Computational Narratology: Extracting Tense Clusters from Narrative Texts

Thomas Bögel, Jannik Strötgen, Michael Gertz

Institute of Computer Science, Heidelberg University, Germany
{thomas.boegel, stroetgen, gertz}@informatik.uni-heidelberg.de

Abstract

Computational Narratology is an emerging field within the Digital Humanities. In this paper, we tackle the problem of extracting temporal information as a basis for event extraction and ordering, as well as further investigations of complex phenomena in narrative texts. While most existing systems focus on news texts and extract explicit temporal information exclusively, we show that this approach is not feasible for narratives. Based on tense information of verbs, we define *temporal clusters* as an annotation task and validate the annotation schema by showing that the task can be performed with high inter-annotator agreement. To alleviate and reduce the manual annotation effort, we propose a rule-based approach to robustly extract temporal clusters using a multi-layered and dynamic NLP pipeline that combines off-the-shelf components in a heuristic setting. Comparing our results against human judgements, our system is capable of predicting the tense of verbs and sentences with very high reliability: for the most prevalent tense in our corpus, more than 95% of all verbs are annotated correctly.

1. Motivation & Introduction

Temporal dependencies reveal interesting insights into the semantic discourse structure of narrative texts for narratologists and are fundamental mechanisms for expressing temporal relations between events. As of today, the investigations of literary scientists are mostly based on laborious manual annotations.

Computational Narratology, a subtopic of the emerging field of the Digital Humanities aims at facilitating annotations and supporting literary scientists with their analyses. According to Mani (2013), one aspect of Computational Narratology focuses on the exploration and testing of literary hypotheses through mining of narrative structures from corpora. Despite the potential value of Natural Language Processing (NLP) to facilitate objectives of narrative scientists with respect to temporal structures, this topic has been mostly neglected in the past.

In this paper, we present our first results obtained from the ongoing project heureCLÉA[1], a collaborative initiative between computational linguists and narrative scientists. The project focuses in particular on the extraction of temporal structures in German literary texts using computational methods. Our approach thus makes a contribution to the combination of literary science and NLP.

In the following, we outline and evaluate an approach to extract *tense clusters* based on verb tenses within documents as a first step towards German event extraction and ordering for narrative texts. By using robust methods that rely on morphological analyses and heuristics, we overcome the domain-dependence of state-of-the-art NLP tools.

Subject of this study are fictional narrative stories. Consider the example in Figure 1: the underlined sentence marks a shift of tense from one sentence to the next one. While the main tense of this text is preterite, the underlined text introduces a short passage that is written in pluperfect (past perfect). From a narratological perspective, this marks the introduction of a passage of *retrospective narration*. These

Ferdinand stand auf der breiten Treppe, den Wandelnden nachsehend und diejenigen betrachtend, welche heraufstiegen, um dem Hochamte beizuwohnen. . . . mit gleichgültigem Herzen und lächelndem Auge hatte er die mannigfaltigen Gestalten betrachtet . . .

Translation: *Ferdinand stood on a wide flight of stairs, his gaze following people walking by to celebrate High Mass. . . . he had inspected the manifold shapes with resigned heart and smiling eyes . . .*

Figure 1: Example narrative text demonstrating temporal shifts.

shifts between tense clusters are a valuable resource for a semantic analysis of the relationship between discourse and corresponding events.

Most approaches that deal with revealing the temporal structure of a text tackle the problem from a viewpoint of explicit *temporal markers* and apply it to news texts, e.g., in the TempEval challenges (Verhagen et al., 2009; Verhagen et al., 2010; UzZaman et al., 2013). While there are tools available that perform temporal tagging for German (Strötgen and Gertz, 2013), we show that this approach in isolation is neither feasible for our text domain nor of particular interest for the problem set of literary scientists – although the results of a temporal tagger can be used as complementary information.

