A Baseline Temporal Tagger for all Languages

Jannik Strötgen and **Michael Gertz**
Institute of Computer Science, Heidelberg University
Im Neuenheimer Feld 348, 69120 Heidelberg, Germany
{stroetgen,gertz}@informatik.uni-heidelberg.de

Abstract

Temporal taggers are usually developed for a certain language. Besides English, only few languages have been addressed, and only the temporal tagger HeidelTime covers several languages. While this tool was manually extended to these languages, there have been earlier approaches for automatic extensions to a single target language. In this paper, we present an approach to extend HeidelTime to all languages in the world. Our evaluation shows promising results, in particular considering that our approach neither requires language skills nor training data, but results in a baseline tagger for 200+ languages.

1 Introduction & Related Work

Temporal tagging (the extraction and normalization of temporal expressions) is a crucial task in NLP, and many applications can benefit from temporal information, e.g., in the context of question answering and information retrieval (Campos et al., 2014). Thus, there has been a lot of research during the last years addressing temporal information extraction, as reflected by manually annotated corpora, approaches to temporal tagging, and research challenges such as the TempEval series (Verhagen et al., 2010; UzZaman et al., 2013).

While, in the meantime, the range of annotated corpora covers several languages different from English, e.g., the French, Portuguese, Italian, and Romanian TimeBank corpora (Bittar et al., 2011; Costa and Branco, 2012; Caselli et al., 2011; Forascu and Tufis, 2012), most approaches to temporal tagging focus on processing English text, e.g., DANTE (Mazur and Dale, 2009; Mazur, 2012) and SUTime (Chang and Manning, 2012). Of course, exceptions exist, e.g., TipSem for English and Spanish (Llorens et al., 2010), and HeidelTime (Strötgen and Gertz, 2013) even covers

13 languages. However, it required a lot of manual effort to extend HeidelTime, and researchers of different institutes and countries have been involved for German (Strötgen and Gertz, 2011), Dutch (van de Camp and Christiansen, 2012), Spanish (Strötgen et al., 2013), French (Moriceau and Tannier, 2014), Croatian (Skukan et al., 2014), Vietnamese and Arabic (Strötgen et al., 2014), Italian (Manfredi et al., 2014), Chinese (Li et al., 2014), Russian, Estonian, and Portuguese.

Obviously, it is possible to manually extend HeidelTime and thus a temporal tagger in general. However, each language is typically addressed separately, and the extension process is time- and labor-intensive, so that a method to automate this process is desirable. While there have been earlier approaches for automatic extensions of temporal taggers to further languages (Saquete et al., 2004; Negri et al., 2006; Spreyer and Frank, 2008), these were limited to a few languages and the results were considered less successful, in particular for the normalization subtask. In contrast, Angeli and Uszkoreit (2013) presented an approach to language-independent parsing of temporal expressions, however, addressing only the normalization and not the extraction subtask.

In this paper, we re-address the task of automatically extending a temporal tagger to further languages. In contrast to previous work, we address both subtasks of temporal tagging, i.e., the extraction and the normalization, and we do not limit our work to a few selected languages, but we aim at extending HeidelTime to cover all languages in the world. In addition of making use of automatic pattern translations based on Wiktionary,[1] we allow for fuzzy matching in the newly addressed languages. This strategy results in promising temporal tagging quality for many languages, although our evaluations show that the quality of a manually extended temporal tagger is not reached, and

---

[1] http://www.wiktionary.org/.

that there are open issues that need to be addressed for specific language characteristics.

By making our newly developed resources publicly available, we establish a baseline temporal tagger for basically all languages in the world – for comparison or as starting point for improvements.

2 HeidelTime

HeidelTime is a rule-based, domain-sensitive, and publicly available temporal tagger that was developed with multilinguality in mind so that the source code and language-dependent resources are strictly separated, and it can be extended to new languages without modifying its source code (Strötgen and Gertz, 2013).

2.1 Input and Output Format

As input, HeidelTime requires linguistic preprocessed documents with sentence, token, and part-of-speech (pos) information – although it is possible to create language resources not making use of pos constraints. For processing news-style documents, a publication date is additionally needed.

Temporal expressions are annotated following the temporal markup language TimeML (Pustejovsky et al., 2005) using TIMEX3 tags with normalization attributes, most importantly *type* (date, time, duration, set) and *value* to represent the main semantics of expressions in standard format.

