Automatic Semantic Classification of Verbs According to their Alternation Behaviour [*]

Sabine Schulte im Walde

May 31, 1999

Abstract

An automatic semantic classification of verbs was performed by first determining the verbs' alternation behaviour and then clustering the verbs on that basis. The alternation behaviour of the verbs was outlined by inducing syntactic subcategorisation frames from maximum probability (Viterbi) parses of a robust statistical parser, completed by assigning WordNet classes to the frames' arguments. The clustering was achieved (a) iteratively by measuring the relative entropy between the verbs' probability distributions over the different types of frames, and (b) by utilising a latent class analysis based on the joint frequencies of verbs and frame types. Using Levin's verb classification [8] as evaluation basis, (a) 61% and (b) 54% of the verbs were classified correctly into semantic classes.

[*] This article is a short version of my diploma thesis [14], published in *Inducing Lexicons with the EM Algorithm*, AIMS Report 4(3). Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart, 1998.

1 Motivation

For this work I assumed that the diathesis alternation of verbs, i.e. the alternation in the expression of the verbs' arguments, is a basis for the comparison of the verbs' meanings. More specifically, I empirically investigated 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 of a semantic classification according to alternation behaviour is related to Levin [8] who defined verb classes on the basis of the verbs' alternation behaviour. Consider, for example, the semantic class of *Vehicle Names* containing verbs like *balloon, bicycle, canoe, skate, ski* because they agree in the following properties:

(1) Intransitive Use, possibly followed by a path:

a. They skated.

b. They skated along the canal/across the lake.

(2) Induced Action Alternation (some verbs):

a sub-type of Transitive Alternation, where the transitive use of the verb can be paraphrased as causing the action named by the verb; the causee is typically an animate volitional entity induced to act by the causer; the verb must be accompanied by a directional phrase

a. He skated Penny around the rink.

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:

an XP which describes the state achieved by the referent of the noun phrase it is predicated of as a result of the action named by the verb

Penny skated her skate blades blunt.

As Levin did, I attempted to derive verb classes from the verbs' behaviour. The information I fed into an automatic deduction process for semantic classes was thereby referring back to Chomsky's [3] demands for the utterance of verbs: the verbs' behaviour was defined by their subcategorisation rules and their selectional rules.

Such a definition of the 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 (by the enrichment of lexical knowledge), word sense disambiguation (by the provision of context information provided by the semantic verb type), or parsing (by the generalisation from verb tokens to verb types and the resulting restriction of syntactic structures). Klavans and Kan [6], for example, discriminate documents by type and semantic properties of the verbs within the documents.

2 Automatic Acquisition of Semantic Verb Classes

I empirically investigated the verbs' behaviour and their meanings by automatically inferring semantic verb classes with the help of data-intensive methods working on data from a large corpus, and by applying statistical methods proved useful for NLP-tasks. The inference process contained three main steps:

1. The induction of subcategorisation frames for verbs from a large corpus
2. The definition of selectional preferences for the subcategorisation frames
3. The clustering of the verbs into semantic verb classes, on account of the verbs' behaviour as defined in steps 1 and 2

Following sections 2.1 to 2.3 present the methods used for the three steps and their realisation.

2.1 Induction of Subcategorisation Frames

Within the first step of inducing purely syntactic subcategorisation frames for verbs I used the robust statistical head-entity parser as described in Carroll and Rooth [2] which utilises an English context-free grammar and a

lexicalised probability model to produce parse forests for sentences, where each sub-tree is annotated with information about the lexical head and the probability. I parsed the heterogeneous *British National Corpus (BNC)* and extracted the maximum probability (Viterbi) parses from the parse forests, for a total of 5.5 million sentences.
Based on the maximum probability parses I determined the main verb and all its arguments as the sentences' subcategorisation frame tokens. For example, the frame token of the sentence *Nobody excelled him in that judgement* would be defined by

`act*excelled subj*nobody obj*him pp*in*judgement`,

describing the full (active) verb form and the subject, object and prepositional phrase arguments as determined by the English grammar, each accompanied by its lexical head, the prepositional phrase accompanied by its lexical head and the head noun of the sub-ordinated noun phrase. I finished the frame description by lemmatising the head information in the subcategorisation frames.

