

Experiments on the Choice of Features for Learning Verb Classes

Sabine Schulte im Walde

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

Abstract

The choice of verb features is crucial for the learning of verb classes. This paper presents clustering experiments on 168 German verbs, which explore the relevance of features on three levels of verb description, purely syntactic frame types, prepositional phrase information and selectional preferences. In contrast to previous approaches concentrating on the sparse data problem, we present evidence for a linguistically defined limit on the usefulness of features which is driven by the idiosyncratic properties of the verbs and the specific attributes of the desired verb classification.

1 Introduction

The verb is central to the meaning and the structure of a sentence, and lexical verb information represents the core in supporting NLP-tasks such as word sense disambiguation (Dorr and Jones, 1996; Prescher et al., 2000), machine translation (Dorr, 1997), document classification (Klavans and Kan, 1998), and subcategorisation acquisition and filtering (Korhonen, 2002). A means to generalise over and predict common properties of verbs is captured by the constitution of verb classes. Levin (1993) has established an extensive manual classification for English verbs; computational approaches adopt the linguistic hypothesis that *verb meaning components to a certain extent determine verb behaviour* as basis for automatically inducing semantic verb classes from corpus-based features (Schulte im Walde, 2000; Merlo and Stevenson, 2001; Joanis, 2002).

Computational approaches on verb classification which take advantage of corpus-based and knowledge-based verb information offered by available tools and resources such as statistical parsers and semantic ontologies, suffer from severe problems to encode and benefit from the information, especially with respect to selectional preferences, cf. Schulte im Walde (2000); Joanis (2002). This paper presents clustering experiments on German verbs which explore the relevance of features on three levels of verb description, purely syntactic frame types, prepositional phrase information and selectional preferences. The clustering results show that the choice and implementation of verb features is crucial for the induction of the verb classes. Intuitively, one might want to add and refine features ad infinitum, but we present evidence for a linguistically defined limit on the usefulness of features which is driven by the idiosyncratic properties of the verbs and the verb classification.

2 German Verb Classes

A set of 168 German verbs is manually classified into 43 concise semantic verb classes. The purpose of the manual classification is (i) to evaluate the reliability and performance of the clustering experiments on a preliminary set of verbs, and (ii) to explore the potential and limit to apply the clustering method to large-scale verb data. The German classes are closely related to the English pendant in (Levin, 1993) and agree with the German verb classification in (Schumacher, 1986) as far as the relevant verbs appear in his semantic ‘fields’.

Table 1 presents the manual verb classification. The class size is between 2 and 7, with an average of 3.9 verbs per class. Eight verbs are am-

biguous and marked by subscripts. The classes include both high and low frequency verbs,¹ in order to exercise the clustering technology in both data-rich and data-poor situations. The class labels are given on two semantic levels; coarse labels such as *Manner of Motion* are sub-divided into finer labels, such as *Locomotion*, *Rotation*. The fine labels are relevant for the clustering experiments, as indicated by the numbering in the left column.

The classification is primarily based on semantic intuition, not on knowledge about the syntactic behaviour. As an extreme example, the *Support* class (34) contains the verb *unterstützen*, which syntactically requires a direct object, together with the three verbs *dienen*, *folgen*, *helfen* which mainly subcategorise an indirect object.

3 Clustering Methodology

Clustering is a standard procedure in multivariate data analysis. It is designed to uncover an inherent natural structure of data objects, and the induced equivalence classes provide a means to generalise over the objects. We perform clustering by the k-Means algorithm (Forgy, 1965), an unsupervised hard clustering method assigning n data objects to k clusters.² Initial verb clusters are iteratively re-organised by assigning each verb to its closest cluster and re-calculating cluster centroids until no further changes take place. The clustering methodology in this work is based on parameter investigations in (Schulte im Walde and Brew, 2002): the *clustering input* is obtained from a hierarchical analysis on the German verbs (Ward's amalgamation method), the *number of clusters* being the number of manual classes; *similarity measure* is performed by the *skew divergence*, a variant of the Kullback-Leibler divergence.

The 168 verbs are associated with probabilistic frame descriptions on various levels of verb information, and assigned to starting clusters by hierarchical clustering. The k-Means algorithm is then allowed to run until no further changes take place, and the resulting clusters are evaluated and interpreted against the manual classes.

