.17	advise	•	•	•	•
0.12	teach	•	•	•	•
0.12	instruct	•	•	•	•

The verbs of *Aspect* alternate between a subject only, realised by an action, an inanimate subject followed by an infinitival phrase, and a living subject followed by a gerund:

Cluster		0.22	0.10	0.09	0.09
		subj(Action)	subj(PhysObject):to	subj(LifeForm):vger	subj(Agent):vger
0.34	start	•	•	•	•
0.19	finish	•	•	•	•
0.18	stop	•	•	•	•
0.16	begin	•	•	•	•

Both approaches established a relationship between alternation behaviour and semantic class by only considering information about the syntactic usage of the subcategorisation frames. The refinement by the frames' selectional preferences allowed further demarcations by the identifying conceptual restrictions on the use of the frames.

Since the latent class analysis is a soft clustering method, it additionally distinguishes between the different verbs' senses and the respective uses of subcategorisation frames. For example, the verb *play* was clustered with *meet* because of the common strong tendency towards a transitive frame illustrating a general meeting, and it was clustered with *fight* because of their common preference for an intransitive frame together with a prepositional phrase headed by *against*, illustrating a more aggressive meeting like a fight:

Cluster		0.55	0.05	0.04	0.03
		subj:obj	subj	subj:obj:pp.with	subj:obj:pp.at
0.49	meet	•	•	•	•
0.20	play	•	•	•	•
Cluster		0.18	0.13	0.09	0.07
		subj:pp-against	subj:obj	subj:obj:pp-against	subj:obj:adv
0.22	fight	•	•	•	•
0.20	play	•	•	•	•

An extensive investigation of the linguistic reliability of the clustered verbs and frames showed that the characterising usages could be underlined by corpus data, for example the above cited transitive use

of the verb *fly* concerning the `subj:obj` frame type with a living subject and an inanimate object can be illustrated by the BNC-sentence *In March the manufacturer's test pilot flew the aircraft for its annual inspection check flight*. The clusters were therefore created on a reliable linguistic basis representing (a selective part of) the verbs' properties.

Comparing the two informational versions, however, showed that refining the frames with selectional preferences points to a problem caused by data sparseness in the verb description. Investigating the automatically created distribution of the verbs over the enriched frame types revealed that, for example, even the high frequent, alternating verb *move* contains 97% (smoothed) zeroes within its distribution. In accordance with this finding even subtle similarities, e.g. the sole fact that two verbs have non-zero values for certain frame types, highly correlates the two verbs. For example, a semantic cluster contained the two verbs *promise* and *love*, because both have non-zero attribute values for the `subj:to` frame, demanding an agent for the subject slot; in their alternation behaviour (including selectional preferences) the two verbs differ, however, so they should not be packed into one cluster. A possible suggestion to handle the problem of data sparseness could be to formulate the conceptual class types in a way which ensures an increased data potential for each type.

Concerning the polysemy of verbs, the (hard) iterative distance clustering failed to model verb senses; a polysemous verb was either not at all assigned to any cluster, or assigned to a cluster describing one of the verb's senses. The (soft) latent class analysis was able to filter the multiple senses and assign them to distinct clusters, but tended to split senses. Low-frequency verbs presented another problem, because the verbs' distributions contained mostly zeroes. They were assigned to clusters nearly randomly.

An investigation of selected WordNet conceptual classes revealed that the selectional preferences within the subcategorisation frames were dominated by a few WordNet classes, mainly `LifeForm` and `Agent`. The demarcation between these two concepts was not obvious when referring to actually appearing nouns within the frames, since both contain a large number of common subordinated nouns. In contrast, some WordNet classes were not chosen at all, e.g. `Unit` or `Anticipation`. Since the WordNet hierarchy in general had turned out to define intuitively correct selectional preferences, an improved classification utilised for my conceptual classification should be substituted by finer synsets, i.e. one should consider using a different cut through the WordNet hierarchy.

