KRAUTS: A German Temporally Annotated News Corpus

Jannik Strötgen^{α}, Anne-Lyse Minard^{β δ}, Lukas Lange^{α}, Manuela Speranza^{β}, Bernardo Magnini^{β}

^{*a*} Max Planck Institute for Informatics, Saarland Informatics Campus, Saarbrücken, Germany

 $^{\beta}$ Fondazione Bruno Kessler, Trento, Italy

^δ Univ Rennes, Inria, CNRS, IRISA, Rennes, France

{jannik.stroetgen,llange}@mpi-inf.mpg.de, {minard,manspera,magnini}@fbk.eu

Abstract

In recent years, temporal tagging, i.e., the extraction and normalization of temporal expressions, has become a vibrant research area. Several tools have been made available, and new strategies have been developed. Due to domain-specific challenges, evaluations of new methods should be performed on diverse text types. Despite significant efforts towards multilinguality in the context of temporal tagging, for all languages except English, annotated corpora exist only for a single domain. In the case of German, for example, only a narrative-style corpus has been manually annotated so far, thus no evaluations of German temporal tagging performance on news articles can be made. In this paper, we present KRAUTS, a new German temporally annotated corpus containing two subsets of news documents: articles from the daily newspaper DOLOMITEN and from the weekly newspaper DIE ZEIT. Overall, the corpus contains 192 documents with 1,140 annotated temporal expressions, and has been made publicly available to further boost research in temporal tagging.

Keywords: temporal tagging, corpus annotation, TIMEX3

1. Introduction

Temporal tagging - the extraction and normalization of temporal expressions from texts - is an important task towards improved natural language understanding. Once temporal expressions have been detected in a text, their semantics can be assigned to them in a standard format so that applications can exploit not only the surface forms of temporal expressions, but also their meaning. For instance, applications in event / timeline extraction (Minard et al., 2015; Cornegruta and Vlachos, 2016; Spitz and Gertz, 2016), question answering (Llorens et al., 2015) and (temporal) information retrieval (Kanhabua et al., 2015) can exploit temporal tagging output. Thus, temporal tagging has become a vibrant research area, and several new temporal taggers have been made available and new strategies have been developed. However, as was shown in previous work (Mazur and Dale, 2010; Strötgen and Gertz, 2013; Bethard et al., 2016; Tabassum et al., 2016), different types of documents pose different challenges for temporal tagging such that domain-sensitive normalization strategies are required (Strötgen and Gertz, 2016). To judge the performance of temporal taggers and new methods, evaluations need to be performed on diverse text types, e.g., on news articles and narrative-style Wikipedia documents. In contrast to many natural language processing tasks, there

In contrast to many natural language processing tasks, there has also been some effort towards multilinguality in the context of temporal tagging, e.g., research competitions were organized not only for English but covered further languages such as Spanish and Italian (Verhagen et al., 2010; Caselli et al., 2014). Despite its importance, German has not been part of any of these challenges so far. In addition, HeidelTime is the only publicly available temporal tagger for German, and only narrative style corpora have been manually annotated so far. Thus, no proper evaluations of German temporal tagging performance on news articles can be carried out. Therefore, HeidelTime's German temporal tagging quality has only been evaluated on narrative texts using the WikiWarsDE and AncientTimes corpora. In this paper, we present our effort in developing KRAUTS, a new temporally annotated corpus in German containing two subsets of news documents: articles from the daily newspaper DOLOMITEN and from the weekly newspaper DIE ZEIT. For annotating temporal expressions in the corpus, we developed annotation guidelines for German temporal tagging by using the guidelines defined for Italian (Caselli and Sprugnoli, 2015) as a starting point. Overall, the corpus contains 192 documents with 1,140 annotated temporal expressions, and the corpus as well as the annotation guidelines have been made publicly available to further boost research in temporal tagging.¹

2. Related Work

The task of temporal processing has gained interest in recent years, in particular thanks to the TempEval tasks at SemEval (Verhagen et al., 2007; Verhagen et al., 2010; Uz-Zaman et al., 2013; Llorens et al., 2015; Bethard et al., 2016). Temporal tagging is a subtask of temporal processing and consists of the identification of temporal expressions in texts and their normalization to some standard format. Strötgen and Gertz (2016) present a complete overview of the task as well as a survey of the resources, tools, etc. They focus on the description of domainsensitive temporal tagging and multilingual taggers.

