# Annotation and Automatic Classification of Aspectual Categories

**Markus Egg** **Helena Prepens** **Will Roberts**
Humboldt-Universität zu Berlin
Berlin, Germany
{markus.egg,helena.prepens,will.roberts}@hu-berlin.de

## Abstract

We present the first annotated resource for the aspectual classification of German verb tokens in their clausal context. We use aspectual features compatible with the plurality of aspectual classifications in previous work and treat aspectual ambiguity systematically. We evaluate our corpus by using it to train supervised classifiers to automatically assign aspectual categories to verbs in context, permitting favourable comparisons to previous work.

## 1 Introduction

The universal linguistic category of aspect describes how a verb or a verbal projection (including sentences, 'predicates' for short) characterises the temporal course of a state of affairs or 'eventuality'. Such information is relevant for tasks that extract temporal information from texts, such as information extraction, question answering, and document summarisation (Costa and Branco, 2012). Further tasks in which aspectual information plays a crucial role include computational semantic analysis (Caselli and Quochi, 2007), zoning (analysing the argumentative and rhetorical structure of texts; Baiamonte et al. 2016), and the analysis of specific textual elements, e.g., captions (Alikhani and Stone, 2019).

Aspect must also be considered in event annotation (Pustejovsky et al., 2010; Bittar et al., 2011; Caselli et al., 2011). Aspect is a universal semantic category; thus, the same aspectual patterns reappear across languages.

We created the first resource of German verbs annotated for aspectual class in context. We use aspectual features compatible with various different previously published aspectual classifications, and model the pervasive phenomenon of aspectual ambiguity. We evaluate the resource by using it in supervised aspectual classifiers for verbs in context.

## 2 Aspectual classes and ambiguity

Aspectual classes are established by feature dichotomies (Vendler, 1967; Moens and Steedman, 1988; Egg, 2005). First, *stative* predicates describe purely static situations (e.g., *be happy* or *love*); *dynamic* ones introduce eventualities with development (e.g., continuous change of place in *move*).

Dynamic predicates can be either *unbounded* (introduce eventualities without inherent boundaries, e.g., *move* or *play the piano*), or *bounded* (e.g., *run a mile* or *build a house*). Bounded predicates (also called 'telic') have four subgroups that are cross-classified by the features *change* - *no change* and *punctual* - *extended*: The first pair distinguishes predicates that express an explicit change of state (e.g., *leave* as change from being present to being away) from predicates that do not (e.g., *play a sonata*).[1] The second pair distinguishes e.g. the no-change predicates *cough* and *play a sonata* or the change predicates *explode* and *build a house*.

The *punctual* - *extended* distinction is gradual (while the others are binary). This will tend to aggravate both the annotation and the automatic classification of aspect.

These features define six aspectual classes: Only dynamic predicates can be bounded or not, and only bounded predicates can be extended or punctual, and introduce an explicit change of state or none. Such aspectual properties are sometimes called 'lexical aspect' or 'aktionsart' to distinguish them from 'morphological aspect', e.g., the progressive or perfective/imperfective markers in Slavic languages.

Also, the aspectual class of a verb may be influenced obligatorily by an argument, in particular, by an 'incremental theme' (Dowty, 1991; Krifka,

[1] This pair appears as 'culminated' and 'non-culminated' in Siegel and McKeown (2000) and as '±conseq[uence]' in Moens and Steedman (1988); in the latter it partitions dynamic predicates.



*Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3335–3341
Florence, Italy, July 28 - August 2, 2019. 

1992) like in *eat an apple* (bounded) vs. *eat apples* (unbounded), or by arguments that specify the path inherent in movement verbs, compare *run a mile* (bounded) vs. *run some laps* (unbounded).[2] Our corpus contains a substantial number of these cases. Their classification in our corpus reflects the aspectual influence of these arguments.

Operators like the progressive and specific kinds of adverbials may exert an aspectual influence on the predicates which they take as arguments. For instance, durative adverbials map unbounded predicates onto extended no-change predicates, and the progressive maps dynamic predicates of all kinds onto stative ones. Consequently, the aspectual class of a full clause or sentence may differ from the one of its main verb (plus its arguments); thus, annotating aspect at the clause or sentence level differs from our annotation task.

