Summary

1. Motivation

The verb is central to the structure and the meaning of a sentence, and therefore lexical verb resources play an important role in supporting computational applications in Natural Language Processing (NLP). But it is tedious and rather impossible to manually define the details of human language. Therefore, especially semantic lexical resources represent a bottleneck in NLP, and methods for the acquisition of large amounts of knowledge with comparably little manual effort have gained importance. In this context, I have investigated the potential and the limits of an automatic acquisition of semantic classes for German verbs.

Semantic Verb Classes

Semantic verb classes are an artificial construct of natural language which generalises over verbs according to their semantic properties. They represent a practical means to capture large amounts of verb knowledge without defining the idiosyncratic details for each verb. The class labels refer to the common semantic properties of the verbs in a class at a general conceptual level, and the idiosyncratic lexical semantic properties of the verbs are either added to the class description or left underspecified. Examples for conceptual structures are *Position* verbs such as *liegen* 'to lie', *sitzen* 'to sit', *stehen* 'to stand', and *Manner of Motion with a Vehicle* verbs such as *fahren* 'to drive', *fliegen* 'to fly', *rudern* 'to row'. Semantic verb classes have been defined for several languages, the most dominant examples concerning English (Levin, 1993; Baker *et al.*, 1998) and Spanish (Vázquez *et al.*, 2000). To my knowledge, no German verb classification is available for NLP applications. Such a classification would therefore provide a principled basis for filling a gap in available lexical knowledge.

What is the usage of verb classes in Natural Language Processing applications? On the one hand, verb classes reduce redundancy in verb descriptions, since they encode the common properties of verbs. On the other hand, verb classes can predict and refine properties of a verb that received insufficient empirical evidence, with reference to verbs in the same class: under this aspect, a verb classification is especially useful for the pervasive problem of data sparseness in NLP, where little or no knowledge is provided for rare events. Previous work on verb classes has proven their usefulness: particularly the English verb classification by Levin (1993) has been used for NLP applications such as word sense disambiguation (Dorr and Jones, 1996), machine translation (Dorr, 1997), and document classification (Klavans and Kan, 1998).

Automatic Induction of German Semantic Verb Classes

But how can we obtain a semantic classification of verbs, avoiding a tedious manual definition of the verbs and the classes? A semantic classification demands a definition of semantic properties, but it is difficult to automatically induce semantic features from available resources, both with respect to lexical semantics and conceptual structure. Therefore, the construction of semantic classes typically benefits from a long-standing linguistic hypothesis which asserts a tight connection between the lexical meaning of a verb and its behaviour: To a certain extent, the lexical meaning of a verb determines its behaviour, particularly with respect to the choice of its arguments, cf. Levin (1993, page 1). We can utilise this meaning-behaviour relationship in that we induce a verb classification on basis of verb features describing verb behaviour (which are easier to obtain automatically than semantic features) and expect the resulting behaviour-classification to agree with a semantic classification to a certain extent.

A common approach to define verb behaviour is captured by the diathesis alternation of verbs. Alternations are alternative constructions at the syntax-semantic interface which express the same or a similar conceptual idea of a verb. In Example (1), the most common alternations for the *Manner of Motion with a Vehicle* verb *fahren* 'to drive' are illustrated. The participants in the conceptual structure are a vehicle, a driver, a driven person, and a direction. In (a), the vehicle is expressed as subject in a transitive verb construction, with a prepositional phrase indicating the direction. In (b), the driver is expressed as subject in a transitive verb construction, with a prepositional phrase indicating the direction. In (c), the driver is expressed as subject in a transitive verb construction, with an accusative noun phrase indicating the vehicle. In (d), the driver is expressed as subject in a ditransitive verb construction, with an accusative noun phrase indicating a driven person, and a prepositional phrase indicating the direction. Even if a certain participant is not realised within an alternation, its contribution might be implicitly defined by the verb. For example, in (a) the driver is not expressed overtly, but we know that there is a driver, and in (b) and (d) the vehicle is not expressed overtly, but we know that there is a vehicle.

(1) (a) *Der Wagen fährt in die Innenstadt.*
'The car drives to the city centre.'

(b) *Die Frau fährt nach Hause.*
'The woman drives home.'

(c) *Der Filius fährt einen blauen Ferrari.*
'The son drives a blue Ferrari.'