The annotation of temporal expressions follows in most cases the TimeML annotation guidelines (Pustejovsky et al., 2003) developed first for English. They have then been adapted to other languages such as Italian (Caselli and Sprugnoli, 2015), Spanish (Saurí et al., 2009) and French (Bittar, 2010). However until now no adaptation of the guidelines to German has been done. The two corpora of narratives in German, AncientTimes (Strötgen et al., 2014) and WikiWarsDE (Strötgen and Gertz, 2011), have been manually annotated but following the English TimeML guidelines without further specifying language-

¹http://github.com/JannikStroetgen/KRAUTS

specific adaptations. WikiWarsDE is the German counterpart of the English WikiWars corpus (Mazur and Dale, 2010) and AncientTimes is a small multilingual corpus containing documents about history.

Driven by the above-mentioned shared tasks, many temporal taggers have been developed. Some of these can process several languages, such as TIPSem (Llorens et al., 2010) for English and Spanish, TimePro (Mirza and Minard, 2014) (a module of TextPro²) for English, Italian and French, and HeidelTime (Strötgen and Gertz, 2013) for 13 languages, including German, as well as its automatic extension as a baseline temporal tagger for more than 200 languages (Strötgen and Gertz, 2015). Strötgen et al. (2014) performed an evaluation of HeidelTime on two German corpora of narratives: WikiWarsDE and AncientTimes. They reported F1-scores of 87.7 and 78.0 for strict match, and value F1-scores³ of 80.4 and 82.2 on Wiki-WarsDE and AncientTimes, respectively.

Our work consists of defining TimeML guidelines for German and annotating a corpus following these guidelines. Using the newly annotated corpus, we report evaluation results for HeidelTime, which to the best of our knowledge is the only temporal tagger available for German.

3. Corpus Description

KRAUTS (Korpus of newspapeR Articles with Underlined Temporal expressionS) consists of two subsets: articles of the daily, regional newspaper DOLOMITEN and articles of the nationwide weekly newspaper DIE ZEIT. The corpus is composed of 192 documents with a total of 75,678 tokens. Table 1 contains some statistics about both subsets. Details about annotated temporal expressions will be given in Section 5.

3.1. Dolomiten

The DOLOMITEN subset consists of 142 articles published between 2009 and 2016. DOLOMITEN is a local newspaper from South Tyrol (Italy) written in the local variant of German. Therefore, the articles contain words and phrases which are not used in High German, including the temporal expression *heuer* which translates to "this year".

Two students, supervised by two expert annotators, performed the manual annotation of temporal expressions. 100 DOLOMITEN articles were annotated starting from raw text, while the remaining 42 were first pre-annotated with the HeidelTime tool and then checked and corrected manually in order to speed-up the annotation process.

3.2. Die Zeit

In the context of a Bachelors thesis on a time-centric analysis of German news articles (Lange, 2017), 50 documents of the German weekly newspaper DIE ZEIT were manually annotated by two annotators – but without first adapting the English annotation guidelines to German in a concise way. This resulted in several discussions about nonuniform annotations, and it was concluded that proper an-

	DOLOMITEN	Die Zeit	KRAUTS
# documents	142	50	192
# tokens	31,422	44,256	75,678
tokens/doc	221	885	394

Table 1: Statistics about the KRAUTS corpus.

notation guidelines for German are required to achieve high quality manual annotations. In the context of the collaboration between the Fondazione Bruno Kessler and the Max Planck Institute for Informatics, the 50 articles have been re-annotated following the newly developed annotation guidelines for German (cf. Section 4).

Compared to the documents in the DOLOMITEN part of the KRAUTS corpus, the DIE ZEIT articles are very long (an average of 885 tokens vs. 221 per article, respectively). In addition, as they are sometimes rather non-standard news articles, annotating temporal expressions in these documents is probably more challenging, even for humans.

4. German-specific Guidelines

As German presents some language-specific phenomena which have to be taken into account when performing any annotation task, it is not possible to apply the English temporal annotation guidelines to German in a straightforward way. The adaptations to be done mainly affect the extent of temporal expressions. In particular, in German there are compounds which can sometimes contain temporal expressions (e.g., Diskussionabende "Evenings of discussion") and there are contractions of prepositions and articles (e.g., im: in + dem "in the"). It is widely accepted that annotations of temporal expressions should always start and end at a token boundary, whereas the specific morphology of German would lead to annotating a subpart of the tokens in the case of compounds and contractions. This illustrates that the English annotation guidelines cannot be directly applied, as these state that articles, when present, are part of temporal expressions while prepositions are not.

