Procesamiento del Lenguaje Natural, Revista nº 59, septiembre de 2017, pp. 15-22 recibido 30-03-2017 revisado 08-05-2017 aceptado 18-05-2017

EusHeidelTime: Time Expression Extraction and Normalisation for Basque

EusHeidelTime: extracción y normalización de expresiones temporales para el euskera

Begoña Altuna, María Jesús Aranzabe, Arantza Díaz de Ilarraza
Universidad del País Vasco/Euskal Herriko Unibertsitatea
Manuel Lardizabal, 1, 20018 Donostia
{begona.altuna,maxux.aranzabe,a.diazdeilarraza}@ehu.eus

Abstract: Temporal information helps to organise the information in texts placing the actions and states in time. It is therefore important to identify the time points and intervals in the text, as well as what times they refer to. We developed EusHeidelTime for Basque time expression extraction and normalisation. For it, we analysed time expressions in Basque, we created the rules and resources for the tool and we built corpora for development and testing. We finally ran an experiment to evaluate EusHeidelTime's performance. We achieved satisfactory results in a morphologically rich language.
Keywords: Time expressions, information extraction, normalisation

Resumen: La información temporal ayuda a organizar la información textual situando las acciones y los estados en el tiempo. Por eso, es importante identificar los puntos e intervalos temporales en el texto, así como los tiempos a los que estos se refieren. Hemos desarrollado EusHeidelTime para la extracción y normalización de expresiones temporales para el euskera. Para ello, hemos analizado las expresiones temporales en euskera, hemos creado las reglas y recursos para la herramienta y hemos construido un corpus para el desarrollo y la evaluación. Finalmente, hemos realizado un experimento para evaluar el rendimiento de EusHeidelTime. Hemos conseguido resultados satisfactorios en una lengua con morfología rica.
Palabras clave: Expresiones temporales, extracción de información, normalización

1 Introduction

Temporal information is a core resource for textual organisation as it structures the discourse along a temporal axis. Its extraction and normalisation is useful and relevant in text comprehension and generation for tasks such as text summarisation (Aramaki et al., 2009), chronology creation (Bauer, Clark, and Graepel, 2015), event prediction (Radinsky and Horvitz, 2013) and event forecasting (Kawai et al., 2010).

Temporal information is composed by the events that happen or occur, the times those events happen in and the relations among those events and times. However, in this work we focus on time expression processing. Time expressions refer to a point in time in which an event takes place, starts or ends, or the duration of an event. For time expression processing, time expressions in texts must be marked and normalised and their features are extracted following a mark-up scheme. The corpora annotated with temporal information can be used for training machine-learning systems or as a gold standard to evaluate the performance of the tools.

Many tools and resources were developed to fulfill the task of identifying and normalising temporal information. On one hand, mark-up languages for temporal information annotation and annotated corpora were created, *e.g.* TimeML (TimeML Working Group, 2010), which was taken as an annotation standard and the TimeBank corpus (Pustejovsky et al., 2006). On the other hand, systems for temporal information extraction and normalisation were developed employing: i) machine-learning methods, e.g. GUTime (Verhagen and Pustejovsky, 2008) and TIPSem (Llorens, Saquete, and Navarro, 2010) ii) rule-based approaches such as CTEMP (Wu et al., 2005) and HeidelTime (Strötgen and Gertz, 2013) and iii) hy-

ISSN 1135-5948

brid tools, for example, TempEX (Mani and Wilson, 2000) and KTX (Jang, Baldwin, and Mani, 2004).

For our experimentation, we analysed Basque time expressions (Section 2), we created the EusTimeBank annotated gold standard corpus (Section 3), we integrated the HeidelTime parser in the Basque processing pipeline and we adapted and created the linguistic resources the system needs (Section 4). Finally, we conducted an annotation experiment (Section 5) and an error analysis for the evaluation of our tool's performance (Section 6). Some final remarks are given in Section 7.