(d) *Der Junge fährt seinen Vater zum Zug.*
'The boy drives his father to the train.'

Assuming that the verb behaviour can be captured by the diathesis alternation of a verb, which are the relevant syntactic and semantic properties for a verb description? The syntactic structures are relevant for the argument functions, the prepositions are relevant to distinguish e.g. directions from locations, and the selectional preferences of the conceptual entities are relevant, since they determine the participant roles. Therefore, I have chosen exactly these three feature levels to describe the verbs by their behaviour.

Assuming that we are provided with a feature description for verb behaviour, how can we obtain a semantic verb classification? I have applied a clustering algorithm which uses the syntactico-semantic descriptions of the German verbs as empirical verb properties and learns to induce a semantic classification from this input data. But sometimes it is something of a black art when applying multivariate clustering to high-dimensional natural language data, since we do not necessarily find out about the relevance of data types or the interpretation of the data by the clustering algorithm. But the data and the clustering technique should be based on the linguistic background of the task. Therefore, I have focused on the following sub-goals of the clustering task: I empirically investigated the definition and the practical usage of the relationship between verb meaning and verb behaviour, i.e. (i) which exactly are the semantic features that define verb classes, (ii) which exactly are the features that define verb behaviour, and (iii) can we use the meaning-behaviour relationship of verbs to induce verb classes, and to what extent does the meaning-behaviour relationship hold? In addition, I investigated the relationship between clustering idea, clustering parameters and clustering result, in order to develop a clustering methodology which is suitable for the demands of natural language. The clustering outcome cannot be a perfect semantic verb classification, since (i) the meaning-behaviour relationship on which we rely for the clustering is not perfect, and (ii) the clustering method is not perfect for the ambiguous verb data. But only if we understand the potential and the limits of the sub-goals, we can develop a methodology which can be applied to large-scale data.

2. Clustering Methodology

The clustering methodology brings together the concept of a German semantic verb classification, empirical data for a verb description at the syntax-semantic interface, a clustering technique, and methods for the evaluation of the clustering experiments. The clustering results are interpreted with respect to the empirical relation between verb meaning and verb behaviour, the development of a methodology for natural language clustering, and the acquisition of semantic verb classes.

German Verb Classes

I manually defined 43 German semantic verb classes containing 168 partly ambiguous German verbs. The construction of the German verb classes is primarily based on semantic intuition: Verbs are assigned to classes according to similarity of lexical and conceptual meaning, and each verb class is assigned a conceptual class label. The class labels are given on two conceptual levels; coarse labels such as *Manner of Motion* are sub-divided into finer labels, such as *Locomotion, Rotation, Rush, Vehicle, Flotation*. Because of the meaning-behaviour relationship at the syntax-semantic interface, the verbs grouped in one class show a certain agreement in their behaviour. The class size is between 2 and 7, with an average of 3.9 verbs per class. Eight verbs are ambiguous with respect to class membership. The classes include both high and low frequency verbs: the corpus frequencies of the verbs range from 8 to 71,604. The manual classification represents a gold standard in order to evaluate the reliability and performance of the clustering experiments.

I provide a detailed class description which is closely related to Fillmore's scenes-and-frames semantics (Fillmore, 1977, 1982), as computationally utilised in *FrameNet* (Baker *et al.*, 1998; Johnson *et al.*, 2002). The frame-semantic class definition contains a prose scene description, predominant frame participant and modification roles, and frame variants describing the scene. The frame roles have been developed on basis of a large German newspaper corpus from the 1990s. They capture the scene description by idiosyncratic participant names and demarcate major and minor roles. Since a scene might be activated by various frame embeddings, I have listed the predominant frame variants as found in the corpus, marked with participating roles, and at least one example sentence of each verb utilising the respective frame. The corpus examples are annotated and illustrate the idiosyncratic combinations of lexical verb meaning and conceptual constructions, to capture the variants of verb senses. The frame variants with their roles marked represent the alternation potential of the verbs, by connecting the different syntactic embeddings to identical role definitions.

Empirical Distributions for German Verbs

I have developed, implemented and trained a statistical grammar model for German which is based on the framework of head-lexicalised probabilistic context-free grammars. The idea originates from Charniak (1995), and this work has used an implementation by Schmid (2000). The statistical grammar model provides empirical lexical information, specialising on but not restricted to the subcategorisation behaviour of verbs.