2.2 Language-dependent Resources

Each language requires its own language resources consisting of patterns, normalization information and rules. Pattern files contain terms that are frequently used to form temporal expressions, e.g., for names of months and weekdays or numbers that may refer to a day or year. Regular expressions can be used in the pattern files, and each pattern file is interpreted as a regular expression pattern, which can be called by HeidelTime's rules.

Normalization files contain normalized information for patterns defined in the pattern files, e.g., that the value of "January" is "01" and that "3" has to be normalized to "03" in case it refers to a day or a month and to "3" in case it refers to a duration.

Finally, for the four types of temporal expressions (date, time, duration, and set), a rule file contains all rules for respective expressions. In addition to a rule name, the extraction part and the value normalization part are obligatory. The former is a regular expression that may contain any number of references to pattern files, and in the latter, normalization functions are called to normalize respective parts of the extracted expressions. In addition, part-of-speech constraints can be specified, negative rules can be formulated, and offset information can be set, e.g., if parts of the extracted patterns are required context information during the extraction process but not part of the temporal expression itself.[2]

2.3 Manual Extension to a New Language

So far, HeidelTime resources for newly supported languages have been manually created (cf. references is Section 1). For this, the following steps had to be completed: (i) linguistic preprocessing for the new language has been provided, (ii) all patterns in the pattern and normalization files have been manually translated, and (iii) rules have been developed iteratively. Starting with simple rules of the source language (typically English), rules have been adapted and added based on false negatives, false positives, partial matches and incorrect value normalizations on the training data for the new language. Then, more complex rules of the source language have been addressed step-by-step.

Note that tasks (ii) and (iii) have been carried out by language experts, and task (iii) was performed until no further improvements could be achieved on the training data. In addition to the fact that this process is time- and labor intensive, further disadvantages are that a language expert and temporally annotated training data are required. Finally, languages are typically addressed by different researchers with differing effort so that quality and completeness varies.

3 Automatic Extension to All Languages

In this section, we present the steps of our automatic extension approach in detail. A graphical summary of our approach is depicted in Figure 1.

3.1 Linguistic Preprocessing

While part-of-speech (pos) information can be helpful to write more general rules and to improve temporal tagging quality, there is no pos tagger covering all languages in the world. In addition,

---

[2]While these explanations are necessary to understand our automated resource development approach that will be presented in Section 3, for further details about HeidelTime and its language-dependent resources we refer to Strötgen and Gertz (2013) and `https://github.com/HeidelTime/heideltime`.
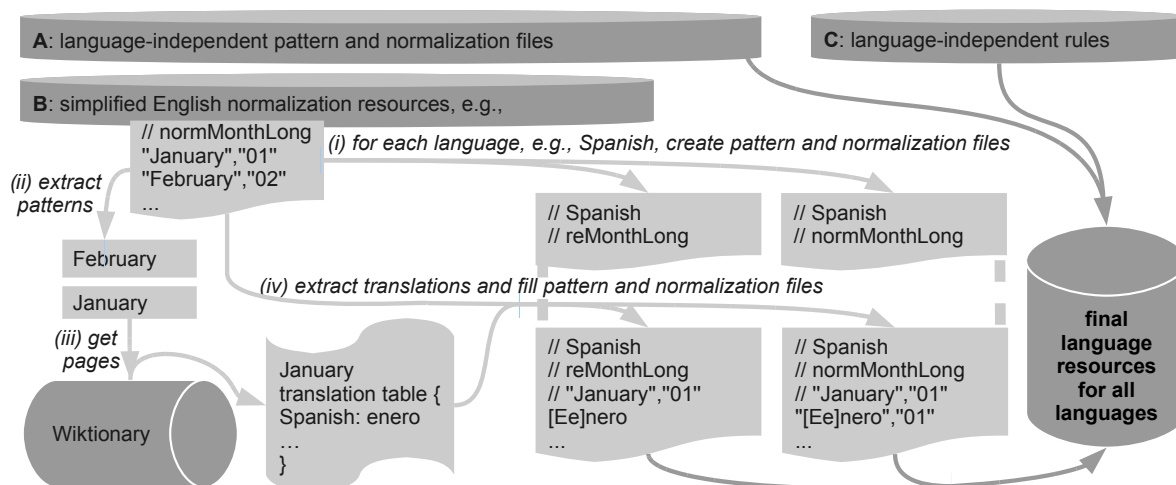
Figure 1: Graphical overview of the automated resource development process. Resources A and C are directly added to the resources of all languages; resource B is processed following steps (i) to (iv).