¹The verb frequency range in 35 million words newspaper data is 8–71,604.

²Hard clustering is an oversimplification for representing ambiguous verbs, but it facilitates interpretation.

(1)	<i>Aspect</i> : anfangen, aufhören, beenden, beginnen, enden
(2)	<i>Propositional Attitude</i> : ahnen, denken, glauben, vermuten, wissen
	<i>Desire</i> :
(3)	<i>Wish</i> : erhoffen, wollen, wünschen
(4)	<i>Need</i> : bedürfen, benötigen, brauchen
(5)	<i>Transfer of Possession (Obtaining)</i> : bekommen, erhalten, erlangen, kriegen
	<i>Transfer of Possession (Giving)</i> :
(6)	<i>Gift</i> : geben, leihen, schenken, spenden, stiften, vermachen, überschreiben
(7)	<i>Supply</i> : bringen, liefern, schicken, vermitteln ₁ , zustellen
	<i>Manner of Motion</i> :
(8)	<i>Locomotion</i> : gehen, klettern, kriechen, laufen, rennen, schleichen, wandern
(9)	<i>Rotation</i> : drehen, rotieren
(10)	<i>Rush</i> : eilen, hasten
(11)	<i>Means</i> : fahren, fliegen, rudern, segeln
(12)	<i>Flotation</i> : fließen, gleiten, treiben
	<i>Emotion</i> :
(13)	<i>Origin</i> : ärgern, freuen
(14)	<i>Expression</i> : heulen ₁ , lachen ₁ , weinen
(15)	<i>Objection</i> : ängstigen, ekeln, fürchten, scheuen
(16)	<i>Face Look</i> : gähnen, grinsen, lachen ₂ , lächeln, starren
(17)	<i>Perception</i> : empfinden, erfahren ₁ , fühlen, hören, riechen, sehen, wahrnehmen
(18)	<i>Manner of Articulation</i> : flüstern, rufen, schreien
(19)	<i>Moaning</i> : heulen ₂ , jammern, klagen, lamentieren
(20)	<i>Communication</i> : kommunizieren, korrespondieren, reden, sprechen, verhandeln
	<i>Statement</i> :
(21)	<i>Announcement</i> : ankündigen, bekanntgeben, eröffnen, verkünden
(22)	<i>Constitution</i> : anordnen, bestimmen, festlegen
(23)	<i>Promise</i> : versichern, versprechen, zusagen
(24)	<i>Observation</i> : bemerken, erkennen, erfahren ₂ , feststellen, realisieren, registrieren
(25)	<i>Description</i> : beschreiben, charakterisieren, darstellen ₁ , interpretieren
(26)	<i>Presentation</i> : darstellen ₂ , demonstrieren, präsentieren, veranschaulichen, vorführen
(27)	<i>Speculation</i> : grübeln, nachdenken, phantasieren, spekulieren
(28)	<i>Insistence</i> : beharren, bestehen ₁ , insistieren, pochen
(29)	<i>Teaching</i> : beibringen, lehren, unterrichten, vermitteln ₂
	<i>Position</i> :
(30)	<i>Bring into Position</i> : legen, setzen, stellen
(31)	<i>Be in Position</i> : liegen, sitzen, stehen
(32)	<i>Production</i> : bilden, erzeugen, herstellen, hervorbringen, produzieren
(33)	<i>Renovation</i> : dekorieren, erneuern, renovieren, reparieren
(34)	<i>Support</i> : dienen, folgen ₁ , helfen, unterstützen
(35)	<i>Quantum Change</i> : erhöhen, erniedrigen, senken, steigern, vergrößern, verkleinern
(36)	<i>Opening</i> : öffnen, schließen ₁
(37)	<i>Existence</i> : bestehen ₂ , existieren, leben
(38)	<i>Consumption</i> : essen, konsumieren, lesen, saufen, trinken
(39)	<i>Elimination</i> : eliminieren, entfernen, exekutieren, töten, vernichten
(40)	<i>Basis</i> : basieren, beruhen, gründen, stützen
(41)	<i>Inference</i> : folgern, schließen ₂
(42)	<i>Result</i> : ergeben, erwachsen, folgen ₂ , resultieren
(43)	<i>Weather</i> : blitzen, donnern, dämmern, nieseln, regnen, schneien

Table 1: German semantic verb classes

4 Clustering Evaluation

Evaluating the result of a cluster analysis against the known gold standard of hand-constructed verb classes requires to assess the similarity between two partitions on the set of n verbs. The evaluation is performed by an adjusted version of the Rand index (Hubert and Arabie, 1985): The Rand index measures the agreement between object pairs in the partitions and is corrected for chance in comparison to the null model that the partitions are picked at random, given the original number of classes and objects.

