Supervised Classification of Aspectual Verb Classes in German
Subcategorization-Frame-Based vs Window-Based Approach: A Comparison

Jürgen Hermes[1], Michael Richter[2] and Claes Neuefeind[1]
[1] *Institute for Digital Humanities, Cologne University, Albertus-Magnus-Platz, Cologne, Germany*
[2] *Department of Automatic Language Processing, Leipzig University, Augustusplatz 10, Leipzig, Germany*

Keywords: Machine Learning, Classification, Aspectual Verb Classes, Subcategorization Frame based Approach, Window based Approach.

Abstract: The present study examines the results of two experiments: the aspectual classification of German verbs within a window-based distributional framework and the classification within a subcategorization-frame-based framework. The predictive power of pure, unstructured co-occurrences of verbs is compared against that of linguistically motivated, well defined co-occurrences which we denote as *informed distributional framework*. Using a support vector machine classifier and a classification into an extended Vendler classification (Vendler, 1967) as the gold standard, we observe excellent results in both frameworks which perform almost on a par.

1 INTRODUCTION

This paper presents experiments on the automatic assignment of German verbs to aspectual classes (Vendler, 1967) using corpus data in a distributional framework (Harris 1954, see also Rubenstein and Goodenough 1965, Schütze and Pedersen 1993, Landauer and Dumais 1997, Pantel 2005, Turney and Pantel 2010). The primary motivation for undertaking this research is the (almost complete) lack of studies on the automatic aspectual classification in typological research on German. In our study, we compare two approaches: i. an aspectual verb classification in a framework which utilizes subcategorization frames of verbs (henceforth *informed distributional* framework), extracting classified nouns in the argument positions of verbs (Richter and Hermes 2015, Hermes et al. 2015) and ii. an aspectual verb classification in a *purely* distributional framework, thus considering co-occurrences of all types.

We pose the question of whether classification in a purely distributional framework would yield better classification results than classification in a linguistically well grounded *informed* distributional framework. The former approach employs verb vectors of very high dimensionality consisting of a considerably higher amount of linguistic material than the vectors used in an *informed* distributional framework. This could be a point in favor of the *purely* distributional framework. On the other hand, the studies of Richter and Hermes (2015) and Hermes et al. (2015) with sets of 35 and 95 German verbs, respectively, achieved promising classification results within an *informed* distributional framework. A classification inspired by Vendler was used as gold standard. This classification is the extension of the Vendlerian typology through the addition of one class (henceforth Vendler + 1): The additional aspectual class *accomplishments with an affected subject*. The studies mentioned above (see also Richter and van Hout 2016) provide evidence for this class. The *accomplishments with an affected subject* class differs from the classical *accomplishments* in the semantic role of the subject. Instead of exclusively assigning the agent role to the subject, the subject in the *accomplishments with an affected subject* class is assigned both a patient role and an agent role. Consider verbs such as *drink*, where an agent subject also has the semantic properties of a patient since the drinker-agent is affected and undergoes a change of state (temporarily puts on weight, gets drunk etc.). Naess (2007) refers to these semantic roles as *volitional undergoers*.

In the present study, we exclusively focus on lexical aspect that is, aspectual properties of bare verbs (or *the fundamental aspectual category* in the terminology of Siegel and McKeown 2000). Thus, aspectual properties of sentences and VPs as results of aspectual coercion or aspectual shift, respectively, are not subject of this study, rather it is our aim to predict the aspectual classes of verbs from their con-

Hermes, J., Richter, M. and Neuefeind, C.
Supervised Classification of Aspectual Verb Classes in German - Subcategorization-Frame-Based vs Window-Based Approach: A Comparison.
DOI: 10.5220/0006728106530662
In *Proceedings of the 10th International Conference on Agents and Artificial Intelligence (ICAART 2018) - Volume 2*, pages 653-662
ISBN: 978-989-758-275-2

texts.[1] The aspectual classification of verb classes from contexts might give indications how language learners manage to build up aspectual verb classes in their mental lexicon. From that perspective, research on verb classes is vital due to their relevance for the processing of natural language by human beings and, in addition, for the theory of natural language acquisition (see Tomasello 2000, Goldberg 1995, Naigles et al. 1992, Naigles et al. 1995, Naigles et al. 1993, Wittek 2002, Richter and van Hout 2013). Research on aspectual classes is of particular relevance because it models the temporal and causal structures of events (see Vendler 1967, Dowty 1979, Dowty 1991, Rothstein 2004, Fernando 2004, Gruender 2008). Theory-driven work by Klein (2009) and experimental studies by Siegel (1997) and Siegel and McKeown (2000) highlight the potential of aspect for classifying linguistic units such as verbs and documents.

2 RELATED WORK

There are hardly any studies which address automatic classifications of the complete Vendlerian typology, let alone studies which compare *subcategorization-frame-based* against *window-based* approaches. By focusing on tense forms of verbs, Klavans and Chodorow (1992) determined gradual state-properties of verbs. Siegel (1997) and Siegel and McKeown (2000) classified verbs into states and events using temporal and modal indicators from contexts such as temporal adverbs, tense forms and manner and evaluation adverbs. Zarcone and Lenci (2008) presented an automatic classification of the four Vendlerian aspect classes in Italian utilizing, amongst others, syntactic and semantic features of the arguments of the target verbs and verb tense. The authors, however, aimed at modelling aspectual shift and consequently focus on aspectual properties of sentences, decomposing the components of sentential event types. Friedrich and Palmer (2014) presented an automatic classification of aspectual verb classes in English using contextual features including tense forms, albeit only distinguishing between stative, dynamic and mixed type verbs.

