Identifying and Quantifying Aspectual Ambiguity with LMs

Reversing the NLP Pipeline

Bachelor Thesis

Date, Year

Samuel Innes

innes@cl.uni-heidelberg.de

Institut für Computerlinguistik

Ruprecht-Karls-Universität Heidelberg

Supervisor Prof. Dr. Michael Herweg
Reviewer Prof. Dr. Katja Markert

Abstract

Abstract in English

Zusammenfassung

Zusammenfasssung auf Deutsch

Contents

List of Figures

List of Tables

1 Related Work

Despite significant efforts to model lexical and grammatical aspects in computational linguistics, this area has received comparatively very little attention from the natural language processing (NLP)[1] part of the field [Friedrich et al., 2023]. This is presumably due to the fact that, on the one hand, it is a high-level semantic task, whose relevance to downstream applications is perhaps not as immediately obvious as other similarly complex tasks, and on the other hand, it is a complex phenomenon which is sometimes hard to capture accurately and faithfully. However, there have been some works in recent years looking at this area and studying how well current models deal with the phenomenon.

1.1 Aspect classification and ambiguity

There have been numerous works in computational linguistics over the last 20 years, using statistical and rule-based techniques to look at aspect. Some of the more prominent examples are listed here:

- Siegel and McKeown [2000] develop a rules-based aspect classification of verbs using certain linguistic indicators as features in a corpus to train decision tree, genetic programming and logistic regression models. The aspect class they aim to predict is an inherent class of a verb (i.e. lexical aspect), and hence they do not take into account the contextual reading, which often has an effect on the aspect of the verb considered. This is one of the issues which makes the aforementioned distinction between lexical and grammatical aspect so difficult in practice. The aspect classification scheme they use was that of Moens and Steedman [1988], a 5-way class distinction building on Vendler [1957]. Interestingly they use their results to gain linguistic insights, such as the fact that stative verbs are characterised by their high verb frequency.
- Friedrich and Palmer [2014] train a 3-way random forest classifier (predicting DYNAMIC, STATIVE or BOTH), crucially for verbs *in context*. They note that there are many cases where the verbs themselves have an ambiguous aspectual reading which is resolved by the

1 As opposed to those with more emphasis on the *linguistics* part of computational linguistics.

context, but that equally there are situations where this ambiguity persists. However, they do not go on to further investigate this fact or the situations where this occurs, or to look at other types of aspectual ambiguity.

- Egg et al. [2019] train several logistic regression classifiers to predict the aspectual class of German verbs in context using a more fine-grained class system, achieving 71.2% accuracy on their 6-way classification. They note the phenomenon of *coercion* and aim to annotate the aspectual class of the element before coercion (therefore in these cases excluding this coercing context). They also stress the importance of dealing with aspectual ambiguity and give the example of the systematic ambiguity of so-called *degree achievements* (from Kennedy and Levin [2008]) such as "den Weg kehren" (Eng. *sweep the path*), which can have an unbounded reading or an extended change reading.

- Croft et al. [2016] propose an annotation scheme for the aspectual structure of events, intended to be integrated into a larger event description framework: Richer Event Descriptions (RED) [Styler et al., 2014]. In doing so, they group together aspectual classes that often get mixed up with each other and use these pairings to create a semantic map (see 1). These more coarse-grained classes are then used for a second level of annotation, where (if possible) these ambiguities are resolved into one of the two finer-grained categories in the pair. However, this is a different type of ambiguity as the one dealt with in this study, where the focus is less on semantic similarity (and the ensuing annotator uncertainty), but more on *true* ambiguity (i.e. where several interpretations are equally possible and valid).