The German verbs are represented by distributional vectors, with features and feature values in the distribution being acquired from the statistical grammar. The distributional description is based on the hypothesis that 'each language can be described in terms of a distributional structure, i.e. in terms of the occurrence of parts relative to other parts', cf. Harris (1968). The verbs are distributionally described on three levels at the syntax-semantic interface, each of them refining the previous level by additional information. The first level $D1$ encodes a purely syntactic definition of verb subcategorisation, the second level $D2$ encodes a syntactico-semantic definition of subcategorisation with prepositional preferences, and the third level $D3$ encodes a syntactico-semantic definition of subcategorisation with prepositional and selectional preferences. So the refinement of verb features starts with a purely syntactic definition and step-wise adds semantic information. The most elaborated description comes close to a definition of the verb alternation behaviour. I have decided on this three step proceeding of verb descriptions, because the resulting clusters and even more the changes in clustering results which come with a change of features should provide insight into the meaning-behaviour relationship at the syntax-semantic interface.

The following table 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. The frame types indicate possible arguments in the frames: 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). Prepositional phrases in the frames are referred to by case and preposition, such as 'mit$_{Dat}$', and 'für$_{Akk}$'. The selectional preferences in the frames on $D3$ refer to the 15 top level nodes of *GermaNet* (Hamp and Feldweg, 1997; Kunze, 2000): *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'. The preferences have been obtained by frequency propagation through GermaNet on basis of nominal fillers for arguments slots in the grammar. The relevant frame slot for selectional preference refinement is underlined. The core part of the verb description, the subcategorisation frames, has been evaluated on levels $D1$ and $D2$: Schulte im Walde (2002b) describes the induction of a subcategorisation lexicon from the grammar model for a total of 14,229 verbs with a frequency between 1 and 255,676. Schulte im Walde (2002a) performs an evaluation of the subcategorisation data against manual dictionary entries and shows that the lexical entries hold a potential for adding to and improving manual verb definitions. The evaluation results justify the utilisation of the subcategorisation frames as a valuable component for supporting NLP-tasks.

Verb	Distribution					
	$D1$		$D2$		$D3$	
beginnen	np	0.43	n	0.28	$\underline{\text{n}}$(Situation)	0.12
'to begin'	n	0.28	np:um$_{Akk}$	0.16	$\underline{\text{n}}$p:um$_{Akk}$(Situation)	0.09
	ni	0.09	ni	0.09	$\underline{\text{n}}$p:mit$_{Dat}$(Situation)	0.04
	na	0.07	np:mit$_{Dat}$	0.08	$\underline{\text{n}}$i(Lebewesen)	0.03
	nd	0.04	na	0.07	$\underline{\text{n}}$(Zustand)	0.03
	nap	0.03	np:an$_{Dat}$	0.06	$\underline{\text{n}}$p:an$_{Dat}$(Situation)	0.03
	nad	0.03	np:in$_{Dat}$	0.06	$\underline{\text{n}}$p:in$_{Dat}$(Situation)	0.03
	nir	0.01	nd	0.04	$\underline{\text{n}}$(Zeit)	0.03
	ns-2	0.01	nad	0.03	$\underline{\text{n}}$(Sache)	0.02
	xp	0.01	np:nach$_{Dat}$	0.01	n$\underline{\text{a}}$(Situation)	0.02
essen	na	0.42	na	0.42	$\underline{\text{n}}$a(Lebewesen)	0.33
'to eat'	n	0.26	n	0.26	n$\underline{\text{a}}$(Nahrung)	0.17
	nad	0.10	nad	0.10	n$\underline{\text{a}}$(Sache)	0.09
	np	0.06	nd	0.05	$\underline{\text{n}}$(Lebewesen)	0.08
	nd	0.05	ns-2	0.02	n$\underline{\text{a}}$(Lebewesen)	0.07
	nap	0.04	np:auf$_{Dat}$	0.02	$\underline{\text{n}}$(Nahrung)	0.06
	ns-2	0.02	ns-w	0.01	$\underline{\text{n}}$(Sache)	0.04
	ns-w	0.01	ni	0.01	$\underline{\text{n}}$d(Lebewesen)	0.04
	ni	0.01	np:mit$_{Dat}$	0.01	n$\underline{\text{d}}$(Nahrung)	0.02
	nas-2	0.01	np:in$_{Dat}$	0.01	n$\underline{\text{a}}$(Attribut)	0.02
fahren	n	0.34	n	0.34	$\underline{\text{n}}$(Sache)	0.12
'to drive'	np	0.29	na	0.19	$\underline{\text{n}}$(Lebewesen)	0.10
	na	0.19	np:in$_{Akk}$	0.05	$\underline{\text{n}}$a(Lebewesen)	0.08
	nap	0.06	nad	0.04	n$\underline{\text{a}}$(Sache)	0.06
	nad	0.04	np:zu$_{Dat}$	0.04	$\underline{\text{n}}$(Ort)	0.06
	nd	0.04	nd	0.04	$\underline{\text{n}}$a(Sache)	0.05
	ni	0.01	np:nach$_{Dat}$	0.04	$\underline{\text{n}}$p:in$_{Akk}$(Sache)	0.02
	ns-2	0.01	np:mit$_{Dat}$	0.03	$\underline{\text{n}}$p:zu$_{Dat}$(Sache)	0.02
	ndp	0.01	np:in$_{Dat}$	0.03	$\underline{\text{n}}$p:in$_{Akk}$(Lebewesen)	0.02
	ns-w	0.01	np:auf$_{Dat}$	0.02	$\underline{\text{n}}$p:nach$_{Dat}$(Sache)	0.02