In order to develop the guidelines needed for the annotation of temporal expressions in German, we selected the It-TimeML guidelines (Caselli and Sprugnoli, 2015) as a reference. The choice was motivated not only by the fact that these guidelines are very well-defined and detailed, but also by the fact that in Italian, as in German, it is also possible to contract articles and prepositions. Thus, the Italian guidelines are a more natural choice than the English ones when adapting annotation guidelines to German.

The new guidelines we produced are summarized in the document *Examples and Guidelines for Annotation of Temporal Expressions (<TIMEX3>) in German*, which is an annex to the It-TimeML guidelines. It is available for download on the It-TimeML website⁴ and linked from the KRAUTS website. The annex contains the extensions needed to adapt the It-TimeML guidelines to the specific morpho-syntactic features of German, as well as many German annotated examples to illustrate the application to the German language of the relevant It-TimeML guidelines.

²http://textpro.fbk.eu

³"value F1-score" consists of evaluating both the recognition of the temporal expressions with relaxed match and the correctness of the normalization value.

⁴https://sites.google.com/site/ittimeml/ documents

The numbering of the examples in the annex is the same as in the It-TimeML guidelines.

4.1. Compounds

Compound words, lexemes that consists of more than one lexical element, are very frequent in German. For example, *Werktag* is composed of *Werk* "work" and *Tag* "day" and means "working day". We can see in this example that compounds can contain lexical elements with temporal meaning.

According to our guidelines, a compound containing a temporal trigger has to be annotated if the syntactic head of the compound is a temporal trigger. If the syntactic head is not a temporal trigger, on the other hand, the compound should not be annotated, even if it contains a temporal trigger. For example, *Diskussionsabende* in (1) is annotated because the syntactic head *abend* "evening" is a temporal trigger, whereas *Monatsblatt* in (2) is not annotated (the syntactic head is *blatt* "leaflet", which is not a temporal trigger).

- Weiters werden im Jugendtreff <TIMEX3>zwei Diskussionsabende</TIMEX3> veranstaltet.
 [Furthermore, two evenings of discussion will be organized in the youth center.]
- (2) Jedes **Monats**blatt behandelt ein eigenes Thema. [Each monthly leaflet addresses a specific subject.]

4.2. Prepositions, Articles, and Contractions

Following the general TimeML rule, articles (whatever their case) are included in the extent of the temporal expressions (3), while prepositions are excluded (4); as for contractions of prepositions and definite articles, we adopted the Italian guidelines and so they are not included in the extent of the temporal expression (5). This decision leaves open the possibility of marking contractions in a future step, as they often include prepositions used as indicators of temporal relations. According to the TimeML framework, these are to be marked as SIGNAL if the full task of temporal annotation and not just temporal tagging is the goal.

- (3) <TIMEX3>der nächste Tag</TIMEX3> [the next day]
- (4) **um** <TIMEX3>15:00 Uhr</TIMEX3> [at 3 pm]
- (5) **im** <TIMEX3>nächsten Jahr</TIMEX3> [in_the next year]

5. Corpus Annotation

In the KRAUTS corpus, we have annotated the following types of temporal expressions: DATE (calendar unit), TIME (time of the day), DURATION (period of time), and SET (reoccurring temporal expression), and the following attributes: tid (identifier), type (type of the temporal expression), value (normalized value), anchorTimeID (ID of the temporal expression to which the marked expression is linked), beginPoint (begin point of a DURATION), end-Point (end point of a DURATION), freq (frequency of a

	DOLOMITEN	Die Zeit	KRAUTS
DATE	376	358	734
TIME	98	12	110
DURATION	94	144	238
SET	19	39	58
Total	587	553	1,140
Empty tag	30	41	71

Table 2: Annotation statistics (in the first part of the table we give the number of text-consuming TIMEX3).

SET), quant (quantifier of a SET) and mod (temporal modifier).

According to the TimeML annotation guidelines, TIMEX3 tags with no extent (i.e., empty TIMEX3 tags) are introduced, for example, to deal with unspecified time points, which are sometimes needed to anchor durations. Nevertheless, in most prior work, empty TIMEX3 tags have not been used, neither in annotated corpora nor by TIMEX3compliant temporal taggers.