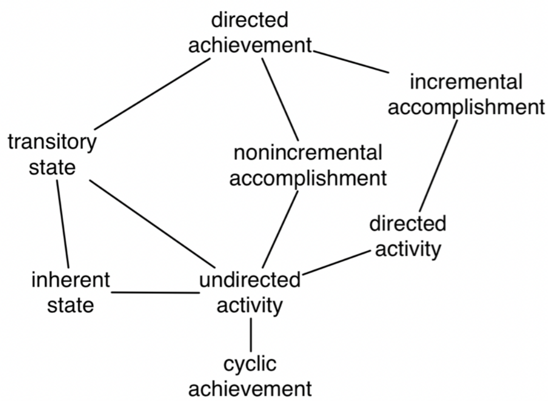

Figure 1: Croft et al.'s semantic map of common aspectual ambiguities Croft et al. [2016]

- Chen et al. [2021] build a rules-based system (or what they call "annotation tool") to predict UMR aspect class, by converting the aspect annotation guidelines into computable steps. Out of 235 gold events distributed across four gold files, the model missed 56 events, successfully identifying 179, and among these 179 identified events, the model accurately labeled 112, resulting in an accuracy rate of 62.57%. This is the only work so far to my knowledge using the same aspect classification schema as this study, UMR (see **??**).

- One other particularly interesting example is Friedrich and Gateva [2017], who use an English-Czech parallel corpus to leverage the fact mentioned in **??** that Slavic languages such as Czech have two forms of each verb, each assigned to a different aspectual reading depending on the context. This makes it possible to extract aspectual information from a Czech translation of an English sentence, assuming an accurate translation, to use as further training data for their linear regression model. They find that incorporating this "silver-standard" data into their training leads to significant increase in the F_1 score, especially when predicting ATELIC examples.

1.1.1 (L)LMs and aspect

There have also been attempts to use more modern NLP techniques such as language models for aspect classification and studies probing their aspectual knowledge.

- Metheniti et al. [2022] carry out a range of experiments with several transformer models (such as BERT, RoBERTa and XLNet) to try and understand to what extent they encode aspectual knowledge. They fine-tune the models on separate classification tasks for telicity and duration, as well as carrying out a qualitative analysis, and come to the conclusion that transformer models *do* understand aspectual phenomena to a sufficient extent, and they find that this knowledge relatively spread across the layers. However, they do not further probe how this information is encoded or which features the model uses to make its decision.
- Conversely, however, Katinskaia and Yangarber [2024] find that in Russian BERT models [Kuratov and Arkhipov, 2019] aspect information is stored in layers higher up. Katinskaia and Yangarber [2024] also find that their model has high uncertainty when predicting aspect in "alternative contexts" (i.e. where both the imperfective and the perfective verb is possible, depending on the implicit context). On the one hand, this is an encouraging find, as it suggests a correlation between machine behaviour and human behaviour, but on the other hand it is not particularly surprising, since "alternative contexts" are by definition contexts

with no tokens containing aspectual information, which are usually used by the model to make a decision.

- Rezaee et al. [2021] look at using generative and discriminative transformers for situation entity classification, which is, however, a slightly different task (see 1.3).

Friedrich et al. [2023] notes however that there have still been very little (if any) work looking at the probing or explainability of LMs for this task.

1.2 Available datasets

Although not small in number or size, the available datasets are relatively sparse since there is a very wide range of aspectual phenomena (/schemata) which can be annotated for. For example while some are annotated for aspectual phenomena such as habituality [Mathew and Katz, 2009, Ikuta et al., 2014] and others purely for telicity [Friedrich and Gateva, 2017, Kober et al., 2020, Metheniti et al., 2022], some (generally smaller) datasets are also annotated for a more general schema such as Vendler's classification [Zarcone and Lenci, 2008, Hermes et al., 2015, Zellers and Choi, 2017] or Universal Dependencies (UD) aspect classes [Kondratyuk and Straka, 2019]. So while there is a relatively wide range of datasets, they are all annotated for slightly different tasks and with different guidelines, which makes matters a little more challenging.

For aspect ambiguity the only currently existing annotated dataset of which I am aware is Friedrich and Palmer [2014], which is however only annotated for ambiguity between DYNAMIC and STATIVE. I therefore had to create my own (see **??**).