$D1$ 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*$_{Akk}$ and *nach*$_{Dat}$, *mit*$_{Dat}$ referring to the begun event, *an*$_{Dat}$ as date and *in*$_{Dat}$ as place indicator. It is obvious that not all PPs are argument PPs, but also adjunct PPs represent 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 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*$_{Akk}$, *zu*$_{Dat}$, *nach*$_{Dat}$) or referring to a means of motion (*mit*$_{Dat}$, *in*$_{Dat}$, *auf*$_{Dat}$). 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'.

Clustering and Evaluation Techniques

The clustering of the German verbs was performed by the k-Means algorithm, a standard unsupervised clustering technique as proposed by Forgy (1965). With k-Means, 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. Applying the k-Means algorithm assumes (i) that verbs are represented by distributional vectors, and (ii) that verbs which are closer to each other in a mathematically defined way are also more similar to each other in a linguistic way.

k-Means includes various cluster parameters: The number of clusters is not known beforehand, so the clustering experiments investigate this parameter. Related to this parameter is the level of conceptual structure: the more verb clusters are found, the more specific the conceptual level, and vice versa. The clustering input was varied according to how much pre-processing we invested. k-Means is sensitive to the input, and the resulting cluster shape should match the idea of verb classes. I therefore tried random cluster input and hierarchically pre-processed cluster input (with amalgamations single-linkage, complete-linkage, average distance between verbs, distance between cluster centroids, Ward's method) to investigate the impact of the input on the output. In addition, we can find various notions of defining the similarity between distributional vectors. But which does best fit the idea of verb similarity? The potential and the restrictions of the natural language clustering approach have been developed with reference to a small-scale German verb classification and discussed and tested on the acquisition of a large-scale German verb classification.

A clustering evaluation demands an independent and reliable measure for the assessment and comparison of clustering experiments and results. In theory, the clustering researcher has acquired an intuition for the clustering evaluation, but in practise the mass of data on the one hand and the subtle details of data representation and clustering algorithms on the other hand make an intuitive judgement impossible. There is no absolute scheme with which to measure clusterings, but a variety of evaluation measures from diverse areas such as theoretical statistics, machine vision and web-page clustering are generally applicable. Based on a series of general evaluation demands, general clustering demands and specific linguistic clustering demands, I compared a number of measures against each other and according to the demands, and determined three measures as the most suitable for the linguistic clustering task: a pair-wise precision and recall measure which has been used in adjective clustering before (Hatzivassiloglou and McKeown, 1993) and provides an easy to understand percentage, an adjusted pair-wise precision measure which introduces a scaling factor based on the size of clusters and comes closest to the linguistic demands on a clustering result (Schulte im Walde and Brew, 2002), and the adjusted Rand index which is a measure of agreement vs. disagreement between object pairs in clusterings that is corrected for chance (Hubert and Arabie, 1985) and provides the most appropriate reference to a null model. The measures compared the results of clustering experiments against the manual verb classification as gold standard.