However, in order to represent durations in a better way, empty TIMEX3 tags should be annotated. We thus followed the organizers of the Italian temporal tagging challenge EVENTI (Caselli et al., 2014) and the developers of the MEANTIME corpus (Minard et al., 2016), who were the first, and, to the best of our knowledge, the only researchers so far who have annotated empty TIMEX3 tags in documents which have resulted in publicly available corpora (the EVENTI corpus for Italian and the MEANTIME corpus contains English, Dutch, Italian, and Spanish temporally annotated news articles). An example of an empty TIMEX3 tag is given in (6): The duration vor einem Monat "one month ago" (vor is outside of the TIMEX3 tag as it is a preposition) is annotated in the text and an empty TIMEX3 tag of type DATE is added which represent the date of one month ago to anchor the duration.

(6) (DCT: 2018-01-17, t0) ... vor <TIMEX3 tid="t1" type="DURATION" value="P1M" beginPoint="t2" endPoint="t0">einem Monat</TIMEX3> ...
<TIMEX3 tid="t2" type="DATE" value="2017-12-17" anchorTimeID="t0"/>
[... one month ago ...]

In Table 2, we provide information about the distribution of the different types of temporal expressions in the corpus. In total, KRAUTS contains 1,140 (text-consuming) temporal expressions and 71 empty TIMEX3 tags. 64% of the temporal expressions are of type DATE.

The DOLOMITEN subset contains 587 temporal expressions, a large proportion of which are dates. We observe a rather high number of temporal expressions of type TIME with regards to the type of texts (newspaper articles). We can explain it by the presence of local event announcements in the DOLOMITEN newspaper.

The DIE ZEIT subset contains 553 temporal expressions, with a majority of dates. Compared to the DOLOMITEN subset, it contains few time expressions and many duration and set expressions. The rather low number of time expressions is probably due to the fact that DIE ZEIT is a weekly

newspaper so that such fine-granular expressions are less important. Duration and set expressions often occur in articles belonging to categories (such as "travel") that are not typical news categories. This also shows that the DIE ZEIT subset contains very diverse articles and several some of the documents can be considered as rather less typical news articles compared to articles of a daily newspaper – which increases the difficulty of temporal tagging the DIE ZEIT subset of the KRAUTS corpus.

6. Evaluating HeidelTime on KRAUTS

In Table 3, we present the evaluation of HeidelTime, performed with the TempEval-3 scorer⁵, on KRAUTS.⁶ We have performed the evaluation on three sections of the corpus separately: DOLOMITEN-42 (the subpart of the Dolomiten articles pre-annotated with HeidelTime and revised manually), DOLOMITEN-100 (the Dolomiten articles annotated manually starting from raw texts), and DIE ZEIT. For comparison, the best system for relaxed matching at TempEval-3 on English news documents, SUTime, obtained an F1-score of 90.32 on relaxed match, 79.57 on strict match and 67.38 on value, and the overall best system, HeidelTime, obtained an F1-score of 90.30 on relaxed match, 81.34 on strict match, and 77.61 on value.

The results obtained on DOLOMITEN-42 are higher than those obtained on the other two sections; this can be explained considering that the DOLOMITEN-42 articles had been pre-annotated with HeidelTime and consequently the final annotation (after manual revision) might still be slightly biased.

The results on the DIE ZEIT articles are the lowest, probably because the articles are very long and thus have a more complex temporal discourse structure. In addition, some articles are written in a narrative rather than news-style fashion – due to the characteristics of the weekly newspaper in general – which led to incorrect normalizations.

In Table 4, we give the detailed results obtained for each type of temporal expression. We can observe that the best results are obtained for temporal expressions of type DATE and DURATION. The results for type SET are low, but it should be noticed that the corpus contains very few of them so that few false positives and false negatives lower the score significantly.

We have analyzed the annotations made by HeidelTime on the corpus DOLOMITEN-42 (corpus of 42 files preannotated with HeidelTime). We counted three false positives: the age of a person (which is not to be marked as a temporal expression), a four digit number that was not a year, and an occurrence of "Christmas" which did not refer to Christmas as a time period but, more in general, as a subject. As false negatives we found many expressions where

⁵www.cs.york.ac.uk/semeval-2013/task1/ index.html

	strict	relaxed	type	value
DOLOMITEN-42	79.4	83.1	82.5	80.0
DOLOMITEN-100	71.4	81.3	76.9	66.2
Die Zeit	69.7	79.3	75.4	62.8

Table 3: HeidelTime evaluation results (in terms of F1-score) on the three subsets of KRAUTS.