1.3 Related areas of work

Situation entities

Situation entity classification is the task of identifying different types of situations, which exist at a clausal level. The task comes more from the tradition of discourse analysis since it is important for discourse representation theory (DRT) to know, for example, which new referents are introduced to a discourse, and also to analyse temporal relationships. Works usually use the original 8 types introduced by Smith [2003]: events, states, generalizing sentences, generic sentences, facts, propositions, questions and imperatives. While incorporating concepts from linguistic aspect, the focus is more on a discourse level and the classification schemata are thus different accordingly. Some works in this area include Palmer et al. [2007], Friedrich et al. [2016], Becker et al. [2017], Friedrich [2017] and Dai and Huang [2018].

Applications: Temporal reasoning

An application of the tasks looked at in this chapter and a related area of work is the more general area of temporal reasoning. While this is evidently closer to the area of situation entities introduced in 1.3, the more linguistically-informed side of aspect also has a role to play. For example, it is clear that the two following sentence, differing only in their aspect, refer to two very different events.

(1) The driver of this vehicle drinks wine.

(2) The driver of this vehicle is drinking wine.

Costa and Branco [2012] analyse the importance of aspect for temporal relation classification, and Derczynski and Gaizauskas [2015] employ a tense and aspect framework to sequence events in text. Allen and UzZaman [2012] explores various aspects of tense and aspect in event processing. Similarly, Kamp and Reyle [1993] discusses the integration of tense and aspect information within discourse representation theory (DRT). Furthermore, Kober et al. [2019] develop a dataset to evaluate the temporal and aspectual entailment capabilities of models.

More broadly, it is clear that understanding verbal aspect is crucial for a variety of more general downstream tasks, including machine translation and question answering systems. See Friedrich et al. [2023] for a more in-depth look at the applications of this area.

Bibliography

Annemarie Friedrich, Nianwen Xue, and Alexis Palmer. A kind introduction to lexical and grammatical aspect, with a survey of computational approaches. In Andreas Vlachos and Isabelle Augenstein, editors, *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pages 599–622, Dubrovnik, Croatia, May 2023. Association for Computational Linguistics. doi: 10.18653/v1/2023.eacl-main.44. URL https://aclanthology.org/2023.eacl-main.44.

Eric V. Siegel and Kathleen R. McKeown. Learning methods to combine linguistic indicators:improving aspectual classification and revealing linguistic insights. *Computational Linguistics*, 26(4):595–627, 2000. URL https://aclanthology.org/J00-4004.

Marc Moens and Mark Steedman. Temporal ontology and temporal reference. *Computational Linguistics*, 14(2):15–28, 1988. URL https://aclanthology.org/J88-2003.

Zeno Vendler. Verbs and times. *The Philosophical Review*, 66(2):143–160, 1957. ISSN 00318108, 15581470. URL http://www.jstor.org/stable/2182371.

Annemarie Friedrich and Alexis Palmer. Automatic prediction of aspectual class of verbs in context. In *Annual Meeting of the Association for Computational Linguistics*, 2014. URL https://api.semanticscholar.org/CorpusID:1832412.

Markus Egg, Helena Prepens, and Will Roberts. Annotation and automatic classification of aspectual categories. In Anna Korhonen, David R. Traum, and Lluís Màrquez, editors, *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers*, pages 3335–3341. Association for Computational Linguistics, 2019. doi: 10.18653/V1/P19-1323. URL https://doi.org/10.18653/v1/p19-1323.

Christopher Kennedy and Beth Levin. Measure of change: The adjectival core of degree achievements. In *Adjectives and Adverbs*, 2008. URL https://semantics.uchicago.edu/kennedy/docs/kl07-measure.pdf.

William Croft, Pavlina Pešková, and Michael Regan. Annotation of causal and aspectual structure of events in RED: a preliminary report. In Martha Palmer, Ed Hovy, Teruko Mitamura, and Tim

O'Gorman, editors, *Proceedings of the Fourth Workshop on Events*, pages 8–17, San Diego, California, June 2016. Association for Computational Linguistics. doi: 10.18653/v1/W16-1002. URL https://aclanthology.org/W16-1002.