Studies on the automatic assignment of non-aspectual verb classes within a distributional framework from Dorr and Jones (1996), Merlo and Stevenson (2001), Joanis et al. (2008), Vlachos et al. (2009), Schulte im Walde and Brew (2002), Schulte im Walde (2003) and Schulte im Walde (2006) for German verbs provide corpus based evidence that argument frames, syntactic subcategorization information and, in addition, aspect (Joanis et al., 2008) are reliable predictors.

3 VENDLER'S TYPOLOGY

Vendler (1967) defines four aspectual classes [2]: *States*, *activities*, *achievements* and *accomplishments*, based on the time schemata of verbs and verb phrases. He gives the following illustrative examples (Vendler, 1967, 149): Activities such as *A was running at time t* are true if the instantiation of t is on a time stretch throughout A was running. An accomplishment such as *A was drawing the circle at t* is true if t is on the time stretch in which A drew that circle. An achievement such as *A won the race between* t_1 *and* t_2 means that the time at which A won that race is between t_1 and t_2; and a statement such as *A loves somebody from* t_1 *and* t_2 means that at any instant between t_1 and t_2 A loved that person.

In (1) below we give Pustejovsky (1991)'s description of the four Vendler classes and in addition a description of the additional class *accomplishments with an affected subject* which extends the description of the accomplishment class with a subject variable. In line with Vendler (1967), Bach (1986), and Dowty (1979), Pustejovsky (1991) distinguished the event types *process*, *state* and *transition* (see also Jackendoff 1972, Lakoff 1970 and Wright 1963). The latter is a function from any event type to its opposite. For instance, *x closes the door* expresses a transition from an event type e_1, the open door, to an event type e_2, the closed door, by acting of agent x and e_2 is the opposite of e_1 ($\neg e_1$). The combinatorial variations within the three event types *process*, *state* and *transition* allows for the formal description of the complete Vendlerian typology.

[1]As an example of aspectual coercion, consider an atelic verb such as *walk*, which can be combined with a PP denoting a destination as in *he walks to the store* and expressing a telic event. The sentence *walks to the store* is no longer an activity, instead, it is an accomplishment. Aspectual coercion can also be triggered by quantification Krifka (1989). A prototypical accomplishment verb such as *kill* can occur in a sentence expressing an activity, as in *he is killing carpet moths*(note the present progressive form of the verb) which stands classical tests of activities, e.g. *he is killing carpet moths for an hour, permanently/forever*. The direct object is a bare plural, expressing cumulative objects (Quine, 1960) which combine well with atelic verbs. With a quantized direct object (Krifka, 1989) however, the sentence is clearly telic: *he kills two carpet moths in one hour/*for hours*.

[2]The Vendlerian quadripartition has been modified and extended: Dowty (1979) added *degree achievements*, Smith (1991) added *semelfactices*, Verkuyl (2005) in contrast defined a tripartition consisting of *states*, *processes* and *events*.

The abbreviation 'LCS' in (1) means the *lexical conceptual structure* which gives a decomposition of predicates (Dowty 1991, Jackendoff 1983, Levin and Rapoport 1992, Pustejovsky 1991). Hence, LCS is the minimal decomposed event structure of verbs.

(1) **Accomplishments**

$ES: \quad process \xrightarrow{transition} state$
$LCS': [act(x,y) \& \neg Q(y)] \qquad [Q(y)]$
$LCS: cause([act(x,y)], become(Q(y)))$

Achievements

$ES: \quad process \xrightarrow{transition} state$
$LCS': [\neg P(y)] \qquad [P(y)]$
$LCS: become(P(y))$

Activities

$ES: \quad process$
$LCS': [act(x)]$
$LCS: act(x)$

States

$ES: \quad state$
$LCS': [Q(x)]$
$LCS: Q(x)$

Accomplishments with an affected subject

$ES: \quad process \xrightarrow{transition} state$
$LCS': [act(x,y) \& \neg P(x) \& \neg Q(y)] \quad [P(x) \& Q(y)]$
$LCS: cause([act(x,y)], become(P(x), become(Q(y)))$

The classes *Accomplishments* and *accomplishments with an affected subject* can be interpreted as a combination of *activities* and *achievements*. The latter express a result, the former a result preceded by an activity. Pulman (1997) gives a slightly different description of the *achievements* class: he formulates a transition from the event type *point* instead of *process* to *state* where *point* is an atomic event whose internal temporal structure (if it may have any) is ignored.[3]

4 METHODOLOGY

In order to answer the question of whether informed distributional methods perform better with regards to the prediction of aspectual verb classes than distributional methods on shallow input, we tested different workflows to classify a selection of 95 common German verbs taken from Schumacher (1986). Schumacher defines seven lexical semantic macrofields: *Verben der allgemeinen Existenz* (verbs of general existence), *Verben der speziellen Existenz* (verbs of special existence), *Verben des sprachlichen Ausdrucks* (verbs of linguistic expression), *Verben der Differenz* (verbs of difference), *Verben der Relation und des geistigen Handelns* (verbs of relation and mental processing), *Verben des Handlungsspielraums* (verbs of freedom of action) and *Verben der vitalen Bedrfnisse* (verbs of vital needs). The macrofields are split into 30 subfields. We chose the verbs randomly from the thirty subfields, the only criterion being the inclusion of every subfield in order to cover the complete semantic range of Schumachers typology.