The aspectual value of a predicate can also be modified in order to fit aspectual selection restrictions of an operator, which is known as aspectual *coercion* (Moens and Steedman, 1988). For instance, if *plötzlich* 'suddenly', which requires a punctual argument, is combined with an unbounded predicate like *laufen* 'walk', this induces an inchoative reinterpretation of the verb in the sense of 'begin to walk'. Our annotation records the aspectual class of the argument before any coercion.

Classifying verbs aspectually must be able to handle the (often systematic) aspectual ambiguity on the token level (5% of the tokens in our corpus), including (1) and (2).[3]

Ambiguity can arise in that a token has no value for a feature, e.g., *abtrennen* 'detach' in (1) for 'punctual-extended', because duration is unclear:

(1) *wenn der Kunde die Karte abtrennt*
'when the client detaches the card'

Other cases have two distinct readings, e.g., many verbs in the semantic field of communication have a stative and a change-of-state reading. E.g., in (2), *zeigen* 'show' can indicate a stative property ('be more successful') or a change of state ('obtain better results'):

(2) *diese Firmen zeigen bessere Ergebnisse*
'these companies show better results'

[2]Incrementality is given a wider definition in Tenny (1992), which goes beyond the phenomena relevant for our annotation initiative.

[3]Croft et al. (2016) also emphasise the importance of aspectual ambiguity in their work on aspectual annotation.

Systematic ambiguity furthermore emerges for so-called 'degree achievements' like *den Weg kehren* 'sweep the path' (Kennedy and Levin, 2008), which systematically have an unbounded reading (continuous development, here, towards cleanliness) and an extended change reading (here, crossing a threshold of cleanliness). We found many instances of these in our corpus as well.

The great level of detail of our classification is novel and addresses the problem that—beyond distinguishing stative predicates—previous work on aspectual classification disagrees widely. Our classification is related to previous ones in Table 1. It can easily be transformed into other classifications by collapsing classes. E.g., uniting the 'unbounded' and the 'extended/no change' class yields Moens and Steedman's system; ignoring the 'change/no change' feature returns Vendler's classes. In this way, our classification is not tied to the limitations imposed by specific aspectual theories. This flexibility is an advantage over preceding annotation initiatives, which typically presuppose a specific aspectual classification.

This flexibility also means that our classification lends itself to tasks of different granularity. As we will show in Section 5, it can be used for coarse two-way distinctions, e.g., between stative and non-stative predicates, as well as for very fine-grained classification tasks.

## 3 Related work

Siegel and McKeown (2000) annotated the main verbs of 1,478 and 615 parsed clauses from medical discharge summaries and novels, respectively, with the classes of Moens and Steedman (1988). These classes are determined lexically (and may be influenced by obligatorily aspectually relevant arguments as discussed above), which they call 'fundamental aspectual class'. Each verb instance is assigned a single aspectual class, which neglects aspectual ambiguity on the token level.

They trained supervised classifiers, using 'linguistic indicators' for aspectual classes as features, e.g., the perfect, the progressive or durative adverbials like *for two hours*. Co-occurrences of these indicators with the verbs were counted in large parsed corpora (supersets of the annotated corpora).

For the first corpus, they distinguished stative vs. dynamic verbs with 93.9% accuracy. The second corpus was used for distinguishing 'culmination'

| Our classes | | | Vendler (1967) | Moens and Steedman (1988) | Egg (2005) |
|---|---|---|---|---|---|
| stative | | | state | state | stative predicate |
| dynamic | unbounded | | activity | process | process predicate |
| | bounded | extended/ no change | accomplishment | | intergressive predicate |
| | | extended/ change | | culminated process | change predicate |
| | | punctual/ no change | achievement | point | intergressive predicate |
| | | punctual/ change | | culmination | change predicate |

Table 1: Comparison of aspectual classes in previous work and our features.

and 'non-culmination'[4] with up to 74% accuracy.

Friedrich and Palmer (2014) took into account aspectual ambiguity of verb tokens. In their first corpus of 6,161 clauses (from MASC, Ide et al. 2008), verb tokens are classified as stative, dynamic, or ambiguous. A second set of 2,667 clauses from the Brown Corpus focused on verbs that are ambiguous for stativity, by sampling sentences containing 20 frequent verbs that had both stative and dynamic senses, and annotating as before.

They trained classifiers on these data, using the type-based indicators of Siegel and McKeown (2000), and vectors from a word space model. To characterise individual verb tokens, they included contextual features like POS tag, tense, voice, and WordNet classes of the verb arguments.