2 *Time expressions in Basque*

We analysed Basque time expressions following (Bittar, 2010) and we have identified five different time expression types:

- dates: expressions referring to a particular period based on the Gregorian calendar, *e.g. martxoaren 8a* (8^{th} of March).
- times: expressions that refer to a particular subdivision of the day, *e.g. bostak* (five o'clock).
- durations: these expressions refer to an extended period of time *e.g. hiru aste* (three weeks).
- frequencies: these constructions express the regularity or re-occurrence of an event *e.g. egunero* (every day).
- temporal quantifications: expressions that consist in the quantification of a temporal unit *e.g. egunean 8 ordu* (8 hours a day).

These time expressions are classified in TimeML in four categories: date, time, duration and set (for frequencies and temporal quantifications). All time expressions are annotated with `TIMEX3` tag in TimeML and its features are normalised by means of a `DATE`, `TIME`, `DURATION` or `SET` type attribute, an ISO-8601 normalised value, as well as other attributes.

We annotated the time expressions in Basque following the EusTimeML guidelines[1], the adaptation of TimeML for Basque, which were used for the annotation of the sentence in (1) as can be seen in Figure 1.

[1] `https://addi.ehu.es/handle/10810/17305`

The time expression (*Iaz*, Last year) appears along with its class (`DATE`) and normalised value (`2016`). An event (*fakturatu zituzten*, turned over) is also displayed as well as the relation between the time expression and the event: the event is included in the time point the time expression refers to.

(1) Iaz 1.167 milioi euro
Last.year 1,167 million euro
fakturatu zituzten.
turn.over 3.PL.PAST
Last year they turned over 1,167 million euros.

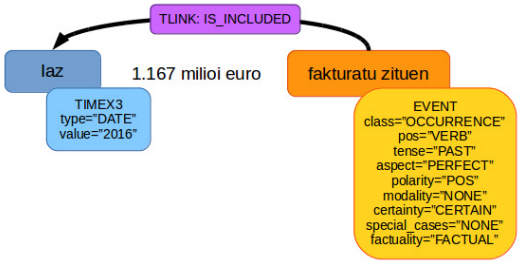

Figure 1: Annotation of example (1) following EusTimeML

An extended description of an annotation process for time expressions is described in Altuna, Aranzabe and Díaz de Ilarraza (2014).

3 *EusTimeBank*

EusTimeBank is a corpus that contains temporal information. It is composed by three subcorpora:

- **FaCor**: a 25 news document corpus on the closure of a company written originally in Basque.
- **WikiWarsEU**: this corpus contains the corresponding Basque Wikipedia articles on 17 of the 20 wars in WikiWars (Mazur and Dale, 2010). The documents are historical texts and have been written by non professional authors or translators.
- **EusMEANTIME**: it is the translation to Basque of the MEANTIME Corpus (Minard et al., 2016), which contains 120 economy news documents.

The documents were manually annotated using the CELCT Annotation Tool (Bartalesi

Lenzi, Moretti, and Sprugnoli, 2012) and following the EusTimeML mark-up scheme. A selection of 67 documents was used for development and evaluation purposes of the temporal information processing tools we created: 25 from FaCor, 17 from WikiWarsEU and 25 from EusMEANTIME. We provide the amount of `TIMEX3` tags, time expression tag, and the size of the annotated corpora for the experiment in Table 1.

Corpora	Size	
	Development (words/TIMEX3)	Test (words/TIMEX3)
FaCor	4,503/142	1,513/59
EusMEANTIME	5,247/200	1,258/53
WikiWarsEU	22,299/701	7,399/343
TOTAL	**32,049/1043**	**10,170/455**

Table 1: Size of the annotated corpora

4 *The EusHeidelTime tool*

We adapted HeidelTime for Basque time expression extraction and normalisation due to the re-usability of the source code and the easiness for linguistic resource creation, as well as the lack of large annotated corpora in Basque. The rules, patterns and normalisation information are language dependent, while the source code is common to all languages. This allows an easy adaptation to new languages. Apart from English, HeidelTime was used for time expression extraction and normalisation in German (Strötgen and Gertz, 2011), Dutch (van de Camp and Christiansen, 2013), French (Moriceau and Tannier, 2014) and Croatian (Skukan, Glavaš, and Šnajder, 2014) among others.