Instead of relying solely on temporal expressions, we focus on shifts in the *tense structure of sub-sentences* to **cluster a narrative text based on its temporal dimension** with respect to tenses. The presented approach to extract tense structures from narrative texts is based on morphological tagging and robust heuristics to counterbalance the domain-dependence of underlying state-of-the-art NLP tools. Our ultimate goal is to extract and investigate temporal orderings of events in our narrative texts similar to Mani and Schiffman (2005). We adopt the view of Reichenbach (1947) that motivates the usage of tense clusters as one fac-

---

[1]Website: `http://heureclea.de/`

	heureCLÉA	WikiWarsDE
Documents	21	22
Token	79431	95604
Sentences p. document	213.4	249.9
Token p. sentence	17.7	17.1
TIMEX3 p. document	15	101.8

Table 1: Characteristics of the data set.

tor of temporal ordering of events. We are currently working on automatically extracting additional temporal phenomena (e.g., deictic time points).

The remainder of the paper is structured as follows: we first describe the nature of our data set and compare it to a standard data set in German that is used for temporal tagging (Section 2.). We then describe our approach (Section 3.), apply a sate-of-the-art temporal tagger, and report first experimental results for both temporal tagging and fine-grained tense annotation (Section 4.).

2. Data Set and Narrative Annotations

The heureCLÉA corpus currently consists of 21 German narrative texts from various authors of the 20th century, each comprising less than ten pages. It includes, for instance, texts from Thomas Mann, Ernest Hemmingway, and Arthur Schnitzler. Due to the diversity of style and text characteristics in literary texts, applying Natural Language Processing is challenging as most systems are optimized for non-fictional texts characterized by stable structures.

Annotation schema. In the corpus, various aspects of narratological research are currently being annotated by literary scientists. Apart from tense annotations that are targeted in this paper, they comprise very fine-grained narratological aspects related to temporal structures, such as relations between discourse and history (e.g., anachrony, prolepsis) or plot organizing sequences and patterns, e.g., Proppian functions (Gius et al., 2013).

With an increasing number of annotated material, we will employ machine learning techniques to automatically reproduce and verify additional manual annotations in the future.

Comparison to factual texts. We initially started our attempt to extract temporal structures by applying a state-of-the-art temporal tagger on the heureCLÉA corpus. Temporal tagging in isolation, however, proved to be insufficient for a deep analysis in particular due to the low number of temporal expressions. To make this issue more clear, we compare the heureCLÉA data set with WikiWarsDE, a German non-fictional corpus that provides manually annotated temporal expressions for parts of Wikipedia documents (Strötgen and Gertz, 2011).

As Table 1 reveals, the characteristics of a *non-fictional* text are very informative with respect to temporal expressions. In contrast, with only 15 TIMEX3 annotations per document on average – which is roughly equivalent to one TIMEX3 expression every 15 sentences – discourse structures of *narrative texts* cannot be derived by simply relying

on temporal expressions in the heureCLÉA corpus. Furthermore, narratological scientists are more interested in structures within the text than explicit temporal anchors.

2.1. Definition of Temporal Clusters

We define the term **temporal cluster** as a contiguous sequence of tokens belonging to the same sub-sentence, where a **sub-sentence** is defined with regard to the output of a constituent parser, referring to every distinct part of a sentence with its own *S* node (e.g., main clause, subordinate clause etc.). We chose sub-sentences as our annotation target based on a discourse perspective: usually, one sub-sentence only refers to one single **event** in a narrative text, while a sentence might contain references to multiple events. In the end, we want to be able to extract and order all events occurring in narrative texts based on temporal aspects (tense, among others) as described similarly in (Mani, 2012). Thus, tagging sub-sentences ensures that the temporal cluster refers to all elements of one specific event, even though most words do not contain explicit tense information.

Adopting this perspective, our system annotates a text as follows: if tense information for a verb t_v can be extracted, all tokens t_1 to t_n within the same sub-sentence S_* of the verb ($S_* = \{t_1, t_2, \cdots, t_v, \cdots, t_n\}$) are annotated with the same tense. In the example in Figure 1, all underlined tokens would be annotated as *pluperf*.