Figure 1 shows the workflow of our analyses, starting with raw sentence data taken from the SdeWaC corpus (Faaß and Eckert, 2013) at the top, to sets of classified verbs at the bottom. The different methods are applied in four experiments, each employing an individual process chain combining a different set of components. The four experiments are depicted as varying paths in figure 1, all starting at the top (sDeWaC) but ending with four different sets of verb classes at the bottom.

An overview showing which method combination is unique in each experiment is given in table 1.

Table 1: Combination of workflow elements.

Combination	- N Cluster	+ N Cluster
- Parsed Input	1	2
+ Parsed Input	3	4

We implemented each process chain on the basis of combined and configurable components within the workflow management tool Tesla[4] so that every experiment performed can be reproduced by other researchers.

For the classification of the 95 verbs, we used a Support Vector Machine classifier (Joachims, 1998) with a non-linear kernel. For 35 verbs, we adapted the aspectual classification as the gold standard which was validated in Richter and van Hout (2016), i.e. Vendler + 1, and we assigned 60 verbs to aspectual

[3] Consider the achievement verb *find*. According to Pulman (1997) the event of finding has a resultant state, the finding itself however is an atomic event, ignoring a possible complex event structure consisting for instance of discovering something on the ground, taking a decision to pick it up, bending down etc.

[4] Tesla (*T*ext *E*ngineering *S*oftware *LA*boratory), see http://tesla.spinfo.uni-koeln.de is an open source virtual research environment, integrating both a visual editor for conducting text-engineering experiments and a Java IDE for developing software components (Hermes and Schwiebert, 2009).

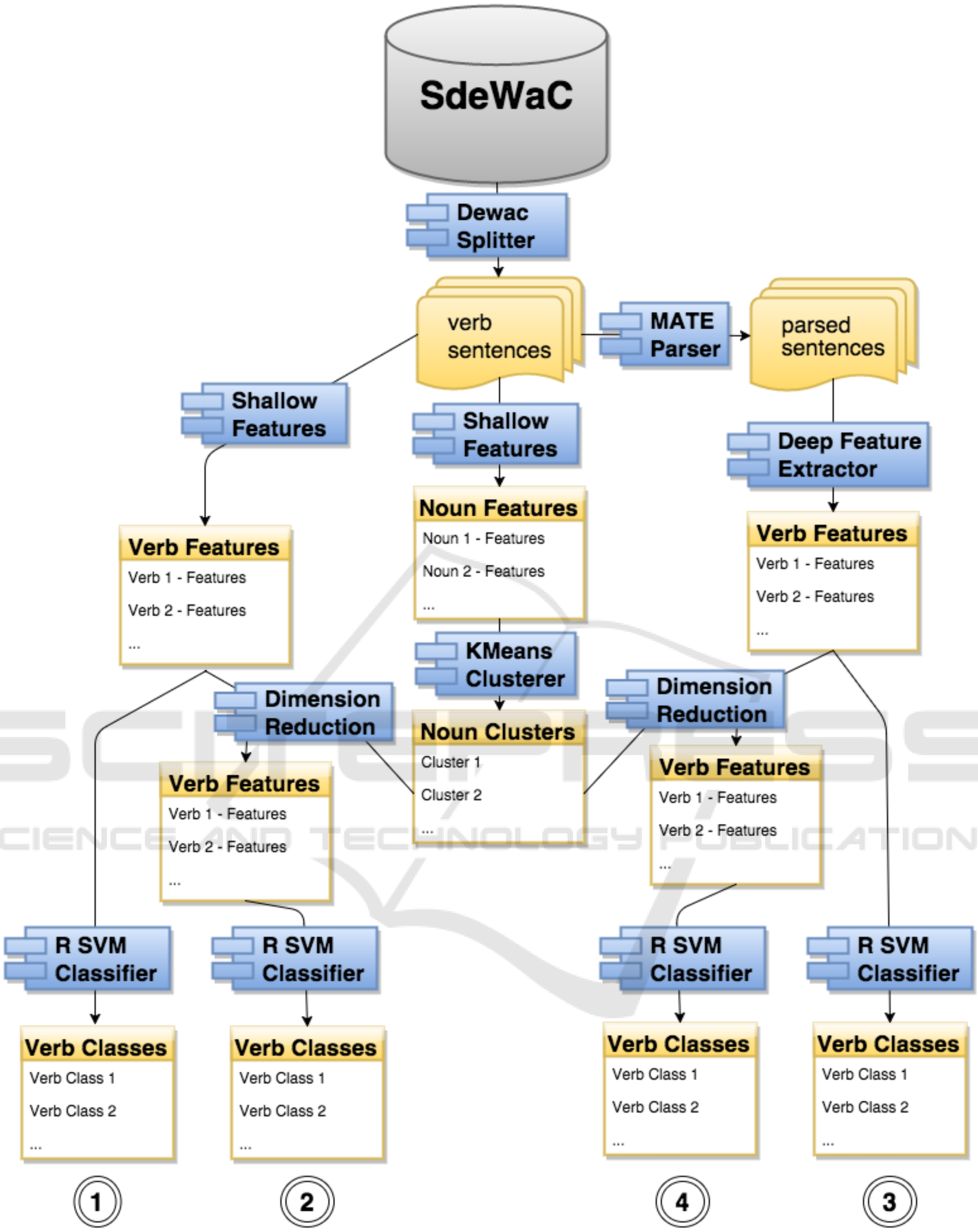

Figure 1: Overview of the complete workflow of all performed experiments. The experiments were realized via 4 process chains which can be identified by the numbers at the bottom of the figure. Each process chain uses a different combination of components (see table 1 for an overview and text for detailed explanations.

classes applying the criteria in Richter and van Hout (2016). We trained the SVM using this aspectual classification as training data and tested it with a 10-fold cross-validation: The data were randomly split into a training and a test set (proportion 90 percent (training set), 10 percent (test set). The classifier was both trained and tested on each of the 10 combinations of training and test set. Thus, in total we obtained 10 sets of class predictions and took the mean accuracy as final result. We used a SVM classifier with a polynomial kernel (which turned out to outperform other kernel types) and a multiclass classifier (instead of training a single classifier per class, we employed just one classifier for the complete set of our aspectual classes.