5 Conclusion

I proposed two algorithms for automatically classifying verbs semantically, based on their alternation behaviour. Taking Levin's classification as a standard for 153 manually chosen verbs with 226 verb senses and their assignment into 30 semantic classes, the iterative distance clustering succeeded for 61% of the verbs considering the syntactic usage of the frames only, and for 38% when adding information about the frame arguments' selectional preferences. The latent class analysis succeeded for 54% and 31%, respectively.

An investigation of the resulting clusters showed that the assignment of the verbs was actually based on their shared linguistic properties: the verbs in a cluster presented common alternation behaviour, refined by adding selectional preferences to the syntactic description of the subcategorisation frames.

It is impressive that as little lexical idiosyncratic verb information as the syntactic use of subcategorisation frames like `subj:to` or `subj:pp.against` suffices as a basis for a semantic class distinction towards Levin's narrow classification system including fine concepts as *Desire* or *Manner of Motion*. The potential is partly characterised by specific frames, but in the majority of cases by successfully combining the frames in order to define the syntactic alternation. Improving the definition and demarcation of conceptual classes should provide further potential concerning the inclusion of selectional preferences into the syntactic description.

References

- Richard Beckwith, Christiane Fellbaum, Derek Gross, and George A. Miller. 1991. Wordnet: A Lexical Database Organized on Psycholinguistic Principles. In Uri Zernik, editor, *Lexical Acquisition – Exploiting On-Line Resources to Build a Lexicon*, chapter 9, pages 211–232. Lawrence Erlbaum Associates, Hillsdale - New Jersey.
- Glenn Carroll and Mats Rooth. 1998. Valence Induction with a Head-Lexicalized PCFG. In *Proceedings of the 3rd Conference on Empirical Methods in Natural Language Processing*, Granada, Spain.
- A. P. Dempster, N. M. Laird, and D. B. Rubin. 1977. Maximum Likelihood from Incomplete Data via the EM algorithm. *Journal of the Royal Statistical Society*, 39(B):1–38.
- 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.
- John Hughes. 1994. *Automatically Acquiring Classification of Words*. Ph.D. thesis, University of Leeds, School of Computer Studies.

- Douglas A. Jones, Robert C. Berwick, Franklin Cho, Zeeshan Khan, Karen T. Kohl, Naoyuki Nomura, Anand Radhakrishnan, Ulrich Sauerland, and Brian Ulicny. 1994. Verb Classes and Alternations in Bangla, German, English, and Korean. Technical Report MIT AI MEMO 1517, Massachusetts Institute of Technology.
- 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.
- S. Kullback and R. A. Leibler. 1951. On Information and Sufficiency. *Annals of Mathematical Statistics*, 22:79–86.
- 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.
- Beth Levin. 1993. *English Verb Classes and Alternations*. The University of Chicago Press, Chicago, 1st edition.
- Malka Rappaport Hovav and Beth Levin. 1998. Building Verb Meanings. In M. Butt and W. Geuder, editors, *Lexical and Compositional Factors*, pages 97–134. CSLI Publications, Stanford, CA.
- Philip Resnik. 1993. *Selection and Information: A Class-Based Approach to Lexical Relationships*. Ph.D. thesis, University of Pennsylvania.
- Philip Resnik. 1997. Selectional Preference and Sense Disambiguation. In *Proceedings of the ACL SIGLEX Workshop on Tagging Text with Lexical Semantics: Why, What, and How?*
- Francesc Ribas. 1994. An Experiment on Learning Appropriate Selectional Restrictions from a Parsed Corpus. In *Proceedings of the 15th International Conference on Computational Linguistics*, pages 769–774.
- Francesc Ribas. 1995. On Learning More Appropriate Selectional Restrictions. In *Proceedings of the 7th Conference of the European Chapter of the Association for Computational Linguistics*, Dublin, Ireland.
- Mats Rooth. 1998. Two-Dimensional Clusters in Grammatical Relations. In *Inducing Lexicons with the EM Algorithm*, AIMS Report 4(3). Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.
- Sabine Schulte im Walde. 1998. Automatic Semantic Classification of Verbs According to Their Alternation Behaviour. Master's thesis, Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.
- Suzanne Stevenson and Paola Merlo. 1999. Automatic Verb Classification Using Distributions of Grammatical Features. In *Proceedings of the 9th*