4.1 Integration of EusHeidelTime in the Basque pipeline

HeidelTime was originally developed as a UIMA (Unstructured Information Management Architecture) (Ferrucci and Lally, 2004) component and integrated as a document processing pipeline. As explained in Strötgen and Gertz (2010), for English, the UIMA pipeline contains a sentence splitter and tokenizer and an OpenNLP PoS tagger to be used by the temporal tagger. For Basque, instead, we defined and integrated the temporal tagger in a document processing pipeline, ixa-pipe-pos-eu, following the Otegi et al. (2016) approach. More specifically, our pipeline (Figure 2) includes, a tokeniser, a robust and wide-coverage morphological analyser and a PoS tagger for Basque and the EusHeidelTime temporal tagger. Ixa-pipe-pos-eu is part of ixaKat[2], a modular chain of NLP tools for Basque where all the modules read and write NAF (Fokkens et al., 2014), a linguistic annotation format designed for complex NLP pipelines. The temporal tagger has these features too, but the core of the module is based on HeidelTime. Thus, the integration of the temporal tagger in a UIMA pipeline would be quite straightforward. In addition, we parametrised the temporal tagger so that it is possible to obtain the temporal information in NAF or TimeML format (Figure 3), which was used for the evaluation of the tool. TimeML format implies XML documents containing XML `TIMEX3` tags that mark time expressions and offer information about their type, normalised value and modifier information if any.

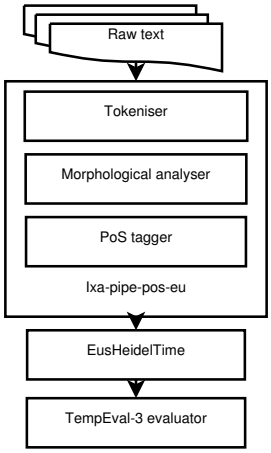

Figure 2: Diagram for time expression extraction in Basque

```
<?xml version="1.0" ?>
<TimeML xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance" xsi:noNamespaceSchemaLocation="http://
timeml.org/timeMLdocs/TimeML_1.2.1.xsd">
<DOCID>1380-World_largest_passenger_
airliner_makes_first_flight.txt.xml</DOCID>

<DCT><TIMEX3 tid="t0" type="TIME"
value="2005-04-27" temporalFunction="false"
functionInDocument="CREATION_TIME">2005-04-27
</TIMEX3></DCT>

<TEXT>
Munduko bidaiari-hegazkinik handienak estreinako
hegaldia egin du . <TIMEX3 type="DATE"
value="2005-04-27" tid="t1">2005eko apirilaren 27a</
TIMEX3> . A380 hegazkina <TIMEX3 type="DATE"
value="2005-01" tid="t2">2005eko urtarrilean</
TIMEX3> aurkeztu zuten .
```

Figure 3: An EusHeidelTime annotation example

[2]http://ixa2.si.ehu.es/ixakat/

4.2 Adapting language dependent resources

As mentioned before, we adapted HeidelTime to Basque. For this, we created three resource sets:

- **Rules**: rules contain the patterns to be extracted and their normalisation, as well as value modifiers and constraints, *e.g.* part-of-speech (PoS) constraint of a token in the pattern. Figure 4 shows a rule for patterns as "Datorren urteko urtarrilean" (On January next year). The rule contains a name (`RULENAME`), the pattern to match (`EXTRACTION`) and the normalisation pattern (`NORM_VALUE`) that will turn the text segment into a TimeML normalised value. There are four rule sets (dates, durations, sets and times), which correspond to the different types of time expressions in EusTimeML[3].
- **Patterns**: pattern resources are regular expressions that gather together patterns of the same kind, *e.g.* months, weekdays etc.
- **Normalisation files**: these contain normalised values of the time expressions. Figure 5 shows weekdays and the normalised value for each string.