We give some examples of the verb classes of the aspectual gold standard classification below; the complete list of all 95 pre-classified verbs used in this study can be found in the appendix.

1. **Accomplishments**: *aufbauen* auf (to build on / to be based on), *herstellen* (to produce), *schneiden* (to cut), *zersägen* (to saw into pieces), *verlängern* (to extend), *mitteilen* (to tell / to inform), *bermitteln* (to communicate / to forward), *verhindern* (to prevent), *abgrenzen* (mark off / to define), *verändern* (to change)
2. **Accomplishments with affected subject**: *untersuchen* (to examine), *bedenken* (to consider), *erörtern* (to debate), *nachprüfen* (to ascertain / to check), *aufessen* (to eat up), *essen* (to eat), *beachten* (to note), *kaufen* (to buy)
3. **Activities**: *laufen* (to walk / to run), *eingehen auf* (to respond to so. / sth.), *hmmern* (to hammer), *sich orientieren an* (to be geared to) *ansteigen* (to increase), *fallen* (to fall), *richten auf* (to direct towards / to focus), *denken* (to think), *stattfinden* (to take place), *wachsen* (to grow)
4. **Achievements**: *einschlafen* (to fall asleep), *vergehen* (to go (by) / to pass / to diasappear), *übersehen* (to overlook), *verlieren* (to lose), *anfangen* (to begin), *abweichen* (to deviate)
5. **States**: *existieren* (to exist), *fehlen* (to lack), *müssen* (to must), *halten fr* (to take so. / sth. for so./ sth.), *folgen aus* (to follow from), *angehören* (to belong to), *übereinstimmen* (to agree), *betreffen* (to concern), *sein* (to be), *vorherrschen* (to predominate)

4.1 Classification of Verbs using Co-occurence Vectors

The first experiment (marked with 1, left side of figure 1) is taken as a benchmark for purely distributional methods: We extracted 2000 sentences for each verb from the SdeWaC corpus and collected the most frequent co-occurrences using the frequency-based heuristics described in Levy and Bullinaria (2001), simply taking the k most frequent types of our corpus as vector features.[5] The vectors were computed in three different configurations. As a baseline, we first took the 200 most frequently occurring elements (mostly closed class function words such as *und* and, *zu* to, *weil* because / since, etc.), and a context window of size 1 (henceforth 1-200), accepting only the direct neighbors as co-occurrences. In the second configuration, co-occurrences were computed against the 2000 most frequently occurring elements within a fixed context window of 5 items to both sides (henceforth 5-2k). In addition, we employed a positional weighting scheme using the HAL model (*Hyperspace Analogue to Language*, Lund and Burgess 1996). In a third configuration, we took the 10.000 most frequently occurring words and a window size of 10 words (henceforth 10-10k), again using the HAL-weighting scheme. While the restriction to function words within a narrow window mainly reflects grammar-related distributional properties, the consideration of content words in combination with a broader window and position weighting emphasizes the more semantically oriented aspects of their distribution. The resulting verb vectors were normalized and weighted with the TF-IDF measure before they were passed to the final classification step.

4.2 Classification of Verbs using Co-occurence Vectors with Reduced Dimensionality

As a first step towards more informed methods, we restricted the vector features to nominal co-occuurrences (tagged as NN and NE in the *SdeWaC* corpus). In order to reduce the feature space and to increase the allocation density of the vectors, we clustered all of these nominal co-occurrences. At this stage we set foot on path 2 in figure 1. Here, we computed co-occurrence vectors based on the same subset of the *SdeWaC* corpus that we used to determine the verb features, again using frequency-based feature selection. The resulting vectors (10-10k) were again weighted by the TF-IDF measure and passed to the cluster analysis.

For cluster analysis we used three different implementations from the *ELKI Data Mining API*[6], namely KmeansLloyd with cluster sizes of $k = 10$.

[5] We decided for the heuristics because of economy considerations (Ockhams razor), giving preference to the simpler method that performs on a par with more complex ones: As Levy and Bullinaria (2001) show in their paper, performance in tasks like synonym detection is comparable to more sophisticated methods of feature selection, such as taking the most variant elements (see Lund and Burgess 1996), the most 'reliable' (see Lowe and McDonald 2000, or to perform a dimensionality reduction (e.g. by singular value decomposition as done in the LSA model, see Landauer and Dumais 1997).