Besides the motivation in terms of a discourse perspective, sub-sentences as annotation targets also constitute a reasonable annotation task for human annotators, which is validated by a high inter-annotator agreement presented next.

2.2. Annotation Process & Annotator Agreement

The corpus is being annotated by multiple students with a background in narratology. After a short calibration phase of collective annotation on different parts of the text that allowed for extensive discussions of debatable cases and considerably increased the detailedness of the annotation guidelines, pairs of two annotators annotated the first 20% of all tokens of each document independently. The result of this independent annotation phase was used to measure inter-annotator agreement (see below). In a final post-processing step, differing annotations were adjudicated, which resulted in a clean version used to evaluating our system.

Fleiss-κ (Fleiss et al., 1981) is used to measure inter-annotator agreement. In our case, we compute the agreement between two annotators and six different categories – one category for each of the five possible tense annotations, as well as a sixth category if a token has not been annotated. The results for all tokens, as well as verbs only, are given in Table 2. While the tense of a sub-sentence depends on the verb, annotators are asked to tag all tokens within the sub-sentence and thus choosing correct sub-sentence boundaries. Naturally, the agreement score is higher on the token-level ($\kappa = 0.897$): if two annotators agree with the tense of a verb (which is the main annotation target), by our definition of temporal clusters, they also agree with all tokens within the same sub-sentence – except for disagreements of sub-sentence boundaries. Nevertheless, a κ score

Annotation target	κ
Tokens	0.897
Verbs	0.84

Table 2: Inter-Annotator agreement for human judgements.

Tense	Token	Percentage
present	1559	13.11%
preterite	9216	77.47%
perfect	223	1.87%
pluperfect	776	6.52%
future	122	1.03%
sum	**11896**	

Table 3: Distribution of annotated tenses in the heureCLÉA corpus.

of 0.84 for all verbs and 6 different categories shows very high agreement and thus validates the annotation task and the corresponding guidelines.

Tense statistics. To get an impression of the distribution of tenses within the corpus and to be able to interpret the impact of individual evaluation scores for each tense, Table 3 lists the resulting distribution of *adjudicated* annotated tokens broken down by individual tenses. Being an instantiation of the so-called *narrative past tense* as a marker of fictionality (Fludernik, 2003), it is not surprising that the *preterite* is by far the most prevalent tense within the corpus.

3. Extracting Temporal Narrative Blocks

To extract tense clusters from our texts, we perform the following three steps that are explained in more detail below: first, we apply a tool chain of NLP components as the foundation for further analysis. We then split sentences into sub-sentences and aggregate the output of a morphological analyser in a set-value representation. Finally, we apply various heuristics to predict the tense of all tokens within a sub-sentence.

3.1. Preprocessing Pipeline

Reproducing and alleviating manual annotations that are related to the temporal structure of texts requires information on multiple levels of the linguistic processing stack. Thus, we implemented a modular pipeline based on Apache UIMA[2] that performs annotations with increasing levels of complexity and allows for easy adaptation and exchange of different base components. We use standard off-the-shelf tools, such as the TreeTagger for German (Schmid, 1995) and HeidelTime (Strötgen and Gertz, 2013). The pipeline architecture is pictured in Figure 2.

CATMA interface. The texts in our corpus are annotated by literary scientists with CATMA[3], a web-based collaborative annotation tool that provides humanists an easy way
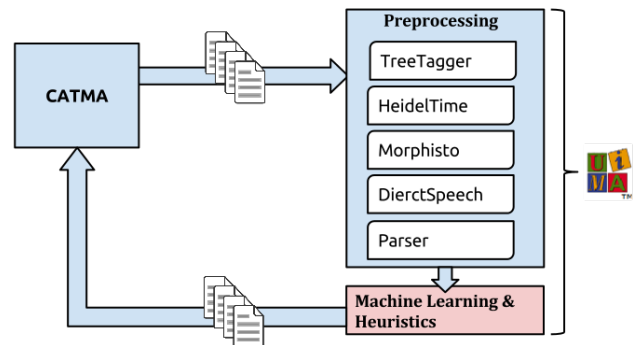
Figure 2: Architecture of our preprocessing pipeline. Documents are imported from CATMA into our pipeline based on UIMA. After preprocessing, tense annotations are predicted and the annotated files are exported to CATMA.