For the development of resources, two main features of Basque were taken into account. First, as Basque is agglutinative, the rich morphology as well as the morphotactics were added. Second, since it is a head-final language, many acquired patterns were reversed to accommodate its syntax. As a consequence, some resources, namely a significant quantity of rules, were created from scratch to accommodate specific Basque temporal constructions. Nonetheless, some rules and patterns for Basque (mainly numeric expressions) were directly transferred from other languages and most of the patterns (*e.g.* month names, weekday names) were translated.

Apart from the relevant linguistic features, the internal architecture of HeidelTime was also taken into account. HeidelTime applies the rules sequentially and when more than one rule matches a time expression,

[3]Rules for intervals were disregarded as intervals are not defined in EusTimeML.

```
//Adibidea: datorren urteko urtarrilean

RULENAME="Data_erl_datorren_year_month",
EXTRACTION="%reDatorren urte%Singularra
%reMonth%reSingularra",
NORM_VALUE="UNDEF-next-year-%normMonthFull(group(4))"
```

Figure 4: An EusHeidelTime rule

```
"[Aa]stelehen","1"
"[Aa]stearte","2"
"[Aa]steazken","3"
"[Oo]stegun","4"
"[Oo]stiral","5"
"[Ll]arunbat","6"
"[Ii]gande","7"
```

Figure 5: Weekday pattern normalisation values

it chooses one following this order: dates, times, durations, sets and intervals[4]. The rules in each category are also ordered and read sequentially.

In Table 2 one can see the amount of resources created for EusHeidelTime. The quantity of rules is due to i) the intention to avoid optional elements in the rules, and ii) grammatical aspects of Basque as word order restrictions with the numeral determiner *bat* (one). This led to defining two different rules for strings containing numerals.

Resource type	Quantity
Rules	
DATE	142
TIME	64
DURATION	101
SET	6
Pattern files	58
Normalisation files	29

Table 2: EusHeidelTime resources

5 Experimentation

We processed a 17 document set of the test corpora (Section 3) and we evaluated the output against our gold standard annotation to evaluate the developed resources[5]. We followed the TempEval-3 (UzZaman et al., 2013) criteria to evaluate the performance of our tool. In Table 3 we present the results for each corpus in these four fields:

- Strict match: the extent of the obtained temporal expression and the correspond-

[4]We do not apply rules for intervals in Basque since they are not a category in EusTimeML.

[5]EusHeidelTime resources and corpora for replication can be downloaded from http://ixa2.si.ehu.es/eusheideltime/

	FaCor			**EusMEANTIME**			**WikiWarsEU**		
	P	R	F1	P	R	F1	P	R	F1
Strict match	79.39	83.64	81.42	81.4	74.47	77.78	77.98	87.8	82.6
Relaxed match	87.93	92.73	90.27	93.02	85.11	88.89	82.67	93.09	87.57
Value			58.41			64.44			74.57
Type			83.19			82.22			86.81

Table 3: Evaluation results for EusHeidelTime

ing one in the gold standard overlap perfectly.

- Relaxed match: partial overlap between the automatically obtained expression and the corresponding one in the gold standard.
- Value: the normalised value of the automatically obtained and the gold standard match.
- Type: the type of the automatically obtained and the gold standard match.

For strict and relaxed matches, precision (P), recall (R) and F-measure (F1) were calculated and for value and type the F-measure was given, in order to be comparable to the TempEval-3 results.

The performance of our tool is in the same range of the best systems for English in TempEval-3. We achieved a F1 of 81.42 for strict match in FaCor and 82.6 in WikiWarsEU, which are close to the best performing tool in TempEval-3, ClearTK-1,2 (Bethard and Martin, 2013) (82.71) and HeidelTime for English (81.34). In what concerns the relaxed match, for which we achieved a F1 score of 90.27 in FaCor corpus, we also get close to the best performing tools, NavyTime-1,2 and SUTime (90.32) and HeidelTime (90.30).

We also got similar results for news (FaCor and EusMEANTIME) and for historical texts (WikiWarsEU). Nevertheles, a high rise on the F1 for value (74.57) can be seen for historical texts, presumably because of the large amount of the absolute dates.

6 Error analysis

We conducted an analysis to identify the nature of the different errors. We classified manually the errors in 8 categories (Table 4) and we tried to solve them.