The agreement in the two partitions is represented by a contingency table $C \times M$: t_{ij} denotes the number of verbs common to classes C_i in the clustering partition C and M_j in the manual classification M ; the marginals $t_{i.}$ and $t_{.j}$ refer to the number of objects in C_i and M_j , respectively. The adjusted Rand index R_{adj} is given in Equation (1); the expected number of common object pairs attributable to a particular cell $\langle C_i, M_j \rangle$ in the contingency table is defined by $\binom{t_{i.}}{2} \binom{t_{.j}}{2} / \binom{n}{2}$. The range of R_{adj} is $0 \leq R_{adj} \leq 1$, with only extreme cases below zero. We choose R_{adj} as evaluation measure compared to e.g. the measures presented in (Schulte im Walde and Brew, 2002), because (a) it does not show a bias towards extreme cluster sizes, and (b) it facilitates the interpretation with its normally used bounds of 0 and 1.

$$R_{adj} = \frac{\sum_{i,j} \binom{t_{ij}}{2} - \frac{\sum_i \binom{t_{i.}}{2} \sum_j \binom{t_{.j}}{2}}{\binom{n}{2}}}{\frac{1}{2} (\sum_i \binom{t_{i.}}{2} + \sum_j \binom{t_{.j}}{2}) - \frac{\sum_i \binom{t_{i.}}{2} \sum_j \binom{t_{.j}}{2}}{\binom{n}{2}}} \quad (1)$$

5 Verb Description

The German verbs are described on three levels of subcategorisation definition $D1$ to $D3$, each refining the previous level by additional information. All information is extracted from a lexicalised probabilistic grammar which is unsupervised trained on 35 million words of a German newspaper corpus, using the EM-algorithm.

$D1$ provides a **coarse syntactic definition of subcategorisation**. The verbs are described by a probability distribution over 38 frame types. 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 next to the obligatory nominative subject is labelled ‘nai’.

On $D2$, the verbs are given a **syntactico-semantic definition of subcategorisation with prepositional preferences**. In addition to the syntactic frame information, $D2$ discriminates between different kinds of pp-arguments. This is done by distributing the probability mass of prepositional phrase frame types over the prepositional phrases, according to their frequencies in the corpus. Prepositional phrases are referred to by case and preposition, such as ‘mit_D’, ‘für_A’, with D=Dative and A=Accusative. We define 30 different PPs, according to the most frequent PPs which appear with at least 10 different verbs.

$D3$ gives a **syntactico-semantic definition of subcategorisation with prepositional and selectional preferences**. The argument slots within a subcategorisation frame type are specified according to which ‘kind’ of argument they require. The grammar provides selectional preference information on a fine-grained level: it specifies argument realisations for a specific verb-frame-slot combination in form of lexical heads. For example, the most prominent nominal argument heads for the verb *verfolgen* ‘to follow’ in the accusative NP slot of the transitive frame type ‘na’ (the considered frame slot is underlined) are *Ziel* ‘goal’, *Strategie* ‘strategy’, *Politik* ‘policy’. Obviously, we would run into a sparse data problem if we tried to incorporate selectional preferences on the nominal level into the verb descriptions. We need a generalisation of the selectional preference definition, for which we use the noun hierarchy in *GermaNet* (Kunze, 2000), the German pendant of the semantic ontology *WordNet* (Fellbaum, 1998).