to create stand-off annotations and share their annotations with other scholars. In order to work with annotated data in CATMA, we implemented a component that interfaces CATMA with our UIMA pipeline. As CATMA is a popular tool in the humanities, we developed the interface as a stand-alone component that can easily be used by others to combine the strengths of CATMA as an annotation framework with the analytical and predictive power of UIMA pipelines. The interface is geared towards literary scientists without programming expertise. To achieve a simple configuration, the user only has to specify mappings between CATMA and UIMA types in a single XML-file. By providing an easy-to-use interface, we want to lower the bar for other projects in the humanities to employ simple yet effective NLP tools and thereby alleviate manual annotations.

Sub-sentence detection and morphological analysis. As motivated in Section 2.1., sentence boundaries are too coarse-grained as a unit for tense clusters. Thus, we use constituent parses obtained from the Stanford parser during preprocessing to extract **sub-sentences** as our target annotation for tense clusters. Morphisto (Zielinski et al., 2009), a morphological analyser based on finite state automata, is used to perform a morphological analysis of all verbs in the document.

3.2. Temporal Heuristics

We distinguish between simple and composite German verb forms for tenses, where **simple** means that the tense is expressed and thus derivable from one word (e.g., "sah", engl. "saw" represents a preterite form) and **composite** refers to tense forms that consist of two or more words. One such complex case is the passive form of the German future perfect that consists of four elements.

For **composite verb forms**, the German tense system employs a past participle or the infinitive and an inflected verb form of the auxiliaries (*aux*) "haben" (to have), "sein" (to be) or "werden" (to become), depending on the verb and the temporal aspect.

Based on the morphologically analysed text, we process it incrementally and create a set of **tense markers** encountered for each sub-sentence, thereby distinguishing between the following categories of markers:

- Verb forms in the past or future (e.g., "sah", engl. "saw")

- Infinitive verb forms (e.g., "sehen", engl. "to see")

- Participle components with a perfective aspect (e.g., "gesehen", engl. "seen")

- Auxiliary verbs (*aux* in Table 4)

Whenever we encounter a marker of one of the above types, we add it to a set of temporal indicators for the current sub-sentence.

In addition, we extract sub-sentences in **direct speech** and **imperative** forms to make the tense extraction more precise. With this set of extraction patterns, we are able to generalize over a wide range of possible instantiations for certain categories of tense markers.

After a sub-sentence has been processed, its final target tense is predicted using the rule set given in Table 4. For simple tenses, the label of the morphology component is adopted as the final tense. Note that passive verb forms are also captured by our rule set.

Following the idea similar to that presented by Bethard et al. (2012), we annotate sub-sentences in direct speech separately as they do not contribute to the overall progression of the story and make the overall analysis of tense clusters much more complex.

Discontinuities. Due to a very diverse vocabulary in the corpus, some verb forms are unknown either to the lexicon of the TreeTagger or the morphology component. Instead of modifying the base components, we adopt a simpler approach and assess whether our procedure is sufficient: if a sub-sentence cannot be annotated, we use the neighbouring sentences to the left and right to transfer the tense annotation if the context both to the left and right has the same tense. If the tense of the left and right context differ, the sentence is left un-annotated. We evaluate the contribution of this heuristic separately.

4.   Experiments

Evaluation setting. In our evaluation, we focus on two different aspects: the percentage of verbs and the percentage of all tokens that are assigned the correct tense. The classification of verbs gives an insight into the trustworthiness of our tense heuristics. However, since the complete contextual content of a sub-sentence is of importance for a narratological analysis, we also take into account the accuracy of all tokens in a sub-sentence (see also Section 2.2.). We only selected annotated tokens that satisfied the following two conditions: (1) they are annotated by at least two annotators and (2) there is no disagreement between any of the annotators for the token. Thus, all annotations are either unambiguous for two annotators or have been adjudicated.