Clustering Examples

For illustrative purposes, I present representative parts of the cluster analysis as based on the following parameters: 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 (43); similarity measure is performed by the skew divergence, a variant of the Kullback-Leibler divergence. The cluster analysis is based on the verb description on $D3$, with selectional roles for '$\underline{n}$', '$\underline{na}$', '$\underline{nd}$', 'n$\underline{a}$d', '$\underline{n}$s-dass'. I compare the respective clusters with their pendants under $D1$ and $D2$. For each cluster, the verbs which belong to the same gold standard class are presented in one line, accompanied by the class label.

(a) beginnen enden – *Aspect*
bestehen existieren – *Existence*
liegen sitzen stehen – *Position*
laufen – *Manner of Motion: Locomotion*

(b) kriechen rennen – *Manner of Motion: Locomotion*
eilen – *Manner of Motion: Rush*
gleiten – *Manner of Motion: Flotation*
starren – *Facial Expression*

(c) klettern wandern – *Manner of Motion: Locomotion*
fahren fliegen segeln – *Manner of Motion: Vehicle*
fließen – *Manner of Motion: Flotation*

(d) festlegen – *Constitution*
bilden – *Production*
erhöhen senken steigern vergrößern verkleinern – *Quantum Change*

(e) töten – *Elimination*
unterrichten – *Teaching*

(f) nieseln regnen schneien – *Weather*

(g) dämmern – *Weather*

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*$_{Dat}$ 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 *Locomotion, Rotation, Rush, Vehicle, Flotation*) 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. *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*$_{Akk}$ and *um*$_{Akk}$. 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.

The same cluster analysis has been applied in a large-scale experiment: I extracted all German verbs from the statistical grammar model which appeared with an empirical frequency between 500 and 10,000 in the training corpus. This selection resulted in a total of 809 verbs, including 94 verbs from the preliminary set of 168 verbs. I added the remaining verbs of the preliminary set (because of evaluation reasons), resulting in a total selection of 883 German verbs. The number of clusters was set to 100, which corresponds to an average of 8.83 verbs per cluster. Some clusters are extremely good with respect to the semantic overlap of the verbs, some clusters contain a number of similar verbs mixed with semantically different verbs, and for some clusters it is difficult to recognise a common semantic aspect of the verbs. For each kind of result I present examples. The verbs which I think semantically similar are marked in bold font.

(a) *anhören* 'to listen', *auswirken* 'to affect', *einigen* 'to agree', *lohnen* 'to be worth', *verhalten* 'to behave', *wandeln* 'to promenade'

(b) *beschleunigen* 'to speed up', ***bilden*** 'to constitute', *darstellen* 'to illustrate', *decken* 'to cover', *erfüllen* 'to fulfil', ***erhöhen*** 'to raise', *erledigen* 'to fulfil', *finanzieren* 'to finance', *füllen* 'to fill', *lösen* 'to solve', *rechtfertigen* 'to justify', ***reduzieren*** 'to reduce', ***senken*** 'to lower', ***steigern*** 'to increase', ***verbessern*** 'to improve', ***vergrößern*** 'to enlarge', ***verkleinern*** 'to make smaller', ***verringern*** 'to decrease', ***verschieben*** 'to shift', ***verschärfen*** 'to intensify', ***verstärken*** 'to intensify', ***verändern*** 'to change'

(c) ***ahnen*** 'to guess', <u>***bedauern***</u> 'to regret', <u>***befürchten***</u> 'to fear', <u>***bezweifeln***</u> 'to doubt', ***merken*** 'to notice', ***vermuten*** 'to assume', *weißen* 'to whiten', ***wissen*** 'to know'

(d) ***anbieten*** 'to offer', *angebieten* is a morphologically mistaken perfect participle of 'to offer', ***bieten*** 'to offer', ***erlauben*** 'to allow', ***erleichtern*** 'to facilitate', ***ermöglichen*** 'to make possible', ***eröffnen*** 'to open', ***untersagen*** 'to forbid', *veranstalten* 'to arrange', ***verbieten*** 'to forbid'

(e) ***basieren*** 'to be based on', ***beruhen*** 'to be based on', ***resultieren*** 'to result from', ***stammen*** 'to stem from'