As one can see from Table 4, the errors identified are quite heterogeneous, but can be divided in human-made and processing

Error	**Quantity**
Absence rule	24
Too general rules	21
Wrong gold standard	6
Wrong tokenisation	11
Wrong rule selection	18
Wrong resolution of relative date	21
Rule not performing well	18
Ambiguous reference	6

Table 4: Classification of errors

errors. The first group is formed by i) the absence of rules for certain time expressions. *E.g.*, "hondarrean" in (2) is not a common term to express the end and we did not consider it when creating the rules; and ii) the too general rules led to false positives. For example, we created restricting rules to treat polysemy as in "urri" (October/scarce), "hil" (month/dead) and "lehen" (past/first) among others, but they proved not to be sufficient. Finally, iii) the errors in the gold standard, mainly typos. These rules can be fixed by adding or correcting the rules and the errors in the gold standard. However, we are aware that we will not be able to address all the possible time expressions in Basque.

(2) `gold annotation: <TIMEX3 type="DATE" value="2014-07" tid="t15">uztailaren hondarrean</TIMEX3> system annotation: <TIMEX3 type="DATE" value="2014-07" tid="t15">uztailaren</TIMEX3> -- relaxed match`

In what concerns errors due to processing, we first noticed the errors due to wrong tokenisation. In example (3), the initials "(UTC)" were wrongly tokenised and this impaired the time expression from being identified although a rule for times containing "(UTC)" existed.

(3) `gold annotation not found in system: <TIMEX3 type="TIME"`

value="2008-09-18T08:00Z" tid="t2">8:00etan (UTC)</TIMEX3>

In what refers to rule selection, as mentioned in section 4.2, the rules are applied sequentially and there is a hierarchy between categories. This has led to the wrong rule selection. In example (4) we got a partial match since the system privileged a date rule instead of a duration rule.

(4) system annotation: <TIMEX3 type="DATE" value="PAST_REF" tid="t7">lehen</TIMEX3> gold annotation: <TIMEX3 type="DURATION" value="PT90M" tid="t7">lehen 90 minutuetan </TIMEX3> -- relaxed match

We also identified some rules not performing well. "Gaur" in example (5) is annotated as a generic present reference, although it refers to the exact date of "today". Both interpretations are possible, but HeidelTime systematically chooses the generic interpretation although the exact one is higher in hierarchy. This may be due to a mistake in the rule and needs further analysis.

(5) system wrong value: <TIMEX3 type="DATE" value="PRESENT_REF" tid="t5">gaur</TIMEX3>

Some annotation errors are much more difficult to correct. Those are the ones that i) involve relative time expressions or ii) ambiguous constructions that can only be resolved through world knowledge or a deep contextual comprehension. For the first, HeidelTime sets the last time expression annotated as a temporal anchor for the next. In example (6) the value for "bihar bertan" (tomorrow) is not well resolved as the temporal anchor is not the right one. The solution for the second is much more complicated because of the difficulty of adding world or contextual knowledge to automatic systems. It is virtually impossible to decide the real duration of "Epe laburreko" (short period) (7), since a short period can be considered hours, days or months in different contexts.

(6) gold value: <TIMEX3 type="DATE" value="2014-10-31" tid="t8">bihar bertan</TIMEX3> system wrong value: <TIMEX3 type="DATE" value="2014-10-28" tid="t8">bihar bertan</TIMEX3>

(7) system wrong value: <TIMEX3 type="DURATION" value="PXD" tid="t3">epe laburreko</TIMEX3>

After the error analysis, we will improve the rules correcting the problems identified.

7 Conclusions and future work

In this paper we presented an experiment on temporal expression annotation in Basque with EusHeidelTime, a rule-based tool based on HeidelTime. Considering Basque is a highly agglutinative and head-final language, we proved that HeidelTime can be used for languages with complex morphology. We also profited the modularity of HeidelTime and we added it to our Basque pipeline.

EusHeidelTime achieved results comparable to those obtained for English. Having reached F1 measures of circa 80% in strict match, we consider the resources created for this experiment are already adequate for the automatic annotation of temporal expressions, although that annotation will have to be supervised by a human annotator. We will also proceed to a final tuning of the resources to correct the flaws identified during the error analysis.