Results. The evaluation results are presented in Table 5. We distinguish between a setting that uses the heuristic for discontinuities described in Section 3.2. (+ *disc. heuristic*) and a system that does not (- *disc. heuristic*).

Overall, despite the lack of training data, our method yields robust results across all annotated tenses. In general, performance is slightly higher when focusing on verbs only,

showing that parser errors make it difficult to find exact sub-sentence boundaries. Nevertheless, using sub-sentences as boundaries for potential shifts of tenses proves to be a reasonable approach.

The second column clearly shows that the addition of our sentence heuristic yields a substantial performance increase. This is an important insight as it uses knowledge transfer for words that are unknown.

The best setting correctly tags about 95% of all tokens/verbs in the preterite – comprising the majority of all annotated tokens with more than 77% – with the correct tense. Even for rare tenses in our corpus, such as the future tense with only 122 annotated tokens, about 90% of all tokens are annotated correctly. Overall, these promising results indicate a high potential for reducing manual annotation efforts.

Error analysis.   Most of the errors are due to erroneous *parser outputs*, leading to wrong spans of sub-sentences. For example, the parser misses to insert a sub-sentence boundary in the following sentence where the first part is written in future tense and the second part in present tense: "niemals werde ich ihn wiedersehen . . . Das Meer ist grau."[4]

Errors in the parser output are often caused by very long sentences that consist of a combination of multiple main clauses without any connectives, mostly for stylistic reasons. The parsers employed in our system are trained on texts from the news domain and are thus incapable of handling these complex structures.

In addition, there are debatable *annotation artefacts* that are challenging for our system. In the sentence "Es ist der Tag, an welchem sie mir allmonatlich die Blumen schickte" (translation: It is the day she used to send me flowers every month), our system correctly identifies the second part as past tense. The manual annotation, however, tags it as present tense, which makes sense from the viewpoint of an informed reader (because the setting is in the present) but does not reflect the correct verb tense.

Finally, the extracted part-of-speech tags are partially erroneous, mostly due to unknown words and archaic verb forms (such as "dünkte", "thun"), a challenging characteristic of any literary, non-current text.

4.1.   Visualization of Temporal Clusters

The information extracted by our system can be used to obtain visual representations of tense clusters in documents. Figure 3 shows two examples of texts from our corpus with a completely different narrative structure. For the sake of clarity, we restrict the graphical representation to three different tenses, although a more fine-grained output is possible: present, past and future tense. Each box in the figure represents one sub-sentence and the sequence of boxes is equivalent to the linear progression of the text.

Figure 3a illustrates tense clusters in "Meine erste Liebe" (a short story written by Ludwig Thoma) and is written almost completely in past tense, except for a short passage where a general statement is given. This is an interesting narrative turning point in the story and might reveal insights for

---

[4]Translation: I will never see him again . . . The sea is grey.

Markers	Tense
participle & $aux_{present}$	perfect
participle & $aux_{present}$ & $aux_{werden,\ partic.}$	perfect. (pass.)
participle & aux_{past}	pluperf.
participle & aux_{past} & $aux_{werden,\ partic.}$	pluperf. (pass.)
participle & $aux_{werden,\ present}$	present (pass.)
participle & $aux_{werden,\ past}$	preterite (pass.)
infinitive & $aux_{werden,\ present}$	future
participle & $aux_{werden,\ present}$ & $aux_{werden,\ infinitive}$	future (pass.)

Table 4: Rule set for predicting the final tense. *aux* refers to auxiliary verbs, the subscript to different forms of the auxiliary. For example, $aux_{werden,\ part.}$ refers to "geworden", the participle form of "werden". *pass.* signalizes passive verb forms.