[6] The open source framework ELKI (Environment for DeveLoping KDD-Applications Supported by Index-Structures) was developed at the LMU Munich, see http://elki.dbs.ifi.lmu.de

4.3 Classification of Verb Vectors with Nominal Fillers and Aspectual Features

For the remaining experiments (paths 3 and 4 in figure 1) we preprocessed the sentences from the SdeWaC corpus with the *Mate Dependency Parser*[7] (Bohnet, 2010) to determine subjects and objects (accusative, dative, and prepositional) for each verb and, in addition, to determine their aspectual features, which Vendler (1967) suggested as a means of distinguishing aspectual verb classes. To that end we collected structures in the sentences such as adverbial fillers in dependent or governing positions, which distinguish the aspectual behaviors of the verbs being investigated. The aspectual features are:

1. verb in imperative form
2. verb complex with *aufhören / stoppen* (to stop / to finish) as governing verbs,
3. verb complex with *überzeugen* (to convince) as governing verb,
4. matrix verb with time adverbials for durations, like *minutenlang* (for minutes), *in einer Minute* (in a minute),
5. matrix verb with time units, like *Minute* (minute), *Jahrhundert* (century),
6. matrix verbs with *seit* (since), combined with unit of time,
7. matrix verb with adverbials *sorgfältig / mit Sorgfalt* (careful / with care),
8. matrix verb with adverbials *absichtlich / mit Absicht* (on purpose),
9. matrix verb with adverbials *fast / beinahe* (almost).

We generated vectors for each feature combination (subjects, direct objects, dative objects, prepositional objects, and adverbials) in order to determine which combination of fillers has the best predictive power concerning the aspectual verb classes defined by Vendler.

4.4 Classification of Verb Vectors with Nominal Fillers and Aspectual Features with Reduced Dimensionality

Finally (path 4 in figure 1), we combined the informed distributional method based on parsed input with the dimensionality reduction based on cluster analysis. We constructed the verb vectors as described in chapter 4.1 and clustered the nouns as described in chapter 4.2. Then we reassembled the verb vectors using the generated noun classes from the noun clustering subworkflow. The verb vectors could then be reduced to 39 or fewer dimensions (up to nine aspectual features complemented by ten clusters for each argument position).

[7]See: https://code.google.com/p/mate-tools/

5 RESULTS

The results are given in reverse order (from path 4 on the right to path 1 on the left of figure 1, because the method described in 4.4 was the starting point of our study.

5.1 Results of the Classification of Verbs using Dimension Reduced Nominal Fillers and Aspectual Features

The workflow is a slightly modified version of the workflow in Hermes et al. (2015). Thus, the results were practically the same: Feature combinations exclusively comprising aspect features yielded high accuracy values (see figure 2).

Input to the classification were vectors with 39 dimensions (each 10 for subjects, direct objects, and prepositional objects and 9 for the adverbial features reflecting the aspectual behaviour). Taking the classification with five aspectual verb classes as the gold standard, ten noun classes per argument position clearly outperform the approaches with fewer features. Additionally, counting every noun token leads to better results than counting only the noun types. Medium length vectors (2000 dimensions), constructed on the basis of a medium context width (window size of five elements) achieve the best outcomes; the verb vectors of the KMeansLloyd noun clustering in particular show the best performance. Figure 2 depicts the accuracy of feature combinations and is subject to the following result description.

The combinations *aspect/subject/direct object* and *aspect/subject/ direct object/prepositional object* outperform the remaining feature combinations with .95 accuracy, and .94 accuracy respectively. To determine the significance values for the accuracy level for the classification in five classes, we calculated Cohens kappa. Kappa values above .81 are characterized as almost perfect agreement and therefore highly significant. With the two feature combinations described above we reached kappa values higher than .90. The feature combinations *aspect/direct object/prepositional object* with .88 accuracy, κ = .84, and *aspect/prepositional object with* .86 accuracy, κ = .81, also achieve almost perfect agreements. Substantial agreements, with above .61, can be observed with the combinations *aspect/subject/prepositional object*, .84 accuracy, κ = .78, *aspect/direct object*, .92 accuracy, κ = .75, and *aspect/subject*, .82 accuracy, κ = .75. None of the feature combinations without aspectual features achieves a satisfactory result. The single features achieve only fair agreements: *Aspect* achieves .57 accuracy, κ = .34, *subject* achieves .52, κ = .27, and *direct object* and *prepositional object* achieve .51 accuracy and κ = .24 each.

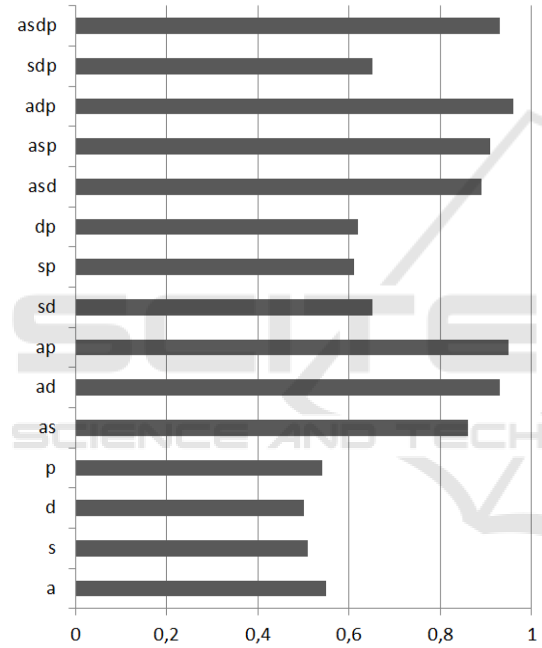

Figure 2: Results of process chain 4.

5.2 Results of the Classification of Verbs using Nominal Fillers and Aspectual Features (no Dimension Reduction)

To reduce the complexity of our workflow, we classified the verbs using feature vectors where all nominal fillers were taken into account instead of clustering them into groups (processing chain 3 in figure 1).

Instead of the 39-dimensional vectors in processing chain 4, we obtained vectors with almost 40000 dimensions (11257 subjects, 12450 direct objects, 16196 prepositional objects, 9 for adverbial fillers).