We achieved similar results both in news and historical documents. Therefore, we presume our tool can annotate documents of different domains. In the future, we aim to perform temporal annotation of clinical texts, due to the relevance of the temporal ordering of events in that field.

Temporal expression extraction and normalisation is only a part of a more extended work on temporal information annotation. It will be combined with event information processing and temporal relation processing for the creation of a system able to treat temporal information in its entirety.

Acknowledgments

This work was financed by the Basque Government scholarship PRE_2016_2_294.

References

Altuna, B., M. J. Aranzabe, and A. Díaz de Ilarraza. 2014. Euskarazko denbora-egiturak. Azterketa eta etiketatze-

esperimentua. *Linguamática*, 6(2):13–24, Dezembro.

Aramaki, E., Y. Miura, M. Tonoike, T. Ohkuma, H. Mashuichi, and K. Ohe. 2009. Text2table: Medical text summarization system based on named entity recognition and modality identification. In *Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing*, BioNLP '09, pages 185–192, Stroudsburg, PA, USA. Association for Computational Linguistics.

Bartalesi Lenzi, V., G. Moretti, and R. Sprugnoli. 2012. CAT: the CELCT Annotation Tool. In N. Calzolari, K. Choukri, T. Declerck, M. U. Doğan, B. Maegaard, J. Mariani, J. Odijk, and S. Piperidis, editors, *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, pages 333–338, Istanbul, Turkey. European Language Resources Association (ELRA).

Bauer, S., S. Clark, and T. Graepel. 2015. Learning to Identify Historical Figures for Timeline Creation from Wikipedia Articles. In L. Aiello and D. E. McFarland, editors, *SocInfo 2014 International Workshops, Barcelona, Spain, November 11, 2014, Revised Selected Papers*, volume 8852 of *Lecture Notes in Computer Science*, pages 234–243, Barcelona, Spain.

Bethard, S. and J. H. Martin. 2013. ClearTK-TimeML: A minimalist approach to TempEval 2013. In S. Manandhar and D. Yuret, editors, *Second Joint Conference on Lexical and Computational Semantics (*SEM) 2: Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, pages 10–14, Stroudsburg, PA, USA. Association for Computational Linguistics.

Bittar, A. 2010. *Building a TimeBank for French: a Reference Corpus Annotated According to the ISO-TimeML Standard*. Ph.D. thesis, Université Paris Diderot, Paris.

Ferrucci, D. and A. Lally. 2004. UIMA: an architectural approach to unstructured information processing in the corporate research environment. *Natural Language Engineering*, 10(3–4):327–348.

Fokkens, A., A. Soroa, Z. Beloki, N. Ockeloen, G. Rigau, W. R. van Hage, and P. Vossen. 2014. NAF and GAF: Linking linguistic annotations. In *Proceedings 10th Joint ISO-ACL SIGSEM Workshop on Interoperable Semantic Annotation*, page 9, Reykjavik, Iceland.

Jang, S. B., J. Baldwin, and I. Mani. 2004. Automatic TIMEX2 Tagging of Korean News. *ACM Transactions on Asian Language Information Processing (TALIP)*, 3(1):51–65, March.

Kawai, H., A. Jatowt, K. Tanaka, K. Kunieda, and K. Yamada. 2010. Chronoseeker: Search engine for future and past events. In *Proceedings of the 4th International Conference on Uniquitous Information Management and Communication*, ICUIMC '10, pages 25:1–25:10.

Llorens, H., E. Saquete, and B. 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, Stroudsburg, PA, USA. Association for Computational Linguistics.

Mani, I. and G. Wilson. 2000. Robust Temporal Processing of News. In *Proceedings of the 38th Annual Meeting on Association for Computational Linguistics*, pages 69–76, Stroudsburg, PA, USA.

Mazur, P. and R. 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, Stroudsburg, PA, USA. Association for Computational Linguistics.