	All tokens		Verbs	
Tense	- disc. heuristic	+ disc. heuristic	- disc. heuristic	+ disc. heuristic
present	89.22	89.03	93.10	93.68
preterite	92.30	94.87	93.36	95.73
perfect	87.44	87.44	96.43	96.43
pluperfect	85.95	86.21	84.71	85.88
future	76.23	86.89	80.00	90.00
weighted avg.	91.23	93.32	92.68	94.77

Table 5: Evaluation results for all tokens and verbs. "with disc." uses the heuristic described in Section 3.2. to counterbalance extraction errors.

narratologists. In Figure 3b, tense clusters for the text "Der Tod" (a novella by Thomas Mann) are given. This story features many different changes of narrative points of view and thus changes the time much more frequently, resulting in a mix of narrative past tense clusters and passages of present/future tense.

We are currently working on a user interface that displays the corresponding text passages when hovering over a tense cluster and lets users tweak settings about the information to show in the output (e.g., a more fine-grained resolution of tenses). Overall, we think that this kind of visual output can provide narratologists with a useful tool as a starting point for further investigations.

5. Ongoing Work

As the evaluation and error analysis shows, our first attempt to extract tense clusters in narrative texts yields promising and especially robust results. In a next step, we are currently annotating the entire corpus with automatic tense annotations and have them redacted by our annotators. In addition, we are implementing a supervised system for additional temporal aspects that are annotated in our corpus (e.g., explicit and implicit time representations) to facilitate deeper narratological research and order events in narratives. Our final goal is to integrate the current heuristic approach into a hybrid system that combines heuristics and machine learning by first proposing automatic annotations based on robust heuristics that are then corrected by annotators. The output of this manual feedback will be used as an input for a supervised system that learns which heuristic rules should be trusted in which scenarios, making it possible to incrementally increase the complexity of predicted

phenomena.

6. Conclusion

In this paper, we presented an approach to automatically annotate the tense of sub-sentences in German narrative texts in order to alleviate manual annotations for the Digital Humanities. We motivated the usefulness of tense annotations as a basis for further narratological research and present and validate tense clusters as a suitable annotation target. Based on a dynamic pipeline of NLP components, we propose a robust rule set accounting for possible combinations of tense markers. The evaluation of our heuristic system shows that we are able to reproduce and predict tense clusters with very high reliability.

Acknowledgements

7. References

Steven Bethard, Oleksandr Kolomiyets, and Marie-Francine Moens. 2012. Annotating story timelines as temporal dependency structures. In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, pages 23–25.

Joseph L Fleiss, Bruce Levin, and Myunghee Cho Paik. 1981. The measurement of interrater agreement. *Statistical methods for rates and proportions*, 2:212–236.
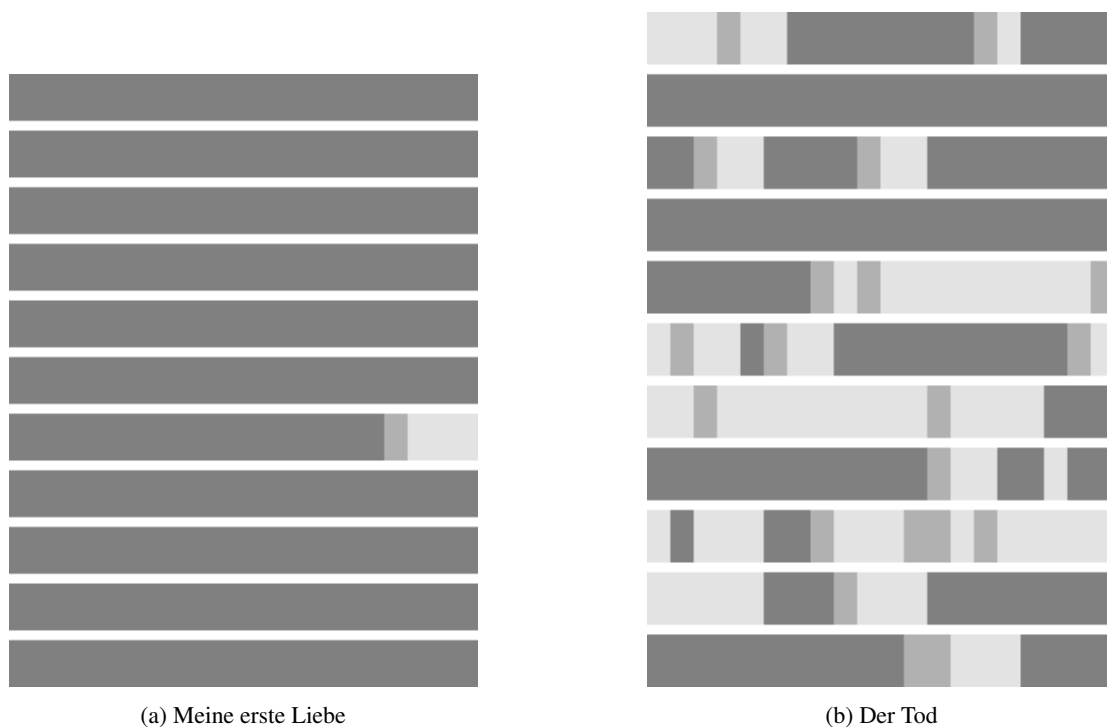
(a) Meine erste Liebe   (b) Der Tod