The hierarchy is realised as synsets, sets of synonymous nouns, which are organised into multiple inheritance hypernym relationships. A noun may appear in several synsets, according to its number of senses. For each nominal argument in a verb-

frame-slot combination, the joint frequency is split over the different senses of the noun and propagated upwards the hierarchy. In case of multiple hypernym synsets, the frequency is split, such that the sum of frequencies over the disjoint top synsets equals the total joint frequency. Repeating the frequency assignment and propagation for all nouns appearing in a verb-frame-slot combination, we define a frequency distribution of the verb-frame-slot combination over all GermaNet synsets. To restrict the variety of noun concepts, we consider only the 15 top GermaNet nodes: *Lebewesen* ‘creature’, *Sache* ‘thing’, *Besitz* ‘property’, *Substanz* ‘substance’, *Nahrung* ‘food’, *Mittel* ‘means’, *Situation* ‘situation’, *Zustand* ‘state’, *Struktur* ‘structure’, *Physis* ‘body’, *Zeit* ‘time’, *Ort* ‘space’, *Attribut* ‘attribute’, *Kognitives Objekt* ‘cognitive object’, *Kognitiver Prozess* ‘cognitive process’.³ Since the 15 nodes exclude each other and the frequencies sum to the total joint verb-frame frequency, we can define a probability distribution over the top nodes representing coarse selectional preferences for the respective verb-frame-slot combination. To obtain *D3*, the verb-frame probability is distributed over those selectional preferences.⁴

Table 2 presents three verbs from different verb classes and their ten most frequent frame types with respect to the three levels of verb definition, accompanied by the probability values. *D1* for *beginnen* ‘to begin’ defines ‘np’ and ‘n’ as the most probable frame types. Even by splitting the ‘np’ probability over the different PP types in *D2*, a number of prominent PPs are left, the time indicating *um_A* and *nach_D*, *mit_D* referring to the begun event, *an_D* as date and *in_D* as place indicator. It is obvious that not all PPs are argument PPs, but also adjunct PPs describe a part of the verb behaviour. *D3* illustrates that typical selectional preferences for beginner roles are *Situation*, *Zustand*, *Zeit*, *Sache*. *D3* has the potential to indicate verb alternation behaviour, e.g. ‘*na*(Situation)’ refers to the same role for the direct object in a

transitive frame as ‘*n*(Situation)’ in an intransitive frame.

essen ‘to eat’ as an object drop verb shows strong preferences for both an intransitive and transitive usage. As desired, the argument roles are strongly determined by *Lebewesen* for both ‘*n*’ and ‘*na*’ and *Nahrung* for ‘*na*’.

fahren ‘to drive’ chooses typical manner of motion frames (‘*n*’, ‘*np*’, ‘*na*’) with the refining PPs being directional (*in_A*, *zu_D*, *nach_D*) or referring to a means of motion (*mit_D*, *in_D*, *auf_D*). The selectional preferences represent a correct alternation behaviour: *Lebewesen* in the object drop case for ‘*n*’ and ‘*na*’, *Sache* in the inchoative/causative case for ‘*n*’ and ‘*na*’.

<i>D1</i>		<i>D2</i>		<i>D3</i>	
<i>beginnen</i> ‘to begin’					
np	0.43	n	0.28	<i>n</i> (Situation)	0.12
n	0.28	np:um _A	0.16	np:um _A (Situation)	0.09
ni	0.09	ni	0.09	np:mit _D (Situation)	0.04
na	0.07	np:mit _D	0.08	<i>ni</i> (Lebewesen)	0.03
nd	0.04	na	0.07	<i>n</i> (Zustand)	0.03
nap	0.03	np:an _D	0.06	np:an _D (Situation)	0.03
nad	0.03	np:in _D	0.06	np:in _D (Situation)	0.03
nir	0.01	nd	0.04	<i>n</i> (Zeit)	0.03
ns-2	0.01	nad	0.02	<i>n</i> (Sache)	0.02
xp	0.01	np:nach _D	0.01	<i>na</i> (Situation)	0.02
<i>essen</i> ‘to eat’					
na	0.42	na	0.42	<i>na</i> (Lebewesen)	0.33
n	0.26	n	0.26	<i>na</i> (Nahrung)	0.17
nad	0.10	nad	0.10	<i>na</i> (Sache)	0.09
np	0.06	nd	0.05	<i>n</i> (Lebewesen)	0.08
nd	0.05	ns-2	0.02	<i>na</i> (Lebewesen)	0.07
nap	0.04	np:auf _D	0.02	<i>n</i> (Nahrung)	0.06
ns-2	0.02	ns-w	0.01	<i>n</i> (Sache)	0.04
ns-w	0.01	ni	0.01	<i>nd</i> (Lebewesen)	0.04
ni	0.01	np:mit _D	0.01	<i>nd</i> (Nahrung)	0.02
nas-2	0.01	np:in _D	0.01	<i>na</i> (Attribut)	0.02
<i>fahren</i> ‘to drive’					
n	0.34	n	0.34	<i>n</i> (Sache)	0.12
np	0.29	na	0.19	<i>n</i> (Lebewesen)	0.10
na	0.19	np:in _A	0.05	<i>na</i> (Lebewesen)	0.08
nap	0.06	nad	0.04	<i>na</i> (Sache)	0.06
nad	0.04	np:zu _D	0.04	<i>n</i> (Ort)	0.06
nd	0.04	nd	0.04	<i>na</i> (Sache)	0.05
ni	0.01	np:nach _D	0.04	np:in _A (Sache)	0.02
ns-2	0.01	np:mit _D	0.03	np:zu _D (Sache)	0.02
ndp	0.01	np:in _D	0.03	np:in _A (Lebewesen)	0.02
ns-w	0.01	np:auf _D	0.02	np:nach _D (Sache)	0.02

