LX-TimeAnalyzer: A Temporal Information Processing System for Portuguese

Francisco Costa and António Branco

DI–FCUL–TR–2012–01

DOI:10455/6796

(http://hdl.handle.net/10455/6796)

February 2012

LX-TimeAnalyzer: A Temporal Information Processing System for Portuguese

Francisco Costa and António Branco

University of Lisbon, Lisbon
fcosta@di.fc.ul.pt
Antonio.Branco@di.fc.ul.pt

Abstract. This paper presents LX-TimeAnalyzer, a tool that extracts the temporal information conveyed by a text and annotates that text with that information. The system recognizes terms that denote events, dates and times and makes explicit the temporal relations that hold between these elements. The interest in this kind of task is quite recent, and there are not many full temporal information systems reported about, nor are there many languages for which temporal information systems have been built. LX-TimeAnalyzer is the first of its kind for Portuguese, and its performance is similar to the state-of-the-art temporal systems for other languages.

1 Introduction

Extracting the temporal information present in a text is relevant to many Natural Language Processing applications, including question-answering, information extraction, and even document summarization, as summaries may be more readable if the information is presented in chronological order.

Recent evaluation challenges have focused on the extraction of temporal information from written text in terms of the temporal information they convey. TempEval [14], in 2007, and more recently TempEval-2 [15], in 2010, were concerned with this problem while providing data that can be used to develop and evaluate systems that can automatically temporally tag natural language text.

Figure 1 shows a small fragment of the data from TempEval. There, terms denoting events, such as the event of releasing the tapes that is described in that text, are annotated using `EVENT` tags, and temporal expressions, such as *today*, are enclosed in `TIMEX3` tags. The attribute `value` of time expressions holds a normalized representation of the date or time they refer to (e.g. the word *today* denotes the date `1998-01-14` in this example). The `TLINK` elements at the end describe temporal relations between events and temporal expressions. For instance, the event of the plane going down is annotated as temporally preceding the date denoted by the temporal expression *today*.

The first TempEval challenge focused solely on the temporal relations. TempEval-2 additionally included tasks related to the identification and normalization of event terms and temporal expressions. Identification is concerned with classifying each word in a text as to whether it is an event term or part of a temporal

```
<s>In Washington <TIMEX3 tid="t53" type="DATE" value="1998-01-14" temporalFunction="true"
functionInDocument="NONE">today</TIMEX3>, the Federal Aviation Administration<EVENT eid="e1"
class="OCCURRENCE" stem="release" aspect="NONE" tense="PAST" polarity="POS"
pos="VERB">released</EVENT> air traffic control tapes from <TIMEX3 tid="t54" type="TIME"
value="1998-XX-XXTNI" temporalFunction="true" functionInDocument="NONE">the night</TIMEX3>
the TWA Flight eight hundred <EVENT eid="e2" class="OCCURRENCE" stem="go" aspect="NONE"
tense="PAST" polarity="POS" pos="VERB">went</EVENT> down.</s>
<TLINK lid="l1" relType="BEFORE" eventID="e2" relatedToTime="t53"/>
<TLINK lid="l2" relType="OVERLAP" eventID="e2" relatedToTime="t54"/>
```

Fig. 1. Sample of the data annotated for TempEval, corresponding to the fragment: *In Washington today, the Federal Aviation Administration released air traffic control tapes from the night the TWA Flight eight hundred went down.*

expression or none of these. Normalization is related to determining the value of the various attributes of `EVENT` and `TIMEX3` elements, specially the `value` attribute of `TIMEX3` elements. Putting together all these tasks, it is possible to fully annotate raw text with temporal information (event terms, temporal expressions and temporal relations) in a way similar to what is shown in Figure 1.

More recently an adaptation of these data has been developed for Portuguese [5], making it possible to train and evaluate temporal taggers for this language. In this paper we present work pursuing this goal.

2 Related Work

The two TempEval challenges were mainly concerned with automatically determining the type of a given temporal relation (`OVERLAP`, `BEFORE`, etc.), and not so much with fully tagging raw text with all these temporal annotations. Nevertheless, TempEval-2 included additional tasks that focused precisely with the identification and normalization of temporal expressions and events. Task A of TempEval-2 was concerned with temporal expressions: determining their extent and the value of the attributes `type` (whether it denotes a calendar date, a clock time, a duration or a set of dates or times) and `value` (its normalized value). Task B focused on event terms: their extent in a text and the value of the attributes `tense` (morphological tense if the event term is a verb, the value `NONE` otherwise), `aspect` (if it is a verb, whether it is part of constructions relevant to time, such as the progressive), `polarity` (whether the event term occurs in a negated syntactic context), and `class` (this attribute includes some information about the semantic class of event terms, distinguishing `REPORTING`, `PERCEPTION` and `ASPECTUAL` terms from the others, and also includes some aspectual distinctions in the spirit of [13, 6], distinguishing `STATE` situations from non-stative events, marked as `OCCURRENCE`s). Table 1 shows the scores obtained by the best participant for each of these problems. The evaluation measures used were the f-measure for the problems of identifying the extents of event and time expressions and accuracy for the tasks dealing with the attributes. Full details can be found in [15].