To generalise over the verbs' usage of subcategorisation frames, I defined as 88 frame types those frames which appeared at least 2,000 times in total in the BNC sentence parses, disregarding the lexical head information. For example, the most frequent frame type was the transitive frame `subj:obj`. On the basis of the frame types I collected information about the joint frequencies of the verbs in the BNC and the subcategorisation frames they appeared with. Appendix A gives a full list of the 88 subcategorisation frame types and an example for the joint frequencies.

2.2 Selectional Preferences for Subcategorisation Frames

The next step after inducing the subcategorisation frame types was to refine the information by identifying 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 as determined in section 2.1, for example the nouns appearing in the object slot of the transitive frame for the verb *drink* included *coffee, milk, beer*, demanding a conceptual class like *beverage* for this argument slot.

I followed Resnik [9]/[10] who defined *selectional preference* as the amount of information a verb provides about its semantic argument classes. He utilised the WordNet taxonomy [1] 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 no synonyms –, but are all sub-ordinated to the synset defined by *beverage, drink, potable*. The goal in this example would therefore be to determine the relevant synset as the most selectionally preferred synset for the object slot of the verb *drink*.

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

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

with the probabilities estimated by maximum likelihood:

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

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

To facilitate the understanding of the equations I briefly interpret the relevant parts:

1. $f(v_s, c_s)$ was defined by how often a certain semantic class appeared in a certain frame slot of a verb's frame type.
2. $f(v_s)$ was defined by the frequency of a certain verb regarding a specific frame type, i.e. the joint frequency of verb and frame type as determined in section 2.1.
3. $f(c_s)$ was defined by how often a certain semantic class appeared in a certain frame slot of a frame type disregarding the verb.
4. $\sum_{c' \in class} f(c'_s)$ 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 was equal to the number of times the slot appeared.
5. $f(s)$ was defined by the number of times the frame type appeared (as determined in section 2.1), 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 [11]/[12]: 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 display the uncertainty about the sense of the noun.[2] The fraction was given each conceptual class in the hierarchy to which the noun belonged and projected upwards until a top node was reached. The result was a numerical distribution over the WordNet classes:

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

To give a further example about the amount of information we were provided with after this process, the verb *swim* with the frame type `subj:pp.in` (indicating a subject and a prepositional phrase headed by *in*) had its strongest preferences for the WordNet class *fish* as subject and *body of water* as prepositional phrase object.
For this work, however, I restricted the possible conceptual classes within the frames' argument slots to 23 WordNet (mostly top) level nodes, to facilitate generalisation and comparison of the verbs' selectional preference behaviour, and defined abbreviations for them. Appendix B gives an overview of those WordNet synsets and its member nouns.

2.3 Clustering Verbs into Semantic Verb Classes

On the basis of the information about subcategorisation frame types and their arguments' conceptual classes I could start to cluster verbs. For that, I selected verbs from Levin's classification. The constraints I required for the verbs were (i) some verbs to be polysemous to investigate the realisation of the phenomenon by the clustering algorithms, and (ii) to distinguish between high and low frequent verbs to see the influence of the frequency onto the algorithms. Therefore I selected 153 different verbs with 226 verb senses which belonged to 30 different semantic classes. Four of the verbs were low-frequent verbs.

[1] For 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.

[2] Intuitively, Ribas' approach was an improvement to Resnik's in this detail, since Resnik split the number of times a certain noun appeared in an argument position by the total number of classes it appeared in, up to the top of the hierarchy. This treatment made the uncertainty dependent on the depth of the hierarchy, though, not from the number of different senses.

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 as acquired in section 2.1 and (B) according to the information about the subcategorisation frames including their selectional preferences as completed in section 2.2.

- *Iterative clustering based on a definition by Hughes [5]:*

  In the beginning, each verb represented an own 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:

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

  with $f(v,t)$ being the joint frequency of verb and frame type, and $f(v)$ being the frequency of the verb, both as determined in section 2.1, and (B) the verb appearing with the frame type and a selectionally preferred class combination C for the argument positions s in t:

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

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

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

  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.