For the most part, the results were comparable to the results of processing chain 4 (see figure 3): Employing all nominal and adverbial fillers (marked as *asdp* in figure 3), we achieved .92 accuracy. In comparison with the +*cluster*-workflow we achieved the same accuracy value when we left out the adverbial fillers (*sdp*), but an even worse value when we left out the objects (*as*). This is not an overly surprising result because the adverbial fillers were limited to only nine dimensions. Within very high-dimensional vectors they should lose ground. We see here a first indication that the paradigm the more the merrier fits well with our results. We consistently achieved better results when we normalized the vectors by cosine. Without length normalization accuracy apparently decreases (at least .07, see figure 3, last bar).

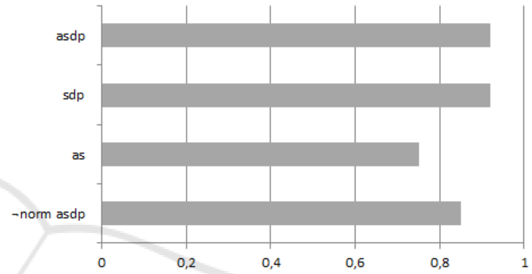

Figure 3: Results of process chain 3.

5.3 Results of the Classification of Verbs using Dimension Reduced Co-occurrences

In the third experiment (path 2 in figure 1) we left out the parsing step. Within process chain 2 we built verb vectors from co-occurrences of nouns that were clustered as illustrated above. Firstly, we generated 39 clusters to provide the same dimensionality as in process chain 4, which gave us a very poor result (see figure 4, first bar). By increasing the number of clusters, the results became better, cumulating in .90 accuracy for 500 clusters. Here again, the maxim seems to be the more the merrier. Consequently, we should try to expand the number of clusters to the number of features. That is actually what we did in process chain 1, which will be described in the next section.

5.4 Results for the Classification of Verbs using Pure Co-occurrences

In the last experiment, we classified the verbs without parsing and without clustering the noun fillers. Instead, we built the vectors by simply collecting the most frequent co-occurrences of a verb, irrespective of whether they are nominal elements or not. Although this was meant to be the baseline analysis,

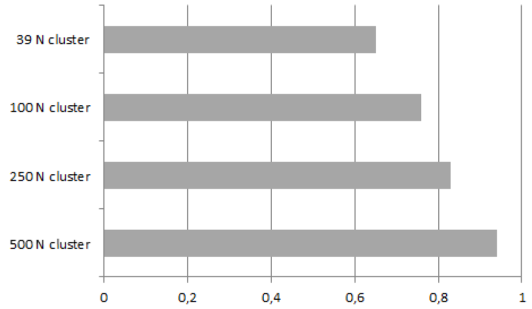

Figure 4: Results of process chain 2.

we obtained excellent results up to .98 accuracy, especially with higher dimensionality of the vectors (see figure 5).

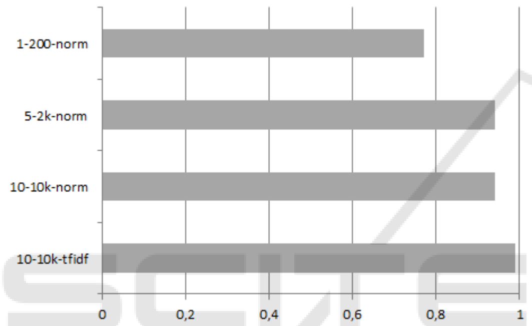

Figure 5: Results of process chain 1.

6 CONCLUSION

The present study provides evidence that Vendlerian aspectual verb classes plus a class of *accomplishments with an affected subject*, i.e. Vendler + 1 classification, can be inferred from the contexts of the target verbs. We observed that two out of our four models outperformed the remaining ones and achieved excellent agreements between the automatically inferred aspectual classes and the Vendler + 1 gold standard classification: (i) a model within an *informed* distributional framework, considering structured, language theoretically well grounded (and thus restricted) context material that is, clustered nouns in the argument positions of verbs in combination with aspectual features and (ii) a model within a *non informed* framework, considering large amounts of data, i.e. completely unstructured co-occurrence material of verbs. This outcome can be interpreted as a manifestation of the principle 'the more, the merrier'. On the one hand, preprocessed linguistic information, i.e. small units of linguistic information, were used for the construction of verb vectors (*informed distributional*). The resulting verb vectors of the latter model have a small dimensionality, however a lot of preprocessing steps were necessary. On the other hand, as much pure data as possible was included in the induction, leading to high dimensional verb vectors (*non informed* distributional). This model needs relatively few preprocessing steps, which means that the experiment's workflow is easy to design. When combining the two approaches, that is using noun clusters and aspectual features in a (restricted) distributional framework or left out the clustering of noun arguments in a (restricted) *informed* distributional framework, the quality of classification decreased.

Which model is preferable depends on the aims and preconditions of the analysis. When a very large set of verbs has to be classified, one might run into trouble with the very high dimensionalities of the verb vectors, making analysis slow, if not impossible. In that case the *informed* distributional method which requires preprocessing tools like parsers would be preferable. The choice of the method thus depends, in general, on the amount of data that have to be analyzed and on the availability of preprocessing tools for the language which is being studied.