Experiment 1 tested performance on verbs during training. The classifier was trained on the first data set, using 10-fold cross validation. Accuracy reached 84.1%, but no feature set statistically outperformed the naïve strategy of memorising each verb's most likely aspect class. Experiment 2 tested the classifier on unseen verbs, by stratifying the cross validation folds by verb lemma.

Falk and Martin (2016) annotated 1,200 French verb tokens, modelling aspectual ambiguity directly in their aspectual classification; this is based on Vendler classes but adds four ambiguity classes, e.g., for verbs ambiguous between 'state' and 'activity' like *penser* 'think'. Also, there is a class of change-of-state verbs unspecified for punctuality, and two classes of degree achievements (with and without preference for the change reading).

We see two problems for their approach. First, aspectual ambiguity is a property of individual verbs, hence, no additional classes are needed. Second, their classification is not general enough, e.g., for *zeigen* 'show', which can be stative or change of state. Since we can handle aspectual ambiguity of verbs, we can replicate their classification (up to the two classes of degree achievements).

Falk and Martin train a classifier on their annotation, which reaches 67% accuracy on a three-way split between unbounded and change-of-state verbs, and those that fall in between the two groups.

Other resources target aspectual classification at the sentence or clause level. Mathew and Katz (2009) annotate 1,816 Penn Treebank sentences with dynamic verbs as episodic (describing actual eventualities) or habitual (referring to habits, a subclass of stative predicates). Friedrich and Pinkal (2015) annotate 10,355 Wikipedia clauses as stative (and non-habitual), episodic, or habitual. Zarcone and Lenci (2008) annotate 3,129 sentences of the Italian Syntactic-Semantic Treebank (Montemagni et al., 2003) for Vendler's (1967) classes. The corpora of Palmer et al. (2007) (6,065 clauses from Brown Corpus and MUC6 dataset) and Friedrich et al. (2016) (45,331 clauses from Brown Corpus, MASC, and Wikipedia) annotate clauses for 'situation entities', which include, but go beyond aspectual classes.

## 4 The resource

### 4.1 Annotation

We compiled a corpus of German verb tokens in their clausal contexts from the SdeWaC corpus (Faaß and Eckart, 2013) and parsed them with `mate-tools` (Bohnet et al., 2013); the aspectual annotation used our six-fold classification.

The corpus has three parts. Part A (3000 clauses) is based on a verb sample balanced for verb frequency. We took 60 verbs drawn at random for annotation, 20 each from the classes with high (65 verbs with counts of $> 10^5$ in SdeWaC), medium (602 verbs, counts $> 10^4$), and low frequency (ca. 2100 verbs, counts $> 10^3$). For each of these 60 verbs, we drew 50 sentences with that verb from SdeWaC. Part B (900 clauses) repeated this procedure without using a verb sample, including 300 sentences each with verbs of high, medium, and low frequency. Part C (300 clauses) has 150 sen-

[4]In their adaption of Moens and Steedman's terms, this emerges as $\pm$conseq, i.e., change verbs vs. the union of unbounded and no-change verbs in our terms.

tences with punctual (e.g., *cough*) and 150 sentences with extended no-change verbs (e.g., *run a mile*), as these are systematically under-represented in the other two parts.

Our annotation tool allowed only feasible combinations of the aspectual features. Annotation guidelines explained the aspectual features and provided tests for assigning values to them. E.g., stative predicates like *glücklich sein* 'be happy' do not combine with adverbials expressing intentionality:

(3) *Max ist freiwillig glücklich.
'Max is voluntarily happy.'

Similar tests guide the annotation of the other three feature pairs, e.g., only unbounded predicates combine with durative adverbials. The guidelines also explain the phenomenon of obligatory aspectual influence by verbal arguments. The annotation paid consideration to metaphorical usages; however, our anecdotal experience suggests that verbal metaphor tends to preserve aspectual class. Disagreements between annotators were subsequently adjudicated.

We annotate aspectual ambiguity on the token level; categories are tagged 'unknown' when a verb has no value for a specific feature like in (1). Cases like (2) get two separate full annotations.