  Starting out with each verb representing an own cluster, I iteratively determined the two closest clusters by applying the information-theoretic measure *relative entropy*[3] [7] to compare the respective distributions. Those were merged into one cluster, and their distributions were merged by calculating a weighted average. Based on test runs I defined heuristics about how often the clustering was 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.

- *Unsupervised latent class analysis as described in Rooth [13], based on the expectation-maximization algorithm:*

  The algorithm identified categorical types among indirectly observed multinomial distributions by applying the EM-algorithm [4] 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: $p(v,t,C)$.

  It needed a fixed number of classes to be built and absolute frequencies of the verbs appearing with the subcategorisation frames. Test runs showed that 80 clusters modeled 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 respective probabilities where 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 within 200 iterations joint events of verbs and subcategorisation frames into the 80 clusters τ, based on the iteratively estimated values

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

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

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

3 Evaluation

The evaluation of the resulting clusters was adjusted to Levin's classification where the verbs had been taken from before. The following tables 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

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

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

Figure 1: Iterative Clustering

Information	Clusters		Verbs(Senses)		Recall	Precision
	Total	Correct	Total	Correct		
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: Latent Classes

of clusters the algorithms had formed (all clusters containing between two and four verbs concerning the iterative algorithm, and a fixed number of 80 clusters concerning the latent class analysis), the share of correct clusters (those 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 table 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:

$$recall = \frac{|verbs_{correct\_clusters}|}{153} \qquad (\frac{|verb\_senses_{correct\_clusters}|}{226}) \tag{14}$$

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:

$$precision = \frac{|verbs_{correct\_clusters}|}{|verbs_{all\_clusters}|} \qquad (\frac{|verb\_senses_{correct\_clusters}|}{|verb\_senses_{all\_clusters}|}) \tag{15}$$

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% comparably, but less successful. With both clustering methods the results became worse when adding information about the selectional preferences for the arguments in the subcategorisation frames.

4 Discussion

Following I present a choice of the correct clusters resulting from the different clustering approaches, in order to demonstrate that the classifications of both approaches illustrate the close relationship between the verbs' alternation behaviour and their affiliation to semantic classes: the resulting clusters which could be annotated by semantic class names show common alternation behaviour of their verbal elements.

The iteratively generated clusters show the verbs in the clusters followed by the five subcategorisation frame types with the highest probabilities in the overall verbs' distributions.
The preferences for verbs in the *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`:[4]

```
need          * subj:to            0.382847629835582 *
              * subj:obj           0.318590601723132 *
              * subj               0.0962654034943192 *
              * subj:obj:to        0.0536333367658669 *
              * subj:obj:pp.for    0.0189647478804105 *

like          * subj:to            0.344067278287462 *
              * subj:obj           0.34302752293578 *
              * subj               0.142110091743119 *
              * subj:obj:adv       0.0364220183486239 *
              * subj:obj:obj       0.0262691131498471 *

want          * subj:to            0.533195075557434 *
              * subj:obj           0.149146676529642 *
              * subj               0.110892423121632 *
              * subj:obj:to        0.102729049984149 *
              * subj:to:adv        0.0163663742999049 *

desire        * subj:obj           0.25 *
              * subj               0.244535519125683 *
              * subj:to            0.203551912568306 *
              * subj:obj:to        0.069672131147541 *
              * subj:s             0.0204918032786885 *
```

[4] It is striking that some wrong subcategorisation frames are listed, especially the intransitive frame type `subj`, which is partly due to underlying sentences containing an NP ellipsis (like in *"Our responsibilities are as follows: you invent, I commercialize."*), partly to parsing mistakes and the frame extraction.

Adding information about the selectional preferences of the verbs' arguments helps to get an idea about their semantic concepts.
The *Manner of Motion* verbs preferably appeared with a subject only, partly followed by an adverb. The subject in both frames was an inanimate object, for *move* it might also be a piece 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:

```
roll          * subj(PhysObject)                  0.241451670685337 *
              * subj(PhysObject):adv              0.104624830989344 *
              * subj(Agent):obj(PhysObject)       0.0722786755339997 *
              * subj(LifeForm):obj(PhysObject)    0.0680756190652667 *
              * subj(Agent):obj(Part)             0.0525121359227189 *

fly           * subj(PhysObject)                  0.335013432064644 *
              * subj(PhysObject):adv              0.123622741498 *
              * subj(LifeForm):obj(PhysObject)    0.0657165877759204 *
              * subj(LifeForm):pp.to(LifeForm)    0.0452314211355251 *
              * subj(LifeForm):pp.to(Agent)       0.0438113663530466 *

move          * subj(PhysObject)                  0.200321615821647 *
              * subj(PhysObject):adv              0.11363088866625 *
              * subj(Part)                        0.0925972119246233 *
              * subj(Group):adv                   0.0442911091963341 *
              * subj(Part):adv                    0.0395279510615529 *
```