the only obligatory linguistic preprocessing steps are sentence splitting and tokenization. Thus, we write a simple and generic, language-independent sentence splitter and tokenizer, which we use for all languages. Note that our tokenization is based on a white-space token boundary assumption so that we have to take care of languages without such token boundaries separately (cf. Section 4).

3.2 Language-independent Resources

Some patterns and normalization information are valid for all (many) languages, e.g., numbers for days and months. For such patterns, we write language-independent pattern and normalization resources (Figure 1, A). These do not contain language-dependent terms and will be part of the resources for each of the languages in the world.

3.3 Simplified English Resources

The main goal of the simplified English resources is to make HeidelTime's original English resources amenable to automatic translation. Thus, rules are addressed in a language-independent fashion (cf. Section 3.4), and the simplified English resources (Figure 1, B) are written as normalization files, from which pattern and normalization files for all languages are derived (Figure 1, i-iv).

For terms directly occurring in English rules, new pattern files are created, and all English patterns are written without regular expressions. While this makes the resources longer (original patterns with regular expressions are expanded to many patterns), these simplified resources are necessary for a smooth translation process.

3.4 Language-independent Rules

Based on HeidelTime's English resources, we also write completely language-independent rules for date, time, duration, and set expressions. In addition, we add "creative" rules that are not necessary for English expressions but might match expressions in other languages. In particular, we allow several orderings of patterns and sometimes even random tokens in between. For instance, instead of just a "month day" pattern (e.g., to match "January 13"), we add "day month", "day X month" and "month X day", with "X" matching any token (e.g., for the Spanish translation "13 de enero").

In the extraction and normalization parts of the rules, we refer to the names of the language-independent and simplified English pattern and normalization files. Note that in contrast to the original rules, these rules do neither contain pos constraints nor any English terms to guarantee language independence. However, fuzzy pattern matching is allowed at the end of some patterns to try to take care of morphology-rich languages.

3.5 Creating Resources for All Languages

As shown in Figure 1, the patterns of the simplified English resources are translated to all languages in the world using Wiktionary as translation resource. For each normalization file, the respective pattern and normalization files for each language in Wiktionary are created (i).[3] Then,

---

[3]Note that Wiktionary exists in about 170 languages but contains translations to many more languages; for details, see http://en.wiktionary.org/wiki/Wiktionary:Statistics.

language: corpus / domain	HeidelTime 1.9 (manual) relaxed extr			value		HeidelTime – automatic relaxed extr			value	
	P	R	F1	F1	acc.	P	R	F1	F1	acc.
English: TE-3 TimeBank / news (UzZaman et al., 2013)	93.1	90.8	91.9	79.6	86.5	95.6	49.2	64.9	54.7	84.3
English: TE-3 platinum / news (UzZaman et al., 2013)	93.1	88.4	90.7	78.1	86.1	98.7	56.5	71.9	54.4	75.7
English: WikiWars / narrative (Mazur and Dale, 2010)	98.3	86.1	91.8	83.1	90.5	97.9	58.4	73.2	53.4	73.0
Arabic: Arabic test-50* / news (Strötgen et al., 2014)	90.9	90.9	90.9	82.2	90.4	91.7	31.8	47.2	38.0	80.5
Chinese: TE-2 test impro. / news (Li et al., 2014)	95.8	89.3	92.4	79.5	86.0	100	9.5	17.3	7.6	44.0
Croatian: WikiWarsHR / narrative (Skukan et al., 2014)	92.6	90.5	91.5	80.8	88.3	87.3	6.8	12.6	9.7	77.0
French: FR-TimeBank / news (Bittar et al., 2011)	91.9	90.1	91.0	73.6	80.9	87.2	59.5	70.8	54.6	77.1
German: WikiWarsDE / narrative (Strötgen and Gertz, 2011)	98.7	89.3	93.8	83.0	88.5	98.4	64.7	78.1	59.7	76.4
Italian: EVALITA'14 test / news (Caselli et al., 2014)	92.7	86.1	89.3	75.0	84.0	98.5	41.2	58.1	49.3	84.9
Spanish: TempEval-3 test / news (UzZaman et al., 2013)	96.0	84.9	90.1	85.3	94.7	95.5	53.8	68.8	58.5	85.0
Vietnamese: WikiWarsVN / narrative (Strötgen et al., 2014)	98.2	98.2	98.2	91.4	93.1	84.0	45.5	59.0	27.1	45.9
Portuguese: PT-TimeBank test / news (Costa and Branco, 2012)	87.3	75.9	81.2	63.5	78.2	91.5	59.3	72.0	59.4	82.5
Portuguese: PT-TimeBank train / news (Costa and Branco, 2012)	83.3	73.1	77.9	54.5	70.0	88.2	51.0	64.6	50.4	78.0
Romanian: Ro-TimeBank / news (Forascu and Tufis, 2012)	–	–	–	–	–	31.9	11.4	16.9	7.8	46.2

