A Hybrid Approach to Extract Temporal Signals from Narratives

Thomas Bögel **Jannik Strötgen** **Michael Gertz**
Institute of Computer Science, Heidelberg University, Germany
{thomas.boegel,stroetgen,gertz}@informatik.uni-heidelberg.de

Abstract

When processing literary narratives, standard temporal annotation specifications – typically developed for processing news-style documents – do not match the expectations of literary scholars. Thus, a different definition of *temporal signals* is required. In this paper, we define this concept from the narratological perspective and present our hybrid approach developed in the context of the heureCLÉA[1] project to extract temporal signals. Our evaluation demonstrates high quality extraction results, making the approach directly applicable to the literary domain.

1 Temporal Signals

The temporal markup language TimeML (Pustejovsky et al., 2005) was developed for temporally annotating text documents such as business news. Annotations include temporal expressions, temporal signals, events, and temporal relations. Besides four types of temporal expressions (date, time, duration, and set expressions) covered by TimeML's TIMEX3 tag, the SIGNAL tag is used to annotate expressions that might be helpful for temporal reasoning, e.g., temporal prepositions and conjunctions (Pustejovsky et al., 2005).

In general, narratologists make a much more fine-grained distinction between different kinds of temporal expressions (Lahn and Meister, 2013, §2.3) and include many more that are neglected by TimeML. Temporal expressions (including temporal signals) from TimeML represent a very small sub-set of the narratological definition, such as narratological time-points ("On July 22, 1848"). Many more expressions, however, are covered by the narratological definition, e.g., event-related time points (e.g., "while the bells were ringing…"). We thus define a *temporal signal* from the narratological perspective as *a phrase capturing temporal semantics excluding tense*.

2 A Hybrid Approach

In contrast to TimeML's TIMEX3 and SIGNAL, the narratological definition of temporal signals implies that one is faced with an open vocabulary. The combination of an open vocabulary and the context of narrative texts with substantial variations in styles and textual content across different texts leads to the issue of data sparsity when trying to predict temporal signals. With a rule-based approach, instances in documents different from the set of documents used for deriving the rule set will most likely not be found. Thus, instead of applying a fully rule-based approach for their extraction – similar to the temporal tagger HeidelTime (Strötgen and Gertz, 2013) for TIMEX3 –, we use a hybrid extraction approach.

After extending HeidelTime to radically increase the recall for extracting narratological temporal signals, we use machine learning to remove incorrectly extracted expressions and to achieve a balanced relation between precision and recall. This technique combines the advantages of both approaches: (i) generalized heuristics yield high recall without the need of copious amounts of training data; (ii) the machine learning component is perfectly suited for improving precision by looking at universal contexts that are preferably general and can be applied across different texts.

Extension of HeidelTime. Instead of applying HeidelTime to extract TIMEX3 annotations only, we extend its rule base by adding general rules to capture a broad range of temporal signals. Note that due to the open vocabulary used to formulate temporal signals, these rules are very general and aim at a high recall ignoring precision issues. The output of HeidelTime's extended version is thus a set of candidate signals $C = \{c_1, \dots c_n\}$.

[1] http://heureclea.de/

Proceedings of the Int. Conference of the German Society for Computational Linguistics and Language Technology,
pages 106–107, University of Duisburg-Essen, Germany, Sep 30–Oct 2 2015.

	# doc.	token	TIMEX3	signals
train	21	79,431	315	3,144
test	4	23,218	11	215

Table 1: Statistics of the training and test set.

Machine Learning. To boost precision, we train a classifier that judges the output of the heuristic system. For each candidate c, a binary classifier determines whether c represents a signal, resulting in a set of final predictions for temporal signals $S = \{s_1, \ldots, s_m\}$, where $S \subseteq C$. The classifier is trained on manual annotations of the training set.

3 Data Sets and Evaluation

We use the corpus of the heureCLÉA project, consisting of 25 narrative texts in German from various authors of the 20th century that comprise less than ten pages (Bögel et al., 2014). To train and evaluate our approach, we split the data into distinct training and test sets. The data set statistics in Table 1 show that narratological temporal signals are much more prevalent in the data than TIMEX3 expressions.

Manual annotation. To extend the rule set and train the classifier, we performed an error-driven evaluation. After a first run of HeidelTime, an expert in narratology manually annotated erroneous, missing, and correct candidates in all training documents. The test set was annotated separately and without prior knowledge about system predictions.

Feature set. Overall, we used 17 features to train the classifier. The feature set comprises information about the length and part-of-speech tags of the candidate, structural properties like the occurrence of the candidate in complex sentence structures, as well as string-based features characterising the subject and verb of the sentence containing the candidate and the presence of temporal adverbs within the sentence. Finally, we investigate changes of verb tense in the surrounding context.

Evaluation. To demonstrate that the narratological perspective on temporal signals differs from TimeML's approach, we first evaluate the performance of HeidelTime on the test set. Then, we show the effects of tackling the problem with a heuristic system by extending and generalizing HeidelTime's rule set. Finally, we report the results of our hybrid approach.

		prec.	rec.	F_1
HeidelTime	*strict*	23.1	1.4	2.6
(TIMEX3 only)	*loose*	84.6	5.1	9.7
Heuristics	*strict*	33.5	78.1	46.9
(ext. HeidelTime)	*loose*	38.5	89.8	53.8
Hybrid	*strict*	74.7	71.7	73.2
(Heuristics + ML)	*loose*	83.5	77.6	80.4

Table 2: Evaluation results on the test set.

The evaluation results in Table 2 confirm the assumption that a temporal tagger in isolation is not sufficient to extract narratological temporal signals, yielding devastating recall. Extending the rule set significantly increases recall but leads to many false positives – as expected. Finally, our hybrid approach that combines heuristics and machine learning achieves the best and most balanced result with a high precision of 74.7% and 83.5% for strict and loose evaluation metrics, respectively. While recall is slightly decreased due to the nature of our setting, it is still solid, especially considering the fundamental textual differences between training and test set. The large drop in recall of the hybrid system for the loose setup can be explained by the fact that we treat overlapping temporal signals as candidates that should be filtered out to boost precision.

4 Future Work

As mentioned above, many recall errors are due to the handling of overlaps as errors when training the classifier. Thus, the next step is to add overlapping candidates as a separate classification outcome to increase the recall of the system. In addition, we are working on a more fine-grained classification of different types of temporal signals by implementing a two-step classification.

References

Thomas Bögel, Jannik Strötgen, and Michael Gertz. 2014. Computational narratology: Extracting tense clusters from narrative texts. In *LREC*, pages 950–955.

Silke Lahn and Jan Christoph Meister. 2013. *Einführung in die Erzähltextanalyse*. J.B. Metzler.

James Pustejovsky, Robert Knippen, Jessica Littman, and Roser Sauri. 2005. Temporal and Event Information in Natural Language Text. *Language Resources and Evaluation*, 39(2–3):123–164.

Jannik Strötgen and Michael Gertz. 2013. Multilingual and Cross-domain Temporal Tagging. *Language Resources and Evaluation*, 47(2):269–298.