The latent class analysis resulted in clusters which are presented with their probability and the verbs with the highest probabilities for the respective cluster, according to cluster membership and combination with the subcategorisation frame types in the columns. The dot indicates whether the verb-frame combination was seen in the data.

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

Cluster PROB 0.0040		0.7455	0.0857	0.0482	0.0158
		subj:obj:to	subj	subj:obj	subj:pp.on
0.1734	advise	•	•	•	•
0.1213	teach	•	•	•	•
0.1198	instruct	•	•	•	•

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

Cluster PROB 0.0208		0.2203	0.1032	0.0942	0.0863
		subj(Action)	subj(PhysObject):to	subj(LifeForm):vger	subj(Agent):vger
0.3382	start	•	•	•	•
0.1945	finish	•		•	•
0.1846	stop	•		•	•
0.1584	begin	•	•		

Both approaches show that the relationship between alternation behaviour and semantic class could already be established when only considering information about the syntactic usage of the subcategorisation frames. The refinement by the frames' selectional preferences allowed further demarcations by the identification of conceptual restrictions on the use of the frames. Since the latent class analysis was able to assign verbs to several clusters, this further distinction can be referred to as distinguishing between the different verbs' senses and the respective uses of subcategorisation frames. For example, consider the following two clusters where the verb *play* was once clustered with *meet* because of the common strong tendency towards a transitive frame illustrating a general meeting, and once it was clustered with *fight* because of their common preference for an intransitive frame together with a prepositional phrase headed by *against*, when illustrating a more aggressive meeting like a match or a fight:

Cluster PROB 0.0095		0.5545	0.0468	0.0366	0.0340
		subj:obj	subj	subj:obj:pp.with	subj:obj:pp.at
0.4947	meet	•	•	•	•
0.1954	play	•	•	•	•

Cluster PROB 0.0018		0.1829	0.1297	0.0894	0.0693
		subj:pp.against	subj:obj	subj:obj:pp.against	subj:obj:adv
0.2212	fight	•	•	•	•
0.1959	play	•	•	•	•

An extensive investigation of the linguistic reliability of the verbs' and clusters' subcategorisation frames showed that the characterising usages could actually be underlined by example sentences, 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 *Today the older pilot flies the aircraft.*

This means that the linguistic properties as modelled for the approaches agree with (a selective part of) the verbs' properties. The clusters were therefore created on a reliable linguistic basis, an important fact to ensure, since an unreliable representation would question the successful relation between alternation behaviour and semantic classes.

A strange result seemed to be the fact that the clustering of the verbs became worse with both algorithms when taking the information about the frames' selectional preferences into account.

This was due partly to the quality of the linguistic basis which has to be differentiated concerning the two informational versions: concerning version `(A)` there was little noise in the descriptions of the verbs' subcategorisation frames, as my study of linguistic reliability showed. Concerning version `(B)` the problems increased. Since the increase of noise correlated with the decrease of precision concerning the clustering success, this seemed an important factor to investigate: considering each argument slot within a subcategorisation frame on its own, the preferred conceptual classes illustrated linguistic reliable possibilities to insert arguments. But by the combination of the classes too many combinatorial possibilities had been created, so the combinations were not always possible to underlie with examples. The solution to this problem should be a different formulation of the conceptual class types, to ensure an improved token per type relation in order to avoid the data sparseness in tokens.