The results of the present study fit well into theories of language acquisition. Naigles et al. (1992), Naigles et al. (1995) and Naigles et al. (1993) describe *frame compliance* as an essential strategy used by young children to interpret sentences. In order to classify verbs into classes, a task which is mastered quite late in the acquisition process (Wittek 2002, Richter and van Hout 2013), children need to build up knowledge about sentence types, for instance about transitivity properties of sentences, and this knowledge is utilized to learn verb classes (Brooks and Tomasello 1999, Brooks et al. 1999). At the beginning of the acquisition process knowledge of sentence types is sometimes presumed to be item based (Tomasello, 2000) meaning that it depends on specific verbs and their contexts, while other studies provide evidence of a linguistic maturation process: A constant development of linguistic knowledge over time. Our study provides evidence for Tomasello's item based approach. Starting with unstructured contexts of verbs, learners identify very early item based prototypical verb/arguments structures and subsequently in the course of the acquisition process they use context materials to come to analogies. When learners recognize a context, they classify a new verb according to that context which is the environment of a previous learned and eventually previously classified verb. In order to manage the task of classifying verbs, linguistic experience is required: The more language learners build up linguistic knowledge and manage to identify arguments of verbs and other verb dependent elements in the sentence, the more they build up the structural knowledge of sentences types which in turn is a prerequisite in order to induce verbs classes from their contexts.

In summary, the results of this study show that, for German, the Vendler + 1 typlogy can be inferred from context materials of verbs. We found that both classification frameworks that is, the informed distributional subcategorization-frame-based framework and the non informed window-based framework performed almost on a par. As future work, our experiments should be complemented by clustering experiments which means switching to unsupervised learning by induction. This would mean a step towards simulation of induction based language acquisition processes (see Bruner et al. 1956) on rule based learning of concepts).

REFERENCES

Bach, E. (1986). The algebra of events. *Linguistics and Philosophy*, 9:5–16.

Bohnet, B. (2010). Very high accuracy and fast dependency parsing is not a contradiction. In *Proceedings of the 23rd International Conference on Computational Linguistics*, COLING '10, pages 89–97, Stroudsburg, PA, USA. Association for Computational Linguistics.

Brooks, P. and Tomasello, M. (1999). How children constrain their argument structure constructions. *Language*, 75(4):720–738.

Brooks, P., Tomasello, M., Dodson, K., and Lewis, L. (1999). Young childrens overgeneralizations with fixed transitivity verbs. *Child Development*, 7(6):1325–1337.

Bruner, J., J., G. J., and A., A. G. (1956). *A Study of Thinking*. Wiley.

Dorr, B. J. and Jones, D. A. (1996). Role of word sense disambiguation in lexical acquisition: Predicting semantics from syntactic cues. In *16th International Conference on Computational Linguistics, Proceedings of the Conference, COLING 1996, Center for Sprogteknologi, Copenhagen, Denmark, August 5-9, 1996*, pages 322–327.

Dowty, D. (1979). *Word Meaning and Montague Grammar*. D. Reidel, Dordrecht.

Dowty, D. (1991). Thematic proto-roles and argument selection. *Language*, 67(3):547–619.

Faaß, G. and Eckert, K. (2013). Sdewac - a corpus of parsable sentences from the web. In *Proceedings of the International Conference of the German Society for Computational Linguistics and Language Technology*, Darmstadt,Germany.

Fernando, T. (2004). A finite-state approach to events in natural language semantics. *Journal of Logic and Computarion*, 14(1):79–92.

Friedrich, A. and Palmer, A. (2014). Automatic prediction of aspectual class of verbs in context. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 517–523, Baltimore, Maryland. Association for Computational Linguistics.

Goldberg, A. E. (1995). *Constructions: a construction grammar approach to argument structure*. University of Chicago Press, Chicago.

Gruender, S. (2008). An algorithm from adverbial aspect shift. In *Proceedings of the 22nd International Conference on Computer Linguistics (Coling08)*, pages 289–296.

Harris, Z. (1954). Distributional structure. *Word*, 10(23):146–162.

Hermes, J., Richter, M., and Neuefeind, C. (2015). Automatic induction of german aspectual verb classes in a distributional framework. In *International Conference of the German Society for Computational Linguistics and Language Technology. Proceedings of the Conference (GSCL 2015)*, pages 122–29.

Hermes, J. and Schwiebert, S. (2009). Classification of text processing components: The tesla role system. In *Advances in Data Analysis, Data Handling and Business Intelligence*, pages 285–294. Springer.

Jackendoff, R. (1972). *Semantic Interpretation in Generative Grammar*. MIT press, Cambridge MA.

Jackendoff, R. (1983). *Semantics and Cognition*. MIT press, Cambridge MA.

Joachims, T. (1998). Text categorization with support vector machines: Learning with many relevant features.

Joanis, E., Stevenson, S., and James, D. (2008). A general feature space for automatic verb classification. *Natural Language Engineering*, 14(3):337–367.

Klavans, J. L. and Chodorow, M. (1992). Degrees of stativity: The lexical representation of verb aspect. In *Proceedings of the 14th Conference on Computational Linguistics - Volume 4*, COLING '92, pages 1126–1131, Stroudsburg, PA, USA. Association for Computational Linguistics.

Klein, W. (2009). How time is encoded. In *The expression of time*, pages 39–82. Mouton de Gruyter, Berlin, New York.

Krifka, M. (1989). Nominal reference, temporal constitution, and quantification in event semantics. In Bartsch, R., van Benthem, J., and van Emde Boas, P., editors, *Semantics and Contextual Expressions*, pages 75–115. Foris.

Lakoff, G. (1970). *Irregularity in Syntax*. Rinehart and Winstons, New York.

Landauer, T. K. and Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological review*, 104(2):211–240.

Levin, B. and Rapoport, T. R. (1992). Lexical subordination. In *Papers from the 24th Regional Meeting of the Chicago Linguistic Society*, pages 275–289.