### 4.2 Annotator agreement

We evaluated inter-annotator agreement after training the annotators and having them annotate ca. 2,200 clauses. Both annotators annotated 248 unseen clauses; nine of these were excluded as invalid. Table 2 shows agreement on the remaining 239 clauses before adjudication. Its first four rows display agreement on specific categories. 'Class' lists agreement on our six-way aspectual classification; the last line shows agreement without the problematic *punctual - extended* category. In Landis and Koch's (1977) terms, agreement on the first three feature pairs is substantial, and fair on the fourth. Agreement on the overall classification is moderate, rising to substantial without the *extended - punctual* distinction. The results confirm our scepticism about the usefulness of this distinction, because deciding whether a predicate is punctual or extended has frequently proven to be extremely hard.

Agreement on the stative/dynamic features is like in Friedrich and Palmer (2014). For the annotation of the first corpus by two annotators, their Cohen's $\kappa$ was 0.7, for their second corpus, 0.6.

| | |
|---|---|
| Stative | $\kappa$ = 0.746 |
| Bounded | 0.735 |
| Change of state | 0.758 |
| Extended | 0.292 |
| Class | 0.548 |
| Class w/o extended | 0.651 |

Table 2: Agreement on the aspectual class annotation.

| Task | Baseline | Classifier |
|---|---|---|
| 6-way | 0.295 | 0.712 |
| 4-way | 0.445 | 0.785 |
| Vendlerian | 0.368 | 0.730 |
| Stativity | 0.719 | 0.877 |
| Stative/Unbounded - Change | 0.443 | 0.817 |
| Culmination | 0.618 | 0.856 |

Table 3: Classifier accuracy on aspect labelling tasks.

## 5 Evaluation

To test the validity and utility of our annotated corpus, we trained supervised classifiers on the dataset. The fine granularity of our classification allows us to define several tasks. We use a logistic regression classifier with L2 regularisation ($\lambda^{-1} = 2.78$) and employ sentence-level features derived from the automatic parse of the clause: the verb lemma; POS; tense; use of the passive; a word embedding for the verb[5]; a bag of words to represent the sentence context; the lemmas of the verb's grammatical dependants; the GermaNet (Hamp and Feldweg, 1997) semantic class for the verb and its subject and object; the adverb modifying the verb, if available; and the subcategorisation frame of the verb, given by the rule-based classifier of Roberts et al. (2014). Training and testing use 10-fold cross validation. Table 3 shows accuracies and baselines, which always predict the training set's most frequent label.

The first classifier predicts the full 6-way classification of the annotation. To handle aspectual ambiguity, each verb instance maps to an *ambiguity class* consisting of one or more aspectual class labels. The distribution of ambiguity classes is long-tailed, and we discard data points with labels

[5]The embedding was built using `word2vec` on the lemmatised SdeWaC with the parameters recommended by Baroni et al. (2014): 400-dimensional CBOW vectors, window size 5, subsampling with $t = 1e^{-5}$, negative sampling with 10 samples.

less frequent than a threshold set to 10. In the case of the 6-way classifier, this removes 40 data points, and results in 10 ambiguity classes in total.

The second and the third classifier test our expectation that our resource is useful for less fine-grained aspectual classifications, too. The second classifier disregards the punctual-extended feature (collapsing the two change and the two non-change classes), i.e., follows Egg's (2005) classification. 18 data points are dropped, leaving 7 possible labels. The third classifier disregards the change/no-change distinction, corresponding to Vendler's (1967) classes. 26 data points are dropped, resulting in 7 possible labels.

These three models achieve similar error rate reductions over the baseline of about 60%. The 4-way classifier, which ignores the extended-punctual distinction, outperforms the Vendlerian classifier, which includes it; this suggests that the extended-punctual distinction is more difficult to identify and to model.

The following three classifiers are motivated by classifications in prior work. The fourth one ('Stativity') predicts whether a token is stative (1077), dynamic (2915), or ambiguous in context (60). This corresponds to Experiment 1 of Friedrich and Palmer (2014, p.520). Their baseline of 0.725 and their classifier accuracy of 0.841 are both similar to our results. We can also replicate their Experiment 2 by stratifying the cross validation folds by verb lemma, showing the performance of the classifier on unseen verbs. Our accuracy here is 0.811, almost identical to their reported 0.819.

The fifth classifier approximates the classification task of Falk and Martin (2016), distinguishing 'atelic' (1707, our stative and unbounded verbs), 'telic' (1794, our change of state verbs), and 'variable telicity' (551, our no-change verbs, plus verbs that are ambiguous between the other two categories). Our results exceed theirs (0.675 accuracy with a 0.484 baseline).