(f) ***befragen*** 'to interrogate', ***entlassen*** 'to release', ***ermorden*** 'to assassinate', ***erschießen*** 'to shoot', ***festnehmen*** 'to arrest', ***töten*** 'to kill', ***verhaften*** 'to arrest'

Cluster (a) is an example cluster where the verbs do not share meaning aspects. In the overall cluster analysis, the semantically incoherent clusters tend to be rather large, i.e. with more than 15-20 verb members. Cluster (b) is an example cluster where a part of the verbs shows overlap in their meaning aspects (quantum change), but the clusters also contain considerable noise. Clusters (c) to (f) are example clusters where most or all verbs show a strong similarity in their conceptual structures: Cluster (c) contains verbs expressing a propositional attitude; the underlined verbs in addition indicate an emotion. The only unmarked verb *weißen* also fits into the cluster, since it is a morphological lemma mistake changed with *wissen* which belongs to the verb class. The verbs in cluster (d) describe a scene where somebody or some situation makes something possible (in the positive or negative sense). Next to a lemmatising mistake (*angebieten* is not an infinitive, but a morphologically mistaken perfect participle of *anbieten*), the only exception verb is *veranstalten*. In cluster (e) all verbs refer to a basis, and in cluster (f) the verbs describe the process from arresting to treating a suspect. A number of semantically coherent classes needs little manual correction as a lexical resource. Semantically diverse verb classes and clustering mistakes need to be split into finer and more coherent clusters, or to be filtered from the classification.

3. Conclusions

I have presented a clustering methodology for German verbs whose results agree with the manual classification in many respects and should prove useful as automatic basis for a large-scale clustering. I did not arbitrarily set the parameters, but tried to find an at least near-optimal compromise between linguistic and practical demands. Without any doubt the cluster analysis needs manual correction and completion, but represents a plausible basis. Key issues of the clustering methodology refer to linguistic aspects on the one hand, and to technical aspects on the other hand.

Linguistic Aspects

The strategy of utilising subcategorisation frames, prepositional information and selectional preferences to define the verb features has proven successful, since the experiments illustrated a tight connection between the induced verb behaviour and the constitution of the semantic verb classes. In addition, each level of representation has generated a positive effect on the clustering and improved the less informative level. The experiments present evidence for a linguistically defined limit on the usefulness of the verb features, which is driven by the dividing line between the common and idiosyncratic features of verbs in a verb class. 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. The meaning of verbs comprises both (a) properties which are general for the respective verb classes, and (b) 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 by step-wise refining the verb description and including lexical idiosyncrasy, the emphasis of the common properties vanishes. From the theoretical point of view, the distinction between common and idiosyncratic features is obvious, but from the practical point of view there is no unique perfect choice and encoding of the verb features. The feature choice depends on the specific properties of the desired verb classes, but even if classes are perfectly defined on a common conceptual level, the relevant level of behavioural properties of the verb classes might differ. The investigated combination within this thesis has proven a useful compromise for feature description.

Technical Aspects

I have investigated the relationship between clustering idea, clustering parameters and clustering result, in order to develop a clustering methodology which is suitable for the demands of natural language. The clustering input plays an important role. k-Means needs similarly-sized clusters in order to achieve a linguistically meaningful classification. The linguistically most successful input clusters are therefore based on hierarchical clustering with complete linkage or Ward's method, since their clusters are comparably balanced in size and correspond to compact cluster shapes. The hierarchical clusterings actually reach similar clustering outputs than k-Means, which is due to the similarity of the clustering methods with respect to the common clustering criterion of optimising the sum of distances between verbs and cluster centroids. The similarity measure used in the clustering experiments is of secondary importance, since the differences in clustering with varying the measure are negligible. For larger object and feature sets, Kullback-Leibler variants show a tendency to outperform other measures, confirming language-based results on distributional similarity by Lee (2001). Both frequencies and probabilities represent a useful basis for the verb distributions. A simple smoothing of the distributions supports the clustering, but to be sure of the effect one would need to experiment with solid smoothing methods. The number of clusters only plays a role concerning the magnitude of numbers. Inducing fine-grained clusters as given in the manual classification seems an ambitious intention, because the feature distinction for the classes is fine-grained, too. Inducing coarse clusters provides a coarse classification which is object to less noise and easier for manual correction.