Table 1: Evaluation results for several languages on public corpora. HeidelTime 1.9 results as reported on `https://github.com/HeidelTime/heideltime/wiki/Evaluation-Results`.

the patterns are extracted from the simplified English normalization resources (ii) and the respective Wiktionary pages are accessed (iii). These contain translation information that is then added to the pattern and normalization files of the occurring languages by extracting respective translations (iv).

Note that for ambiguous words, Wiktionary contains one translation table for each meaning. For instance, there are two translation tables for "November" for the meanings "eleventh month of the Gregorian calendar" and "N in the ICAO spelling alphabet".[4] Obviously, it is crucial to use only the translations of the intended meaning. Thus, in the case of multiple meanings of English patterns, we provide the names of Wiktionary's correct translation tables to our translation script, so that only translations of the correct meaning are generated. In addition, if patterns in the new language start with a lower case character, we allow upper case and lower case for the respective character as shown in Figure 1, (iv).

Unfortunately, Wiktionary is of course not complete and does not contain translations to all languages for all English patterns. However, to allow backtracking, we add the English information to the resource files of all languages. If, at a later point, the automatically created resources are used as basis for developing better resources for specific languages, missing translations can be easily detected and manually added.

---

[4] `http://en.wiktionary.org/wiki/November#Translations`.

3.6 Improving and Finalizing the Resources

In particular for the rule development process, we also create the automatic resources for English. These are repeatedly evaluated on the temporally annotated English corpora TimeBank (Pustejovsky et al., 2003), TempEval-3 platinum (UzZaman et al., 2013), and WikiWars (Mazur and Dale, 2010). Analyzing the errors on these corpora, we iteratively improve the simplified English resources and the rules. In Table 1, we compare HeidelTime's original English resources (HeidelTime 1.9) with the simplified resources (HeidelTime – automatic). While details about evaluation measures will be described in Section 4, the results of the automatic approach are obviously lower, as expected due to the simplification process.

Resources for all languages are finally built by combining the language-independent resources, the automatically created pattern and normalization files, and the language-independent rules. Note that depending on the completeness of Wiktionary, the coverage of the resources – and thus also their quality – varies between languages.

4 Evaluation

For a subset of languages, temporally annotated corpora exist. Thus, we can directly evaluate our approach for these languages. For German, Spanish, Italian, French, Portuguese, Croatian, Arabic, Vietnamese, and Chinese, we compare our automatic approach with HeidelTime's manually developed resources. In addition, first results are reported for Romanian.
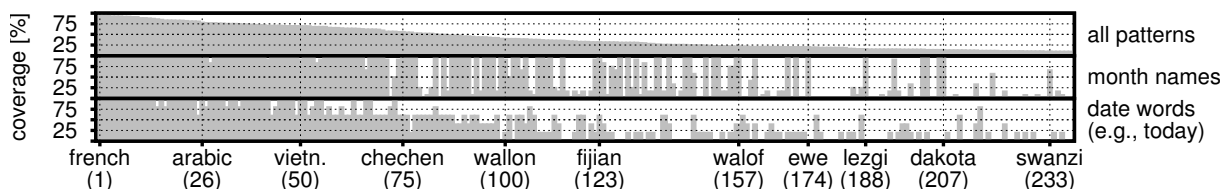
Figure 2: Language completeness statistics showing how many of the simplified English patterns have translations in Wiktionary for the 238 languages with at least 10% coverage (ordered based on "all patterns" coverage). A detailed plot with all language names is available on HeidelTime's GitHub page.