Minard, A.-L., M. Speranza, R. Urizar, B. na Altuna, M. van Erp, A. Schoen, and C. van Son. 2016. MEANTIME, the NewsReader Multilingual Event and Time Corpus. In N. Calzolari, K. Choukri, T. Declerck, S. Goggi, M. Grobelnik, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, and S. Piperidis, editors, *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*,

Paris, France, may. European Language Resources Association (ELRA).

Moriceau, V. and X. Tannier. 2014. French Resources for Extraction and Normalization of Temporal Expressions with HeidelTime. In N. Calzolari, K. Choukri, T. Declerck, H. Loftsson, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, and S. Piperidis, editors, *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC 2014)*, Reykjavik, Iceland, may. European Language Resources Association (ELRA).

Otegi, A., N. Ezeiza, I. Goenaga, and G. Labaka. 2016. A Modular Chain of NLP Tools for Basque. In P. Sojka, A. Horák, I. Kopeček, and K. Pala, editors, *Proceedings of the 19th International Conference on Text, Speech and Dialogue — TSD 2016, Brno, Czech Republic*, volume 9924 of *Lecture Notes in Artificial Intelligence*, pages 93–100. Springer International Publishing.

Pustejovsky, J., M. Verhagen, R. Saurí, J. Littman, R. Gaizauskas, G. Katz, I. Mani, R. Knippen, and A. Setzer. 2006. TimeBank 1.2. Technical report, Linguistic Data Consortium.

Radinsky, K. and E. Horvitz. 2013. Mining the web to predict future events. In *Proceedings of the sixth ACM international conference on Web search and data mining*, pages 255–264. ACM.

Skukan, L., G. Glavaš, and J. Šnajder. 2014. HeidelTime.Hr: Extracting and Normalizing Temporal Expressions in Croatian. In *Proceedings of the Nineth Language Technologies Conference*, pages 99–103. Information Society.

Strötgen, J. and M. Gertz. 2010. HeidelTime: High Quality Rule-based Extraction and Normalization of Temporal Expressions. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, SemEval '10, pages 321–324, Stroudsburg, PA, USA. Association for Computational Linguistics.

Strötgen, J. and M. Gertz. 2011. WikiWarsDE: a German Corpus of Narratives Annotated with Temporal Expressions. In H. Hedeland, T. Schmidt, and K. Wörner, editors, *Multilingual Resources and Multilingual Applications. Proceedings of the Conference of the German Society for Computational Linguistics and Language Technology (GSCL) 2011*, pages 129–134, Hamburg University.

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

TimeML Working Group. 2010. TimeML Annotation Guidelines version 1.3. Manuscript. Technical report, Brandeis University.

UzZaman, N., H. Llorens, J. F. Allen, L. Derczynski, M. Verhagen, and J. Pustejovsky. 2013. TempEval-3: Evaluating Events, Time Expressions, and Temporal Relations. In S. Manandhar and D. Yuret, editors, *Second Joint Conference on Lexical and Computational Semantics (*SEM) 2: Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, volume 2, pages 1–9, Stroudsburg, PA, USA. Association for Computational Linguistics.

van de Camp, M. and H. Christiansen. 2013. Resolving relative time expressions in Dutch text with Constraint Handling Rules. In *Revised Selected Papers of the 7th International Workshop on Constraint Solving and Language Processing - Volume 8114*, CSLP 2012, pages 166–177, New York, NY, USA. Springer-Verlag New York, Inc.

Verhagen, M. and J. Pustejovsky. 2008. Temporal Processing with the TARSQI Toolkit. In *22n d International Conference on on Computational Linguistics: Demonstration Papers*, COLING '08, pages 189–192, Stroudsburg, PA, USA. Association for Computational Linguistics.

Wu, M., W. Li, Q. Lu, and B. Li. 2005. CTEMP: A Chinese Temporal Parser for Extracting and Normalizing Temporal Information. In R. Dale, K.-F. Wong, J. Su, and O. Y. Kwong, editors, *Natural Language Processing – IJCNLP 2005: Second International Joint Conference, Jeju Island, Korea, October 11-13, 2005. Proceedings*, pages 694–706, Berlin, Heidelberg. Springer.