Table 2: Examples of most probable frame types

6 Feature Variation

The previous section introduced the verb description in an ‘as is’ fashion, but obviously one can

³Little manual intervention was necessary to define a coherent set of top level nodes, since GermaNet had not been completed.

⁴Strictly speaking, we do not have a probability distribution any longer, since multiple frame slots may be refined. The skew divergence still works well.

find multiple variations. In order to illustrate that the most plausible variations have been considered, we describe and use linguistically intuitive mutations of the verb descriptions.⁵

- On *D1*, there is little room to vary the verb information, since the valency encoding is close to standard German grammar, cf. Helbig and Buscha (1998).

- On *D2*, we vary the amount of PP information: (a) Following standard German grammar books we define a more restricted set of prepositional phrases for argument usage, and (b) ignoring any frequency constraint on the PP information increases the kinds of PPs in the relevant frame types up to 140.

- On *D3*, there is most room for variation:

Role Choice: Instead of using the 15 top level nodes in GermaNet, (a) we use selectional preferences on a more fine-grained level, the word level, and (b) we define a more generalised description of selectional preferences, by merging the frequencies of the 15 top level nodes in GermaNet to only 2 (Lebewesen, Objekt) or 3 (Lebewesen, Sache, Abstraktum).

Role Integration: To integrate the selectional preferences into the verb description, either (a) each argument slot in a subcategorisation frame is substituted by selectional roles separately, e.g. the joint frequency of a verb and transitive ‘na’ is distributed over the nominative slot preferences ‘na(Lebewesen)’, ‘na(Sache)’, etc. and also over the accusative slot preferences ‘na(Lebewesen)’, ‘na(Sache)’, etc. (as in Table 2). In this case, the argument slots of frame types with several arguments are considered independently, but the number of features remains in a reasonable magnitude, 15 per frame slot. Or (b) the subcategorisation frames are substituted by the combinations of selectional preferences for the argument slots, e.g. the joint probability of a verb and ‘na’ is distributed over ‘na(Lebewesen:Nahrung)’, ‘na(Lebewesen:Sache)’, ‘na(Sache:Nahrung)’, etc. This encoding would directly represent the linguistic idea of alternations, but no direct frequencies are available, and the number of features explodes (15 features for an intransitive,

⁵We do not attempt to optimise the feature set algorithmically, because that would lead to overfitting.

15² for a transitive, 15³ for a ditransitive) and leads to differing magnitudes of probabilities.

Role Means: We could use a different means for selectional role representation than GermaNet. But since the ontological idea of WordNet has been widely and successfully used and we do not have any comparable source at hand, we have to exclude this variation.

7 Clustering Results

The **baseline** for the clustering experiments is $R_{adj} = -0.004$ and refers to 50 random clusterings: The verbs are randomly assigned to a cluster (with a cluster number between 1 and the number of manual classes 43), and the resulting clustering is evaluated. The baseline value is the average value of the 50 repetitions. The **upper bound** is $R_{adj} = 0.909$ and calculated on a hard version of the manual classification, i.e. multiple senses of verbs are reduced to a single class affiliation, which represents the optimum for the hard clustering algorithm.