For all other languages, it is obviously difficult to judge the temporal tagging quality of the newly developed resources. However, for an estimation of the quality for languages without temporally annotated corpora, we provide completeness statistics of the pattern translations. The fewer translations are available for a language, the more likely it is that the temporal tagging quality is rather low.

4.1 Evaluation Measures

Since the task of temporal tagging is two-fold, we report precision, recall, and F1-score for the extraction task, and for the full task (relaxed extraction plus value normalization) *value F1*. Similar as the TempEval-3 organizers (UzZaman et al., 2013), we consider *value F1* with relaxed matching as most important. However, when processing large amounts of data, *value accuracy* becomes also important. Since it directly shows independent of the recall if extracted expressions are normalized correctly, it is reported, too.

4.2 Evaluation Results and Discussion

As shown in Table 1, the results of our automatically created resources are worse than those of HeidelTime with manually developed resources. For Spanish, German, French, and Portuguese, high precision, moderate recall, and good normalization results are achieved. In contrast, recall for Chinese, Croatian, and Romanian are quite low, precision is low for Romanian, and normalization accuracy for Chinese, Vietnamese and Romanian. Clearly, more work is necessary to achieve temporal tagging quality that can be considered as applicable for other applications. In particular, issues with morphology-rich languages and those without white-space tokenization have to be addressed.

Nevertheless, the results look promising for several languages taking into account that the resources are developed without any language-specific information, and that our approach has

been the first step towards more sophisticated temporal tagging of all languages in the world. Note that recall is much worse than precision for all languages and that the normalization of extracted expressions works quite well (*value acc.*).

4.3 Completeness Statistics

To estimate the quality of our automatically created resources for languages without annotated corpora, we create completeness statistics. As Figure 2 shows, the completeness differs significantly between languages. However, Wiktionary translations exist for at least 75% and 50% of the patterns for 34 and 83 languages, respectively, and even for many languages with low overall coverage, important patterns such as month names and date words are well-covered. Note that even for languages with few translations, our approach can be considered as baseline since temporal tagging of most languages has not been addressed before.

5 Conclusions and Ongoing Work

In this paper, we presented our approach to automatically extend the temporal tagger HeidelTime to all languages in the world. We thus establish new temporal tagging baselines for many languages, and anyone working on any of those languages in the future can either use our resources as a starting point or as a baseline for comparison. The resources for 200+ languages are available with HeidelTime 2.0 at `http://github.com/HeidelTime/heideltime`.

Currently, we are improving the simplified English resources and the translation process by exploiting further resources such as Wikipedia. In the future, we thus plan to constantly update HeidelTime's automatically created resources.

Acknowledgments

References

Gabor Angeli and Jakob Uszkoreit. 2013. Language-Independent Discriminative Parsing of Temporal Expressions. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL '13)*, pages 83–92. ACL.

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 (ACL '11)*, pages 130–134. ACL.

Ricardo Campos, Gaël Dias, Alípio M. Jorge, and Adam Jatowt. 2014. Survey of Temporal Information Retrieval and Related Applications. *ACM Computing Surveys*, 47(2):15:1–15:41.

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 (LAW '11)*, pages 143–151. ACL.

Tommaso Caselli, Rachele Sprugnoli, Manuela Speranza, and Monica Monachine. 2014. EVENTI. EValuation of Events and Temporal INformation at Evalita 2014. In *Proceedings of the Forth International Workshop EVALITA*, pages 27–34.

Angel X. Chang and Christopher D. Manning. 2012. SUTime: A Library for Recognizing and Normalizing Time Expressions. In *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC '12)*, pages 3735–3740. ELRA.

Francisco Costa and António Branco. 2012. TimeBankPT: A TimeML Annotated Corpus of Portuguese. In *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC '12)*, pages 3727–3734. ELRA.

Corina Forascu and Dan Tufis. 2012. Romanian TimeBank: An Annotated Parallel Corpus for Temporal Information. In *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC '12)*, pages 3762–3766. ELRA.