	DATE	TIME	DURAT.	SET
DOLOMITEN-42	84.8	77.1	87.5	50.0
DOLOMITEN-100	83.3	51.7	66.1	54.6
Die Zeit	78.8	76.2	74.5	37.5

Table 4: HeidelTime evaluation results (relaxed match in terms of F-measure) on the three subsets of KRAUTS.

the time of day (e.g., "um 20.30 Uhr" [at 20:30]) was not an hour on the dot, as well as expressions with the ordinal number of the week (e.g., "der dritten Woche" [the third week]). Introducing a few new rules will prevent the tool from leaving out these temporal expressions in the future.

7. Conclusions

In this paper, we defined specific TimeML guidelines for German starting from It-TimeML, the Italian TimeML guidelines. Following these new guidelines we annotated a corpus of newspaper articles: the KRAUTS corpus. It is the first news corpus for temporal tagging in German. It consists of 192 articles from a daily, regional newspaper and a weekly newspaper, with 1,140 annotated temporal expressions.

As a benchmark for the evaluation of automatic systems, we have exploited KRAUTS to evaluate HeidelTime, the temporal tagger for German, which has only been evaluated against narrative-style corpora for German so far. On two of the three subparts of the corpus (DOLOMITEN-100 and DIE ZEIT) it obtained F1-scores of around 70 and 80, respectively for strict and relaxed match.

KRAUTS contains different kinds of news articles which differ from the length of the documents and the proportion of each type of TIMEX3. This is not enough in order to develop and evaluate a generic temporal tagger. For German, narrative-style temporal annotated documents are also available. It will now be interesting to also annotate some colloquial texts, such as tweets or emails.

8. Acknowledgments

We thank Sara Baino and Martina Coser for their contribution, which consisted of manually annotating the DOLOMITEN articles. This work has been partially funded by the EUCLIP_RES project under the FESR Program of the Autonomous Province of Bolzano - South Tyrol.

9. Bibliographical References

Bethard, S., Savova, G., Chen, W.-T., Derczynski, L., Pustejovsky, J., and Verhagen, M. (2016). SemEval-2016 Task 12: Clinical TempEval. In *Proceedings of the 10th International Workshop on Semantic Evaluation*, SemEval'16, pages 1052–1062. Association for Computational Linguistics.

⁶"Strict" and "relaxed" refer to the evaluation of the extent of the temporal expressions; "type" and "value" represent the evaluation of the respective attributes. The F1-score is computed taking into account the recognition of the temporal expressions with relaxed match and the identification of the attribute. The TempEval-3 scorer does not evaluate the empty TIMEX3 tags, so they are not part of the presented evaluation.

- Bittar, A. (2010). ISO-TimeML Annotation Guidelines for French, version 1.0. Technical report.
- Caselli, T. and Sprugnoli, R. (2015). It-TimeML, TimeML Annotation Guidelines for Italian, version 1.4. Technical report.
- Caselli, T., Sprugnoli, R., Speranza, M., and Monachini, M. (2014). EVENTI: EValuation of Events and Temporal INformation at Evalita 2014. In *Proceedings of the Fourth International Workshop EVALITA*, EVALITA 2014, pages 27–34.
- Cornegruta, S. and Vlachos, A. (2016). Timeline Extraction Using Distant Supervision and Joint Inference. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP'16, pages 1936–1942. Association for Computational Linguistics.
- Kanhabua, N., Blanco, R., and Nørvåg, K. (2015). Temporal Information Retrieval. *Found. Trends Inf. Retr.*, 9(2):91–208.
- Lange, L. (2017). Time in Newspaper A Largescale Analysis of Temporal Expressions in News Corpora. Bachelor's thesis, Universität des Saarlandes, Max Planck Institute for Informatics, Saarland Informatics Campus.
- Llorens, H., Saquete, E., and Navarro, B. (2010). TIPSem (English and Spanish): Evaluating CRFs and Semantic Roles in TempEval-2. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, SemEval'10, pages 284–291. Association for Computational Linguistics.
- Llorens, H., Chambers, N., UzZaman, N., Mostafazadeh, N., Allen, J., and Pustejovsky, J. (2015). SemEval-2015 Task 5: QA TempEval - Evaluating Temporal Information Understanding with Question Answering. In Proceedings of the 9th International Workshop on Semantic Evaluation, SemEval'15, pages 792–800. Association for Computational Linguistics.
- Mazur, P. and Dale, R. (2010). WikiWars: A New Corpus for Research on Temporal Expressions. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, EMNLP'10, pages 913–922. Association for Computational Linguistics.
- Minard, A.-L., Speranza, M., Agirre, E., Aldabe, I., van Erp, M., Magnini, B., Rigau, G., and Urizar, R. (2015).
 SemEval-2015 Task 4: TimeLine: Cross-Document Event Ordering. In *Proceedings of the 9th International Workshop on Semantic Evaluation*, SemEval'15, pages 778–786. Association for Computational Linguistics.
- Minard, A.-L., Speranza, M., Urizar, R., Altuna, B., van Erp, M., Schoen, A., and van Son, C. (2016). MEANTIME, the NewsReader Multilingual Event and Time Corpus. In *Proceedings of the 10th International Conference on Language Resources and Evaluation*, LREC'16, pages 4417–4422. ELRA.
- Mirza, P. and Minard, A.-L. (2014). FBK-HLT-time: A Complete Italian Temporal Processing System for EVENTI-Evalita 2014. In *Proceedings of the Fourth International Workshop EVALITA*, EVALITA 2014.
- Pustejovsky, J., Castaño, J. M., Ingria, R., Saurí, R., Gaizauskas, R. J., Setzer, A., Katz, G., and Radev, D. R.