The sixth classifier predicts whether a verb token is 'culminated' or 'non-culminated', corresponding to the task of Experiment 2 of Siegel and McKeown (2000, Table 16, p.618). Culminated verbs (1834) are our change verbs, and non-culminated verbs (1077), the union of our unbounded and no-change verbs; 59 verbs are ambiguous in context. Siegel and McKeown report a baseline of 0.633, similar to ours, and their classifier achieves 0.740, which we outperform.

These experiments support several conclusions. First, we have shown our resource can be used to build machine learning classifiers of high quality, speaking to the validity of our corpus. While we can only draw indirect comparisons to previous work in English and French, the accuracies achieved by our classifiers suggest that we go beyond the state of the art in our work.

Second, our resource has proven to be very flexible in that it can be broken down in different ways to capture different aspectual distinctions, which is very welcome considering the wide range of aspectual classifications.

Finally, the better performance of the 4-way classifier compared to the Venderian classifier, combined with the $\kappa$ value for the extended-punctual distinction (Table 2) seems to indicate that both machines and human annotators find it hard to judge the length of time of a reported event. As hypothesised, this distinction has proved to be the most difficult of our four aspectual features; this finding accords with Zarcone and Lenci (2008), who report that durativity is the hardest aspectual feature to classify.

## 6 Conclusion and future work

We present the first aspectually annotated resource for German verb tokens. We report substantial inter-annotator agreement, and validate our resource by training automatic aspectual classifiers, permitting favourable comparisons to prior work. The annotated corpus, the source code for the annotation tool, and the annotation guidelines are available at `https://github.com/wroberts/annotator`.

Future work will offer a more principled account of aspectual classification for specific verb classes, among them speech act and communication verbs (e.g., *promise* or *call*) that occur frequently in corpora but have hitherto been neglected in aspectual analyses.

On a more general scale, we envisage examining the interplay of verb class (e.g., the classes of Levin 1993), verb sense, and aspectual class, with the purpose of estimating the influence of the sentential context on the aspectual value of the predicate. We also intend to develop a more principled treatment for the aspectual classification of metaphors, which are frequent in other corpora.

## References


Malihe Alikhani and Matthew Stone. 2019. "Caption" as a coherence relation: evidence and implications. In *Second Workshop on Shortcomings in Vision and Language (SiVL)*.

Daniela Baiamonte, Tommaso Caselli, and Irina Prodanof. 2016. Annotating content zones in news articles. In *Proceedings of Third Italian Conference on Computational Linguistics (CLiC-it 2016)*.

Marco Baroni, Georgiana Dinu, and Germán Kruszewski. 2014. Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 238–247, Baltimore, MD. Association for Computational Linguistics.

André Bittar, Pascal Amsili, Pascal Denis, and Laurence Danlos. 2011. French TimeBank: An ISO-TimeML annotated reference corpus. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 130–134, Portland, Oregon, USA. Association for Computational Linguistics.

Bernd Bohnet, Joakim Nivre, Igor Boguslavsky, Richárd Farkas, Filip Ginter, and Jan Hajič. 2013. Joint morphological and syntactic analysis for richly inflected languages. *Transactions of the Association for Computational Linguistics*, 1:415–428.

Tommaso Caselli, Valentina Bartalesi Lenzi, Rachele Sprugnoli, Emanuele Pianta, and Irina Prodanof. 2011. Annotating events, temporal expressions and relations in Italian: The It-Timeml experience for the Ita-TimeBank. In *Proceedings of the 5th Linguistic Annotation Workshop*, pages 143–151, Portland, Oregon, USA. Association for Computational Linguistics.

Tommaso Caselli and Valeria Quochi. 2007. Inferring the semantics of temporal prepositions in italian. In *Proceedings of the Fourth ACL-SIGSEM Workshop on Prepositions*, pages 38–44. Association for Computational Linguistics.

Francisco Costa and António Branco. 2012. Aspectual type and temporal relation classification. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 266–275. Association for Computational Linguistics.

William Croft, Pavlina Peskova, and Michael Regan. 2016. Annotation of causal and aspectual structure of events in RED: A preliminary report. In *Proceedings of the Fourth Workshop on Events*, pages 8–17, San Diego, CA. Association for Computational Linguistics.

David Dowty. 1991. Thematic proto-roles and argument selection. *Language*, 67:547–619.