Both algorithms were confronted with two further problems:

- Polysemy:
  The different verb senses were hidden in the representation for one verb. That is, it was not obvious how to filter the uncertain number of senses out of the word-form. The iterative distance clustering completely failed to model verb senses; a polysemous verb was because of its opaque representation either not at all assigned to a cluster, or assigned to one cluster to which one of the verb's senses belongs. The latent class analysis was able to filter the multiple senses and assign them to distinct clusters, but tended to over-interpret.

- Low Frequency:
  Verbs which rarely appeared were difficult to cluster, since the necessary background was missing. The latent class analysis suffered from this sparse data, since those verbs were always assigned low probabilities. Distance clustering suffered even more, since – in addition to the sparse data concerning the verb's usage – also the information about the co-occurrence with subcategorisation frames was missing, so the verb's distribution contained mostly zeroes, a difficult mathematical basis.

Turning to the specific problems of the clustering algorithms, I first investigated the iterative clustering: letting each verb point to the closest verb as measured by relative entropy showed that 61%/36% in the respective versions chose a verb from the same semantic class. The conclusions from this investigation are two-fold: (i) the percentages can be considered as an upper boundary for what could have been achieved by the clustering method, since not more verbs than those pointing to a verb from the same class could be clustered correctly, so to achieve a better result other distance measures should be considered, and (ii) there is a loss of correct assignments when taking into account that – as table 1 shows – only 36%/20% of the verbs were finally found in correct clusters, which had to be caused by the merging process and the limit on the size of the clusters, so those were less than optimal and worth to be developed further.

Investigating the latent class analysis could underline that the data sparseness as mentioned before caused problems for the training process. In total there were only 6,873 verb-frame types for version `(B)` which was a too narrow basis. For version `(A)` I had 27,016 verb-frame types, but differently to `(B)` only 88 different frames, so creating 80 different clusters had the tendency to result in some classes where only one frame was favoured.

5 Conclusion

I proposed two algorithms for automatically classifying verbs semantically, based on their alternation behaviour. Taking Levin [8] as golden 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 frames' 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 a common alternation behaviour. The common properties within one cluster were refined when adding information about the selectional preferences to the syntactic description of the subcategorisation frames.

The discussion demonstrated that some problems in the classification process still have to be solved:

- An obvious problem in the clustering was the fact that the results were worse when incorporating the definition of the frames' selectional preferences. The representation of the subcategorisation frames including information about their selectional preferences should be improved to ensure a better token per type relation.
- The polysemy of verbs presented a problem, especially for the distance clustering, which could not distinguish between the multiple senses.
- Both approaches had difficulties in clustering low-frequency verbs, since the data could not be delimited in the clustering process.

Considering the overall motivation of this work, a successful step into the direction of presenting the connection between the verbs' alternation behaviour and their semantics by automatic means is done. Naturally, there are possibilities to improve the process.

A Subcategorisation Frames

Part A.1 contains a list of the 88 subcategorisation frame types which built the basis for the syntactic description of the verbs. The frames are numbered from 0 to 87. Explanations about the syntactic features within the frames can be found in part A.2. The appendix is concluded in part A.3 by the joint frequencies of the verb *give* concerning the frame types.

A.1 Frame Types

0	subj
1	subj:adv
2	subj:ap
3	subj:obj
4	subj:obj:adv
5	subj:obj:ap
6	subj:obj:as
7	subj:obj:obj
8	subj:obj:obj:adv
9	subj:obj:obj:pp.at
10	subj:obj:obj:pp.for
11	subj:obj:obj:pp.in
12	subj:obj:obj:pp.on
13	subj:obj:obj:pp.to
14	subj:obj:obj:pp.with
15	subj:obj:pp.about
16	subj:obj:pp.after
17	subj:obj:pp.against
18	subj:obj:pp.as
19	subj:obj:pp.at
20	subj:obj:pp.before
21	subj:obj:pp.between
22	subj:obj:pp.by
23	subj:obj:pp.during
24	subj:obj:pp.for
25	subj:obj:pp.from
26	subj:obj:pp.in
27	subj:obj:pp.in:adv
28	subj:obj:pp.in:pp.in
29	subj:obj:pp.into
30	subj:obj:pp.like
31	subj:obj:pp.of
32	subj:obj:pp.on
33	subj:obj:pp.out_of
34	subj:obj:pp.over