Table 3 presents the clustering results for *D1* and *D2*, with *D2* distinguishing the amount of PP information (*arg* for arguments only, *chosen* for the manually defined PPs, *all* for all possible PPs). As stated by Schulte im Walde and Brew (2002), refining the syntactic verb information by prepositional phrases is helpful for the clustering; and the usage is not restricted to argument PPs, but extended by the more variable PP information.

Distribution		R_{adj}
<i>D1</i>		0.094
<i>D2</i>	pp _{arg}	0.151
	pp _{chosen}	0.151
	pp _{all}	0.160

Table 3: Clustering results on *D1* and *D2*

Underlying the results in Table 4, the argument roles for selectional preference information in *D3* are varied. The left part presents the results when refining only a single argument within a single frame, in addition to *D2*. Obviously, the results do not match linguistic intuition. For example, we would expect the arguments in the two highly frequent intransitive ‘n’ and transitive ‘na’ to provide valuable information with respect to their selectional preferences, but only those in ‘na’ improve

D2. On the other hand, ‘ni’ which is not expected to provide variable definitions of selectional preferences for the nominative slot, does work better than ‘n’. The right part in Table 4 illustrates the clustering results for example combinations of argument slots refined by selectional preferences, e.g. $\underline{n}/\underline{na}$ means that the nominative slot in ‘n’, and both the nominative and accusative slot in ‘na’ are refined by selectional preferences. The combined information does not necessarily improve the single slot clustering results, e.g. $\underline{n}/\underline{na}$ achieves results below the ones for refining only \underline{na} or \underline{n} . The overall best result (including non-illustrated experiment results) is achieved by defining selectional preferences on $\underline{n}/\underline{na}/\underline{nd}/\underline{nad}/\underline{ns-dass}$, better than refining all NP slots or all NP and all PP slots in the frame types. Summarising, Table 4 illustrates that a linguistic choice of features is worthwhile, but linguistic intuition and algorithmic clustering results do not necessarily align. On selected argument roles, the selectional preference information in *D3* once more improves the clustering results compared to *D2*, but the improvement is not as persuasive as *D2* improving *D1*.

Single Slots		Slot Combinations	
\underline{n}	0.125	\underline{na}	0.137
\underline{na}	0.176	$\underline{n}/\underline{na}$	0.128
\underline{n}	0.164	\underline{nad}	0.088
\underline{nad}	0.144	$\underline{n}/\underline{na}/\underline{nad}$	0.118
\underline{nad}	0.115	\underline{nd}	0.150
\underline{nad}	0.161	$\underline{n}/\underline{na}/\underline{nd}$	0.124
\underline{nd}	0.152	$\underline{n}/\underline{na}/\underline{nad}/\underline{nd}$	0.161
\underline{nd}	0.143	$\underline{n}/\underline{na}/\underline{nd}/\underline{nad}/\underline{ns-dass}$	0.182
\underline{np}	0.133	$\underline{np}/\underline{ni}/\underline{nr}/\underline{ns-2}/\underline{ns-dass}$	0.131
\underline{ni}	0.148	all NP	0.158
\underline{nr}	0.136	all NPs+PPs	0.176
$\underline{ns-2}$	0.121		
$\underline{ns-dass}$	0.156		

Table 4: Clustering results on varying *D3*

With respect to further feature variation, merging the frequencies of the 15 top level nodes in GermaNet to 2 or 3 roles results in noisy distributions and destroys the coherence of the cluster analyses. Experiment setups which either include a nominal level of selectional preference information or an alternation-like combination of selectional roles were tried, but they suffer from their time demands and result in far worse analyses.

Finally, we present representative parts of the

cluster analysis based on *D3*, with selectional roles ‘ \underline{n} ’, ‘ \underline{na} ’, ‘ \underline{nd} ’, ‘ \underline{nad} ’, ‘ $\underline{ns-dass}$ ’, and compares the respective clusters with their pendants under *D1* and *D2*. The manual class numbers as defined in Table 1 are given as subscripts.