Markus Egg. 2005. *Flexible semantics for reinterpretation phenomena*. CSLI Publications, Stanford.

Gertrud Faaß and Kerstin Eckart. 2013. SdeWaC - A corpus of parsable sentences from the Web. In *Language processing and knowledge in the Web*, pages 61–68. Springer, Berlin, Heidelberg.

Ingrid Falk and Fabienne Martin. 2016. Automatic identification of aspectual classes across verbal readings. In *Proceedings of the Fifth Joint Conference on Lexical and Computational Semantics*, pages 12–22, Berlin, Germany. Association for Computational Linguistics.

Annemarie Friedrich and Alexis Palmer. 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, MD. Association for Computational Linguistics.

Annemarie Friedrich, Alexis Palmer, and Manfred Pinkal. 2016. Situation entity types: Automatic classification of clause-level aspect. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1757–1768, Berlin, Germany. Association for Computational Linguistics.

Annemarie Friedrich and Manfred Pinkal. 2015. Automatic recognition of habituals: A three-way classification of clausal aspect. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 2471–2481, Lisbon, Portugal. Association for Computational Linguistics.

Birgit Hamp and Helmut Feldweg. 1997. GermaNet: A lexical-semantic net for German. In *Proceedings of ACL Workshop on Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications*, pages 9–15.

Nancy Ide, Collin Baker, Christiane Fellbaum, Charles Fillmore, and Rebecca Passonneau. 2008. Masc: The manually annotated sub-corpus of American English. In *6th International Conference on Language Resources and Evaluation, LREC 2008*, pages 2455–2460. European Language Resources Association (ELRA).

Chris Kennedy and Beth Levin. 2008. Measure of change: The adjectival core of degree achievements. In Louise McNally and Chris Kennedy, editors, *Adjectives and adverbs: Syntax, semantics and discourse*, pages 156–182. Oxford University Press, Oxford.

Manfred Krifka. 1992. Thematic roles as links between nominal reference and temporal constitution. In Ivan Sag and Anna Sabolcsi, editors, *Lexical matters*, pages 29–53. CSLI, Stanford.

J. Richard Landis and Gary G. Koch. 1977. The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174.

Beth Levin. 1993. *English verb classes and their alternation: A preliminary investigation*. University of Chicago Press, Chicago.

Thomas Mathew and Graham Katz. 2009. Supervised categorization for habitual versus episodic sentences. In *Sixth Midwest Computational Lingustics Colloquium*, Bloomington. Indiana University.

Marc Moens and Mark Steedman. 1988. Temporal ontology and temporal reference. *Computational Linguistics*, 14(2):15–28.

Simonetta Montemagni, Francesco Barsotti, Marco Battista, Nicoletta Calzolari, Ornella Corazzari, Alessandro Lenci, Antonio Zampolli, Francesca Fanciulli, Maria Massetani, Remo Raffaelli, et al. 2003. Building the Italian syntactic-semantic treebank. In *Treebanks*, pages 189–210. Springer.

Alexis Palmer, Elias Ponvert, Jason Baldridge, and Carlota Smith. 2007. A sequencing model for situation entity classification. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 896–903, Prague, Czech Republic. Association for Computational Linguistics.

James Pustejovsky, Kiyong Lee, Harry Bunt, and Laurent Romary. 2010. ISO-TimeML: An international standard for semantic annotation. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta. European Language Resources Association (ELRA).

Will Roberts, Markus Egg, and Valia Kordoni. 2014. Subcategorisation acquisition from raw text for a free word-order language. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, pages 298–307, Gothenburg, Sweden. Association for Computational Linguistics.

Eric V. Siegel and Kathleen R. McKeown. 2000. Learning methods to combine linguistic indicators: Improving aspectual classification and revealing linguistic insights. *Computational Linguistics*, 26(4):595–628.

Carol Tenny. 1992. The aspectual interface hypothesis. In Ivan Sag and Anna Sabolcsi, editors, *Lexical matters*, pages 1–27. CSLI, Stanford.

Zeno Vendler. 1967. Verbs and times. In Z. Vendler, editor, *Linguistics in philosophy*, pages 97–121. Cornell University Press, New York.

Alessandra Zarcone and Alessandro Lenci. 2008. Computational models for event type classification in context. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, pages 1232–1238, Marrakech, Morocco. European Language Resources Association (ELRA).