W. Styler, K. Crooks, T. O'Gorman, and M. Hamang. Richer Event Description (RED) Annotation Guidelines, 2014. Unpublished manuscript, University of Colorado at Boulder.

Daniel Chen, Martha Palmer, and Meagan Vigus. AutoAspect: Automatic annotation of tense and aspect for uniform meaning representations. In Claire Bonial and Nianwen Xue, editors, *Proceedings of the Joint 15th Linguistic Annotation Workshop (LAW) and 3rd Designing Meaning Representations (DMR) Workshop*, pages 36–45, Punta Cana, Dominican Republic, November 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.law-1.4. URL https://aclanthology.org/2021.law-1.4.

Annemarie Friedrich and Damyana Gateva. Classification of telicity using cross-linguistic annotation projection. In Martha Palmer, Rebecca Hwa, and Sebastian Riedel, editors, *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2559–2565, Copenhagen, Denmark, September 2017. Association for Computational Linguistics. doi: 10.18653/v1/D17-1271. URL https://aclanthology.org/D17-1271.

Eleni Metheniti, Tim Van De Cruys, and Nabil Hathout. About time: Do transformers learn temporal verbal aspect? In Emmanuele Chersoni, Nora Hollenstein, Cassandra Jacobs, Yohei Oseki, Laurent Prévot, and Enrico Santus, editors, *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics*, pages 88–101, Dublin, Ireland, May 2022. Association for Computational Linguistics. doi: 10.18653/v1/2022.cmcl-1.10. URL https://aclanthology.org/2022.cmcl-1.10.

Anisia Katinskaia and Roman Yangarber. Probing the category of verbal aspect in transformer language models, 2024.

Yuri Kuratov and Mikhail Arkhipov. Adaptation of deep bidirectional multilingual transformers for russian language, 2019.

Mehdi Rezaee, Kasra Darvish, Gaoussou Youssouf Kebe, and Francis Ferraro. Discriminative and generative transformer-based models for situation entity classification, 2021.

Thomas A. Mathew and E. Graham Katz. Supervised categorization of habitual and episodic sentences. In *Sixth Midwest Computational Linguistics Colloquium*, Bloomington, Indiana, 2009. Indiana University.

Rei Ikuta, Will Styler, Mariah Hamang, Tim O'Gorman, and Martha Palmer. Challenges of adding causation to richer event descriptions. In Teruko Mitamura, Eduard Hovy, and Martha Palmer, editors, *Proceedings of the Second Workshop on EVENTS: Definition, Detection, Coreference, and Representation*, pages 12–20, Baltimore, Maryland, USA, June 2014. Association for Computational Linguistics. doi: 10.3115/v1/W14-2903. URL https://aclanthology.org/W14-2903.

Thomas Kober, Malihe Alikhani, Matthew Stone, and Mark Steedman. Aspectuality across genre: A distributional semantics approach. In Donia Scott, Nuria Bel, and Chengqing Zong, editors, *Proceedings of the 28th International Conference on Computational Linguistics*, pages 4546–4562, Barcelona, Spain (Online), December 2020. International Committee on Computational Linguistics. doi: 10.18653/v1/2020.coling-main.401. URL https://aclanthology.org/2020.coling-main.401.

Alessandra Zarcone and Alessandro Lenci. Computational models for event type classification in context. In Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, and Daniel Tapias, editors, *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco, May 2008. European Language Resources Association (ELRA). URL http://www.lrec-conf.org/proceedings/lrec2008/pdf/315_paper.pdf.

Jürgen Hermes, Michael Richter, and Claes Neuefeind. Automatic induction of german aspectual verb classes in a distributional framework. 09 2015.