- (a) beginnen₁ bestehen₃₇ enden₁ existieren₃₇
laufen₈ liegen₃₁ sitzen₃₁ stehen₃₁
- (b) eilen₁₀ gleiten₁₂ kriechen₈ rennen₈ starren₁₆
- (c) fahren₁₁ fliegen₁₁ fließen₁₂ klettern₈ segeln₁₁
wandern₈
- (d) bilden₃₂ erhöhen₃₅ festlegen₂₂ senken₃₅
steigern₃₅ vergrößern₃₅ verkleinern₃₅
- (e) töten₃₉ unterrichten₂₉
- (f) nieseln₄₃ regnen₄₃ schneien₄₃
- (g) dämmern₄₃

The weather verbs in cluster (f) strongly agree in their syntactic expression on *D1* and do not need *D2* or *D3* refinements for a successful class constitution. *dämmern* in cluster (g) is ambiguous between a weather verb and expressing a sense of understanding; this ambiguity is idiosyncratically expressed in *D1* frames already, so *dämmern* is never clustered together with the other weather verbs on *D1* – 3.

Manner of Motion, *Existence*, *Position* and *Aspect* verbs are similar in their syntactic frame usage and therefore merged together on *D1*, but adding PP information distinguishes the respective verb classes: *Manner of Motion* verbs primarily demand directional PPs, *Aspect* verbs are distinguished by patient *mit_D* and time and location prepositions, and *Existence* and *Position* verbs are distinguished by locative prepositions, with *Position* verbs showing more PP variation. The PP information is essential for successfully distinguishing these verb classes, and the coherence is partly destroyed by *D3*: *Manner of Motion* verbs (from the sub-classes 8-12) are captured well by clusters (b) and (c), since they inhibit strong common alternations, but cluster (a) merges the *Existence*, *Position* and *Aspect* verbs, since verb-idiosyncratic demands on selectional roles destroy the *D2* class demarcation. Admittedly, the verbs in cluster (a) are close in their semantics, with a common sense of (bringing into vs. being in) existence. Schumacher (1986) actually classifies most of the verbs into one existence class. *laufen* fits into the cluster with its sense of ‘to function’.

Cluster (d) contains most verbs of *Quantum Change*, together with one verb of *Production* and *Constitution* each. The semantics of the cluster is therefore rather pure. The verbs in the cluster typically subcategorise a direct object, alternating with a reflexive usage, ‘nr’ and ‘npr’ with mostly *auf_A* and *um_A*. The selectional preferences help to distinguish this cluster: the verbs agree in demanding a thing or situation as subject, and various objects such as attribute, cognitive object, state, structure or thing as object. Without selectional preferences (on *D1* and *D2*), the change of quantum verbs are not found together with the same degree of purity.

There are verbs as in cluster (e), whose properties are correctly stated as similar on *D1* – 3, so a common cluster is justified; but the verbs only have coarse common meaning components, in this case *töten* and *unterrichten* agree in an action of one person or institution towards another.

Summarising the cluster description, some verbs and verb classes are distinctive on a coarse feature level, some need fine-grained extensions, and some are not distinctive with respect to any combination of features.

8 Discussion and Conclusion

We have presented a clustering methodology for German verbs whose results agree with a manual classification in many respects and should prove useful as automatic basis for a large-scale clustering. Without any doubt the cluster analysis would need manual correction and completion, but represents a plausible basis.

The various verb descriptions illustrate that step-wise refining the features does improve the clustering. But the linguistic feature refinements not necessarily align with expected changes in clustering. This effect could be due to (i) noisy or (ii) sparse data, but (i) the example distributions in Table 2 demonstrate that –even if noisy– our basic verb descriptions appear reliable with respect to their desired linguistic content. In addition, the subcategorisation information on *D1* and *D2* has been evaluated against manual definitions in a dictionary and proven useful (Schulte im Walde, 2002). And (ii) Table 4 illustrates that even with adding little information (e.g. refining a single argument by 15 selectional roles results

in 186 instead of 171 features) linguistic intuition and clustering results do not necessarily align.