Hui Li, Jannik Strötgen, Julian Zell, and Michael Gertz. 2014. Chinese Temporal Tagging with HeidelTime. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL '14)*, pages 133–137. ACL.

Hector Llorens, Estela Saquete, and Borja Navarro. 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. ACL.

Giulio Manfredi, Jannik Strötgen, Julian Zell, and Michael Gertz. 2014. HeidelTime at EVENTI: Tuning Italian Resources and Addressing TimeML's Empty Tags. In *Proceedings of the Forth International Workshop EVALITA*, pages 39–43.

Pawel Mazur and Robert Dale. 2009. The DANTE Temporal Expression Tagger. In *Proceedings of the 3rd Language and Technology Conference (LTC '07)*, pages 245–257. Springer.

Pawel Mazur and Robert Dale. 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. ACL.

Pawel Mazur. 2012. *Broad-coverage Rule-based Processing of Temporal Expressions*. Ph.D. thesis, Macquarie University and Wroclaw University of Technology.

Véronique Moriceau and Xavier Tannier. 2014. French Resources for Extraction and Normalization of Temporal Expressions with HeidelTime. In *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC '14)*, pages 3239–3243. ELRA.

Matteo Negri, Estela Saquete, Patricio Martínez-Barco, and Rafael Muñoz. 2006. Evaluating Knowledge-based Approaches to the Multilingual Extension of a Temporal Expression Normalizer. In *Proceedings of the Workshop on Annotating and Reasoning about Time and Events (ARTE '06)*, pages 30–37. ACL.

J. Pustejovsky, P. Hanks, R. Saurí, A. See, R. Gaizauskas, A. Setzer, D. Radev, B. Sundheim, D. Day, L. Ferro, and M. Lazo. 2003. The TIMEBANK Corpus. In *Proceedings of Corpus Linguistics 2003*, pages 647–656.

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

Estela Saquete, Patricio Martínez-Barco, and Rafael Muñoz. 2004. Automatic Multilinguality for Time Expression Resolution. In *Proceedings of the Third Mexican International Conference on Artificial Intelligence (MICAI '04)*, pages 458–467.

Luka Skukan, Goran Glavaš, and Jan Šnajder. 2014. HeidelTime.Hr: Extracting and Normalizing Temporal Expressions in Croatian. In *Proceedings of the 9th Language Technologies Conference (LTC '14)*, pages 99–103. Information Society.

Kathrin Spreyer and Anette Frank. 2008. Projection-based Acquisition of a Temporal Labeller. In *Proceedings of the 3rd International Joint Conference on Natural Language Processing (IJCNLP '08)*, pages 489–496. ACL.

Jannik Strötgen and Michael Gertz. 2011. Wiki-WarsDE: 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 '11)*, pages 129–134.

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

Jannik Strötgen, Julian Zell, and Michael Gertz. 2013. HeidelTime: Tuning English and Developing Spanish Resources for TempEval-3. In *Proceedings of the 7th International Workshop on Semantic Evaluation (SemEval '13)*, pages 15–19. ACL.

Jannik Strötgen, Ayser Armiti, Tran Van Canh, Julian Zell, and Michael Gertz. 2014. Time for More Languages: Temporal Tagging of Arabic, Italian, Spanish, and Vietnamese. *ACM Transactions on Asian Language Information Processing (TALIP)*, 13(1):1:1–1:21.

Naushad UzZaman, Hector Llorens, Leon Derczynski, James F. Allen, Marc Verhagen, and James Pustejovsky. 2013. SemEval-2013 Task 1: TempEval-3: Evaluating Time Expressions, Events, and Temporal Relations. In *Proceedings of the 7th International Workshop on Semantic Evaluation (SemEval '13)*, pages 1–9. ACL.

Matje van de Camp and Henning Christiansen. 2012. Resolving Relative Time Expressions in Dutch Text with Constraint Handling Rules. In *Proceedings of the 7th International Workshop on Constraint Solving and Language Processing (CSLP '12)*, pages 74–85. Springer.

Marc Verhagen, Roser Saurí, Tommaso Caselli, and James Pustejovsky. 2010. SemEval-2010 Task 13: TempEval-2. In *Proceedings of the 5th International Workshop on Semantic Evaluation (SemEval '10)*, pages 57–62. ACL.