35	subj:obj:pp.through
36	subj:obj:pp.to
37	subj:obj:pp.under
38	subj:obj:pp.with
39	subj:obj:pp.within
40	subj:obj:pp.without
41	subj:obj:ppart
42	subj:obj:s
43	subj:obj:sub
44	subj:obj:that
45	subj:obj:to
46	subj:obj:vbase
47	subj:obj:vger
48	subj:pp.about
49	subj:pp.across
50	subj:pp.after
51	subj:pp.against
52	subj:pp.as
53	subj:pp.at
54	subj:pp.at:adv
55	subj:pp.between
56	subj:pp.by
57	subj:pp.for
58	subj:pp.for:adv
59	subj:pp.from
60	subj:pp.from:pp.to
61	subj:pp.in
62	subj:pp.in:adv
63	subj:pp.into
64	subj:pp.like
65	subj:pp.of
66	subj:pp.on
67	subj:pp.on:adv
68	subj:pp.out_of
69	subj:pp.over
70	subj:pp.through
71	subj:pp.to
72	subj:pp.to:adv
73	subj:pp.towards
74	subj:pp.under
75	subj:pp.up_to
76	subj:pp.upon
77	subj:pp.with
78	subj:pp.with:adv
79	subj:ppart
80	subj:s
81	subj:sub
82	subj:that
83	subj:to
84	subj:to:adv
85	subj:vbase
86	subj:vbase:adv
87	subj:vger

A.2 Frame Features

Syntactic features of the frame types, as defined by the English grammar:

adv	adverb
ap	adjectival phrase
as	as-expression
pp	prepositional phrase
ppart	stranded preposition
s	sentence
that	subordinated that-phrase
to	infinitive form of verb after 'to'
vbase	base form of verb
vger	gerund

and additional identifiers:

subj	subject of the sentence
obj	object of the sentence

A.3 Joint Frequencies of the Verb *give* concerning the Frame Types

The following list displays the joint frequencies of the verb *give* concerning the frame types in column two. For frame types defined in appendix A.1 which do not appear here the joint frequency was zero.

give	subj	758
give	subj:adv	105
give	subj:ap	58
give	subj:obj	9,982
give	subj:obj:adv	498
give	subj:obj:ap	60
give	subj:obj:as	53
give	subj:obj:obj	13,430
give	subj:obj:obj:adv	158
give	subj:obj:obj:pp.at	59
give	subj:obj:obj:pp.for	144
give	subj:obj:obj:pp.in	238
give	subj:obj:obj:pp.on	68
give	subj:obj:obj:pp.to	240

give	subj:obj:obj:pp.with	39
give	subj:obj:pp.about	57
give	subj:obj:pp.after	42
give	subj:obj:pp.against	14
give	subj:obj:pp.as	171
give	subj:obj:pp.at	220
give	subj:obj:pp.before	24
give	subj:obj:pp.between	5
give	subj:obj:pp.by	40
give	subj:obj:pp.during	30
give	subj:obj:pp.for	566
give	subj:obj:pp.from	56
give	subj:obj:pp.in	936
give	subj:obj:pp.in:adv	16
give	subj:obj:pp.in:pp.in	11
give	subj:obj:pp.into	17
give	subj:obj:pp.like	8
give	subj:obj:pp.of	198
give	subj:obj:pp.on	234
give	subj:obj:pp.out_of	16
give	subj:obj:pp.over	35
give	subj:obj:pp.through	15
give	subj:obj:pp.to	3,735
give	subj:obj:pp.under	26
give	subj:obj:pp.with	103
give	subj:obj:pp.within	15
give	subj:obj:pp.without	36
give	subj:obj:ppart	98
give	subj:obj:s	35
give	subj:obj:sub	16
give	subj:obj:that	67
give	subj:obj:to	277
give	subj:obj:vbase	15
give	subj:obj:vger	35
give	subj:pp.about	4
give	subj:pp.across	3
give	subj:pp.after	5
give	subj:pp.against	1
give	subj:pp.as	10
give	subj:pp.at	17
give	subj:pp.at:adv	3
give	subj:pp.between	1
give	subj:pp.by	2
give	subj:pp.for	34
give	subj:pp.for:adv	6
give	subj:pp.from	5
give	subj:pp.from:pp.to	1
give	subj:pp.in	50
give	subj:pp.into	9
give	subj:pp.of	31
give	subj:pp.on	14
give	subj:pp.out_of	6
give	subj:pp.over	1
give	subj:pp.through	2
give	subj:pp.to	288
give	subj:pp.to:adv	17
give	subj:pp.towards	2
give	subj:pp.under	3
give	subj:pp.up_to	6
give	subj:pp.upon	3
give	subj:pp.with	14
give	subj:pp.with:adv	1
give	subj:ppart	6
give	subj:s	280
give	subj:sub	1
give	subj:that	18
give	subj:to	38
give	subj:vbase	36
give	subj:vbase:adv	1
give	subj:vger	15