Related work on automatic verb classes confirms the difficulty of selecting and encoding verb features. Schulte im Walde (2000) clusters 153 English verbs into 30 verb classes as taken from (Levin, 1993), using a hierarchical clustering method. The clustering is most successful when utilising syntactic subcategorisation frames enriched with PP information (comparable to our *D2*); selectional preferences are encoded by role combinations taken from WordNet. Schulte im Walde claims the detailed encoding and therefore sparse data to make the clustering worse with than without the selectional preference information. Merlo and Stevenson (2001) classify a smaller number of 60 English verbs into three verb classes, by utilising supervised decision trees. The features of the verbs are restricted to those which should capture the basic differences between the verb classes, and the feature values are approached by corpus-based heuristics (e.g. measuring the degree of animacy by personal pronoun realisation in the transitive subject slot). An extension of their work by Joanis (2002) uses 802 verbs from 14 classes in (Levin, 1993). He defines an extensive feature space with 219 core features (such as part of speech, auxiliary frequency, syntactic categories, animacy as above) and 1,140 selectional preference features taken from WordNet. As in our approach, the selectional preferences do not improve the clustering.

Why do we encounter such unpredictability concerning the encoding and effect of verb features, especially with respect to selectional preferences? In contrast to previous approaches concentrating on the sparse data problem, we have presented evidence for a linguistically defined limit on the usefulness of the verb features, driven by the **idiosyncratic properties of the verbs**. Recall the underlying idea of verb classes, that the meaning components of verbs to a certain extent determine their behaviour. This does not mean that all properties of all verbs in a common class are similar and we could extend and refine the feature description endlessly, still improving the clustering. The meaning of verbs comprises both (i) properties which are general for the respective verb

classes, and (ii) idiosyncratic properties which distinguish the verbs from each other. As long as we define the verbs by those properties which represent the common parts of the verb classes, a clustering can succeed. But with step-wise refining the verb description by including lexical idiosyncrasy, the emphasis of the common properties vanishes.

The exemplary description of cluster outcomes in the previous section confirms that it is impossible to determine an overall appropriate level of feature specification which suffices all kinds of verb classes defined in Table 1. Some verbs and verb classes are distinctive on a coarse feature level, some need fine-grained extensions, some are not distinctive with respect to any combination of features. There is no unique perfect choice and encoding of the verb features, even more with respect to a potential large-scale extension of verbs and classes. Further work on the verb classes should concern a choice of verb features with respect to the **specific properties of the desired verb classification**. We could think of either (i) performing several cluster analyses on the same set of verbs, but with different choices of verb features, and then find a way to merge the results to a unique classification, or (ii) not aiming for a fine-grained clustering, but create fewer but larger clusters on coarse features, which classify the verbs on a more general level. Both solutions should facilitate the demarcation of common and idiosyncratic verb features and improve the clustering results.

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*, pages 322–327, Copenhagen, Denmark.
- Bonnie J. Dorr. 1997. Large-Scale Dictionary Construction for Foreign Language Tutoring and Interlingual Machine Translation. *Machine Translation*, 12(4):271–322.
- Christiane Fellbaum, editor. 1998. *WordNet – An Electronic Lexical Database*. Language, Speech, and Communication. MIT Press, Cambridge, MA.
- 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.
- Lawrence Hubert and Phipps Arabie. 1985. Comparing Partitions. *Journal of Classification*, 2:193–218.
- Eric Joanis. 2002. Automatic Verb Classification using a General Feature Space. Master's thesis, Department of Computer Science, University of Toronto.
- 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*, pages 680–686, Montreal, Canada.
- Anna Korhonen. 2002. *Subcategorization Acquisition*. Ph.D. thesis, University of Cambridge, Computer Laboratory. Technical Report UCAM-CL-TR-530.
- Claudia Kunze. 2000. Extension and Use of GermaNet, a Lexical-Semantic Database. In *Proceedings of the 2nd International Conference on Language Resources and Evaluation*, pages 999–1002, Athens, Greece.
- Beth Levin. 1993. *English Verb Classes and Alternations*. The University of Chicago Press.
- Paola Merlo and Suzanne Stevenson. 2001. Automatic Verb Classification Based on Statistical Distributions of Argument Structure. *Computational Linguistics*, 27(3):373–408.
- 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*, pages 649–655, Saarbrücken, Germany.
- Sabine Schulte im Walde and Chris Brew. 2002. Inducing German Semantic Verb Classes from Purely Syntactic Subcategorisation Information. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 223–230, Philadelphia, PA.
- 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. 2002. 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*, pages 187–197, Copenhagen, Denmark.
- Helmut Schumacher. 1986. *Verben in Feldern*. de Gruyter, Berlin.