Figure 3: Visualization of tense clusters. Dark grey represents sub-sentences in past tense, medium grey in present tense and light grey stands for sentences in future tense.

Monika Fludernik. 2003. Chronology, time, tense and experientiality in narrative. *Language and Literature*, 12(2):117–134.

Evelyn Gius, Janina Jacke, Jan Meister, and Marco Petris. 2013. Catma annotation guidelines. Technical report, Institut für Germanistik II, Hamburg University.

Inderjeet Mani and Barry Schiffman. 2005. Temporally anchoring and ordering events in news. *Time and Event Recognition in Natural Language. John Benjamins*.

Inderjeet Mani. 2012. Computational modeling of narrative. *Synthesis Lectures on Human Language Technologies*, 5(3):1–142.

Inderjeet Mani. 2013. Computational Narratology. The living handbook of narratology. `http://www.lhn.uni-hamburg.de/article/computational-narratology`.

Hans Reichenbach. 1947. *Elements of Symbolic Logic*. Macmillan.

Helmut Schmid. 1995. Improvements in part-of-speech tagging with an application to German. In *In Proceedings of the ACL SIGDAT-Workshop*.

Jannik Strötgen and Michael Gertz. 2011. WikiWarsDE: A German corpus of narratives annotated with temporal expressions. In *Proceedings of the conference of the German society for computational linguistics and language technology (GSCL 2011)*, pages 129–134.

Jannik Strötgen and Michael Gertz. 2013. Multilingual and cross-domain temporal tagging. *Language Resources and Evaluation*, 47(2):269–298.

Naushad UzZaman, Hector Llorens, Leon Derczynski, Marc Verhagen, James Allen, and James Pustejovsky. 2013. Semeval-2013 task 1: Tempeval-3: Evaluating time expressions, events, and temporal relations. In *Second joint conference on lexical and computational semantics (* SEM)*, volume 2, pages 1–9.

Marc Verhagen, Robert Gaizauskas, Frank Schilder, Mark Hepple, Jessica Moszkowicz, and James Pustejovsky. 2009. The tempeval challenge: identifying temporal relations in text. *Language Resources and Evaluation*, 43(2):161–179.

Marc Verhagen, Roser Sauri, Tommaso Caselli, and James Pustejovsky. 2010. Semeval-2010 task 13: Tempeval-2. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, pages 57–62. Association for Computational Linguistics.

Andrea Zielinski, Christian Simon, and Tilman Wittl. 2009. Morphisto: Service-oriented open source morphology for german. In Cerstin Mahlow and Michael Piotrowski, editors, *State of the Art in Computational Morphology*, volume 41 of *Communications in Computer and Information Science*, pages 64–75. Springer Berlin Heidelberg.