Rowan Zellers and Yejin Choi. Zero-shot activity recognition with verb attribute induction. In Martha Palmer, Rebecca Hwa, and Sebastian Riedel, editors, *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 946–958, Copenhagen, Denmark, September 2017. Association for Computational Linguistics. doi: 10.18653/v1/D17-1099. URL https://aclanthology.org/D17-1099.

Dan Kondratyuk and Milan Straka. 75 languages, 1 model: Parsing Universal Dependencies universally. In Kentaro Inui, Jing Jiang, Vincent Ng, and Xiaojun Wan, editors, *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th*

International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pages 2779–2795, Hong Kong, China, November 2019. Association for Computational Linguistics. doi: 10.18653/v1/D19-1279. URL https://aclanthology.org/D19-1279.

Carlota S. Smith. *Modes of Discourse: The Local Structure of Texts*. Cambridge Studies in Linguistics. Cambridge University Press, 2003.

Alexis Palmer, Elias Ponvert, Jason Baldridge, and Carlota Smith. A sequencing model for situation entity classification. In Annie Zaenen and Antal van den Bosch, editors, *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 896–903, Prague, Czech Republic, June 2007. Association for Computational Linguistics. URL https://aclanthology.org/P07-1113.

Annemarie Friedrich, Alexis Palmer, and Manfred Pinkal. Situation entity types: automatic classification of clause-level aspect. In Katrin Erk and Noah A. Smith, editors, *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1757–1768, Berlin, Germany, August 2016. Association for Computational Linguistics. doi: 10.18653/v1/P16-1166. URL https://aclanthology.org/P16-1166.

Maria Becker, Michael Staniek, Vivi Nastase, Alexis Palmer, and Anette Frank. Classifying semantic clause types: Modeling context and genre characteristics with recurrent neural networks and attention. In Nancy Ide, Aurélie Herbelot, and Lluís Màrquez, editors, *Proceedings of the 6th Joint Conference on Lexical and Computational Semantics (*SEM 2017)*, pages 230–240, Vancouver, Canada, August 2017. Association for Computational Linguistics. doi: 10.18653/v1/S17-1027. URL https://aclanthology.org/S17-1027.

Annemarie Silke Friedrich. States, events, and generics: computational modeling of situation entity types. 2017.

Zeyu Dai and Ruihong Huang. Building context-aware clause representations for situation entity type classification. In Ellen Riloff, David Chiang, Julia Hockenmaier, and Jun'ichi Tsujii, editors, *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3305–3315, Brussels, Belgium, October-November 2018. Association for Computational Linguistics. doi: 10.18653/v1/D18-1368. URL https://aclanthology.org/D18-1368.

Francisco Costa and António Branco. Aspectual type and temporal relation classification. In Walter Daelemans, editor, *Proceedings of the 13th Conference of the European Chapter of* *the Association for Computational Linguistics*, pages 266–275, Avignon, France, April 2012. Association for Computational Linguistics. URL https://aclanthology.org/E12-1027.

Leon Derczynski and Robert J. Gaizauskas. Temporal relation classification using a model of tense and aspect. In *Recent Advances in Natural Language Processing*, 2015. URL https://aclanthology.org/R15-1017.pdf.

James F. Allen and Naushad UzZaman. Interpreting the temporal aspects of language. 2012. URL https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=f0b9a86a80a5bc09e448db250a97ade218e4d529.

Hans Kamp and Uwe Reyle. *Tense and Aspect*, pages 483–689. Springer Netherlands, Dordrecht, 1993. ISBN 978-94-017-1616-1. doi: 10.1007/978-94-017-1616-1_6. URL https://doi.org/10.1007/978-94-017-1616-1_6.

Thomas Kober, Sander Bijl de Vroe, and Mark Steedman. Temporal and aspectual entailment. In Simon Dobnik, Stergios Chatzikyriakidis, and Vera Demberg, editors, *Proceedings of the 13th International Conference on Computational Semantics - Long Papers*, pages 103–119, Gothenburg, Sweden, May 2019. Association for Computational Linguistics. doi: 10.18653/v1/W19-0409. URL https://aclanthology.org/W19-0409.