B WordNet (Top) Synsets

There are 11 top level nodes of 11 hierarchies in WordNet. Since the concept of `Entity` seemed too general as conceptual class, I replaced it by the next lower levels (13 different synsets). Each WordNet synset is defined by an identifying abbreviation, followed by the nouns which are member of that synset:

```
Entity:      entity
             => LifeForm:      life form, organism, being, living thing
             => Cell:          cell
             => Agent:         causal agent, cause, causal agency
             => PhysObject:    object, inanimate object, physical object
             => Thing:         thing
             => Whole:         whole, whole thing, unit
             => Content:       subject, content, depicted object
             => Unit:          unit, building block
             => Part:          part, piece
             => Essential:     necessity, essential, requirement,
                               requisite, necessary, need
             => Inessential:   inessential
             => Variable:      variable
             => Anticipation:  anticipation
Psycho:      psychological_feature
Abstract:    abstraction
Location:    location
Shape:       shape, form
State:       state
Event:       event
Action:      act, human_action, human_activity
Group:       group, grouping
Possession:  possession
Phenomenon:  phenomenon
```

References

[1] Richard Beckwith, Christiane Fellbaum, Derek Gross, and George A. Miller. 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, 1991.

[2] Glenn Carroll and Mats Rooth. Valence Induction with a Head-Lexicalized PCFG. In *Proceedings of the 3rd Conference on Empirical Methods in Natural Language Processing*, Granada, Spain, 1998.

[3] Noam Chomsky. *Aspects of the Theory of Syntax.* MIT Press, Cambridge, MA, 1965.

[4] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum Likelihood from Incomplete Data via the EM algorithm. *Journal of the Royal Statistical Society*, 39(B):1–38, 1977.

[5] John Hughes. *Automatically Acquiring Classification of Words.* PhD thesis, University of Leeds, School of Computer Studies, 1994.

[6] Judith L. Klavans and Min-Yen Kan. The Role of Verbs in Document Analysis. In *Proceedings of the 17th International Conference on Computational Linguistics (COLING-ACL '98)*, Montreal, Canada, August 1998.

[7] S. Kullback and R. A. Leibler. On Information and Sufficiency. *Annals of Mathematical Statistics*, 22:79–86, 1951.

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

[9] Philip Resnik. *Selection and Information: A Class-Based Approach to Lexical Relationships.* PhD thesis, University of Pennsylvania, 1993.

[10] Philip Resnik. Selectional Preference and Sense Disambiguation. In *Proceedings of the ACL SIGLEX Workshop on Tagging Text with Lexical Semantics: Why, What, and How?*, 1997.

[11] Francesc Ribas. An Experiment on Learning Appropriate Selectional Restrictions from a Parsed Corpus. In *Proceedings of the 15th International Conference on Computational Linguistics (COLING-94)*, pages 769–774, 1994.

[12] Francesc Ribas. On Learning More Appropriate Selectional Restrictions. In *Proceedings of the 7th Conference of the European Chapter of the Association for Computational Linguistics*, Dublin, Ireland, 1995.

[13] Mats Rooth. Two-Dimensional Clusters in Grammatical Relations. In *Inducing Lexicons with the EM Algorithm*, AIMS Report 4(3). Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart, 1998.

[14] Sabine Schulte im Walde. Automatic Semantic Classification of Verbs According to Their Alternation Behaviour. Master's thesis, Universität Stuttgart, Institut für Maschinelle Sprachverarbeitung, 1998.