(2003). TimeML: Robust Specification of Event and Temporal Expressions in Text. In *New Directions in Question Answering*, pages 28–34.

- Saurí, R., Saquete, E., and Pustejovsky, J. (2009). Annotating Time Expressions in Spanish TimeML Annotation Guidelines. Technical report.
- Spitz, A. and Gertz, M. (2016). Terms over LOAD: Leveraging Named Entities for Cross-Document Extraction and Summarization of Events. In *Proceedings of the* 39th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR'16, pages 503–512. ACM.
- Strötgen, J. and Gertz, M. (2011). WikiWarsDE: A German Corpus of Narratives Annotated with Temporal Expressions. In In Proceedings of the Conference of the German Society for Computational Linguistics and Language Technology, GSCL'11, pages 129–134.
- Strötgen, J. and Gertz, M. (2013). Multilingual and Crossdomain Temporal Tagging. *Language Resources and Evaluation*, 47(2):269–298.
- Strötgen, J. and Gertz, M. (2015). A Baseline Temporal Tagger for all Languages. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, EMNLP'15, pages 541–547. Association for Computational Linguistics.
- Strötgen, J. and Gertz, M. (2016). *Domain-sensitive Temporal Tagging*. Synthesis Lectures on Human Language Technologies, Morgan & Claypool Publishers, San Rafael, CA.
- Strötgen, J., Bögel, T., Zell, J., Armiti, A., Canh, T. V., and Gertz, M. (2014). Extending HeidelTime for Temporal Expressions Referring to Historic Dates. In Proceedings of the 9th International Conference on Language Resources and Evaluation, LREC'14, pages 2390–2397. ELRA.
- Tabassum, J., Ritter, A., and Xu, W. (2016). TweeTime : A Minimally Supervised Method for Recognizing and Normalizing Time Expressions in Twitter. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP'16, pages 307–318. Association for Computational Linguistics.
- UzZaman, N., Llorens, H., Derczynski, L., Allen, J., Verhagen, M., and Pustejovsky, J. (2013). SemEval-2013 Task 1: TempEval-3: Evaluating Time Expressions, Events, and Temporal Relations. In *Proceedings* of 7th International Workshop on Semantic Evaluation, SemEval'13, pages 1–9. Association for Computational Linguistics.
- Verhagen, M., Gaizauskas, R., Schilder, F., Hepple, M., Katz, G., and Pustejovsky, J. (2007). SemEval-2007 Task 15: TempEval Temporal Relation Identification. In Proceedings of the 4th International Workshop on Semantic Evaluations, SemEval'07, pages 75–80. Association for Computational Linguistics.
- Verhagen, M., Saurí, R., Caselli, T., and Pustejovsky, J. (2010). SemEval-2010 Task 13: TempEval-2. In Proceedings of the 5th International Workshop on Semantic Evaluation, SemEval'10, pages 57–62. Association for Computational Linguistics.