Levy, J. P. and Bullinaria, J. A. (2001). Learning lexical properties from word usage patterns: Which context words should be used? In *Connectionist Models of Learning Development and Evolution: Proceedings of the 6 th Neural Computation and Psychology Works hop*, pages 273–282. Springer.

Lowe, W. and McDonald, S. (2000). The direct route: mediated priming in semantic space. In *Proceedings of the annual Conference of the Cognitive Science Society (CogSci-2000)*.

Lund, K. and Burgess, C. (1996). Hyperspace analogue to language (HAL): A general model semantic representation. In *Brain and Cognition*, volume 30, page 265.

Merlo, P. and Stevenson, S. (2001). Automatic verb classification based on statistical distributions of argument structure. *Comput. Linguist.*, 27(3):373–408.

Naess, A., editor (2007). *Prototypical Tranisitivity*. Typological Studies in Language 72. Benjamins, Amsterdam.

Naigles, L., Fowler, A., and Helm, A. (1995). Syntactic bootstrapping from start to finish with special reference to Down syndrome. *Beyond names for things: Young childrens acquisition of verbs*, pages 299–330.

Naigles, L., Gleitman, L., and Gleitman, H. (1993). Children acquire word meaning components from syntactic evidence. *Language and cognition: A developmental perspective. Norwood, NJ: Ablex*, 5:104–140.

Naigles, L. G., Fowler, A., and Helm, A. (1992). Developmental shifts in the construction of verb meanings. *Cognitive Development*, 7(4):403–427.

Pantel, P. (2005). Inducing ontological co-occurrence vectors. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, ACL '05, pages 125–132, Stroudsburg, PA, USA. Association for Computational Linguistics.

Pulman, S. G. (1997). Aspectual shift as type coercion. In *Transactions of the Philological Society 95(2)*, pages 279–317.

Pustejovsky, J. (1991). The synatax of event structure. *Cognition*, 41:47–81.

Quine, W. v. O. (1960). *Word and Object*. MIT press, Cambridge MA.

Richter, M. and Hermes, J. (2015). Classification of german verbs using nouns in argument positions and aspectual features. *NetWordS 2015 Word Knowledge and Word Usage*, pages 177–181.

Richter, M. and van Hout, R. (2013). Interpreting resultative sentences in German: Stages in l1 acquisition. *Linguistics*, 51(1):117–144.

Richter, M. and van Hout, R. (2016). A classification of german verbs using empirical language data and concepts of vendler and dowty. *Sprache und Datenverarbeitung - International Journal for language data processing*, 38(1-2):81–117.

Rothstein, S. (2004). *Structuring Events: A Study in the Semantics of Aspect*. Explorations in Semantics. Wiley.

Rubenstein, H. and Goodenough, J. B. (1965). Contextual correlates of synonymy. *Commun. ACM*, 8(10):627–633.

Schulte im Walde, S. (2003). Experiments on the choice of features for learning verb classes. In *Proceedings of the Tenth Conference on European Chapter of the Association for Computational Linguistics - Volume 1*, EACL '03, pages 315–322, Stroudsburg, PA, USA. Association for Computational Linguistics.

Schulte im Walde, S. (2006). Experiments on the automatic induction of german semantic verb classes. *Computational Linguistics*, 32(2):159–194.

Schulte im Walde, S. and Brew, C. (2002). Inducing german semantic verb classes from purely syntactic subcategorisation information. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, ACL '02, pages 223–230, Stroudsburg, PA, USA. Association for Computational Linguistics.

Schumacher, H. (1986). *Verben in Feldern: Valenzwörterbuch zur Syntax und Semantik deutscher Verben*, volume 1. Walter de Gruyter.

Schütze, H. and Pedersen, J. (1993). A vector model for syntagmatic and paradigmatic relatedness. pages 104–113.

Siegel, E. V. (1997). Learning methods for combining linguistic indicators to classify verb. In *Proceedings of the 2nd Conference on Empirical Methods in Natural Language Processing, EMNLP*, Brown University, Providence, RI. cmp-lg/9707015.

Siegel, E. V. and McKeown, K. R. (2000). Learning methods to combine linguistic indicators: Improving aspectual classification and revealing linguistic insights. *Comput. Linguist.*, 26(4):595–628.

Smith, C. (1991). *The Parameter of Aspect*. Kluwer, Dordrecht.

Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition*, 74(3):209 – 253.

Turney, P. D. and Pantel, P. (2010). From frequency to meaning: Vector space models of semantics. *J. Artif. Int. Res.*, 37(1):141–188.

Vendler, Z. (1967). *Linguistics in philosophy*. G - Reference, Information and Interdisciplinary Subjects Series. Cornell University Press.

Verkuyl, H. J. (2005). Aspectual composition: Surveying the ingredients. pages 19–39.

Vlachos, A., Korhonen, A., and Ghahramani, Z. (2009). Unsupervised and constrained dirichlet process mixture models for verb clustering. In *Proceedings of the Workshop on Geometrical Models of Natural Language Semantics*, GEMS '09, pages 74–82, Stroudsburg, PA, USA. Association for Computational Linguistics.

Wittek, A. (2002). *Learning the meaning of change-of-state verbs: A case study of German child language*, volume 17. Walter de Gruyter.

Wright, G. H. v. (1963). *Norm and Action*. Routledge and Kegan Paul, London.

Zarcone, A. and Lenci, A. (2008). Computational models of event type classification in context. In *Proceedings of the 6th International Conference on Language Resources and Evaluation*, pages 1232–1238, Marrakech, Morocco.