Task	Scores	
	English	Spanish
Temporal expressions		
Extents	0.86	0.91
`type`	0.98	0.99
`value`	0.85	0.83
Events		
Extents	0.83	0.88
`class`	0.79	0.66
`tense`	0.92	0.96
`aspect`	0.98	0.89
`polarity`	0.99	0.92

Table 1. Best system results for the various identification and normalization tasks of TempEval-2.

Before TempEval, the 2004 Temporal Expression Recognition and Normalization evaluation (TERN 2004—`http://timex2.mitre.org`) and the Automatic Content Extraction (ACE) evaluations [11] had been contemplating the problem of temporal expression identification in the context of named entity recognition.

TERN 2004 was much more limited than TempEval and TempEval-2: it was exclusively about temporal expressions, and there was no concern about events or temporal relations. The systems at TERN 2004 were evaluated against a test corpus of around 50,000 words. The TempEval-2 systems were evaluated on a much smaller data set: around 53,450 words for training and only 9,613 words for testing.

The best performing system at TERN 2004 was [10]. This system achieved a 0.93 f-measure for the problem of identifying temporal expressions. With respect to the normalization of temporal expressions, this system scored between 0.69 and 0.87 for the various attributes of the elements used to tag temporal expressions, with the highest score (0.87 f-measure) being seen with respect to the `val` attribute.[1] The authors used a symbolic approach. It starts by tagging the input text with parts-of-speech. The expressions are then detected and normalized by a set of over 1000 hand-crafted rules.

Since TERN 2004, machine learning approaches have been able to match these results. [1] replaces the large set of hand-crafted rules typical of systems for this task by a series of machine learned classifiers and a much smaller set of rules. The system begins with parsed documents as input. The sentences are parsed with the constituency parser of [3]. The strategy is then to use a succession of several machine learned classifiers to arrive at the final result. First, a machine learned classifier decides which of the phrases detected by the parser are temporal expressions. The classifier features used include character type

[1] This attribute `val` is identical to the `value` attribute used in the TempEval annotation scheme.

patterns, a context window of two words to the left, and syntactic information produced by the parser. Another classifier then assigns a semantic class to the recognized timexes. These classes distinguish *inter alia* durations, dates, and times. The system proceeds in this spirit, with many classifiers, each performing a small subtask of the problem. As a result, at the end only a small set of 89 hand-crafted rules is used for normalization.

The system was evaluated with the data of TERN 2004. It achieved an f-measure of 0.87 for the identification of temporal expressions, not very far from the best system of TERN 2004, that scored 0.93. The f-measure for correctly assigning the `val` attribute of a temporal expression is 0.89, compared to the 0.87 score for the best system at TERN 2004 (which was fully rule-based).

For Portuguese, the second HAREM [9] challenge of named entity recognition included a track for temporal expressions. The data used for HAREM included 14,056 words and 193 normalized temporal expressions. It covered both recognition and normalization, and the best system was XIP [7]: for recognition, the system obtained an f-measure of 0.76; for normalization the f-measure was 0.74. This system is rule based.

3 Approach and Evaluation

The data that was used for the first TempEval has recently been adapted to Portuguese, as reported in [5]. The documents that make up this corpus were translated to Portuguese, and the annotations adapted to the language. Figure 2 shows a fragment of the data, corresponding to the English fragment presented above in Figure 1. The training subset contains 68,351 words, 6,790 events, 1,244 temporal expressions and 5,781 temporal relations. The evaluation set has 9,829 words, 1,097 events, 165 temporal expressions and 758 temporal relations.

```
<s>Em Washington, <TIMEX3 tid="t53" type="DATE" value="1998-01-14" temporalFunction="true"
functionInDocument="NONE">hoje</TIMEX3>, a Federal Aviation Administration <EVENT eid="e1"
class="OCCURRENCE" stem="publicar" aspect="NONE" tense="PPI" polarity="POS"
pos="VERB">publicou</EVENT> gravações do controlo de tráfego aéreo da <TIMEX3 tid="t54"
type="TIME" value="1998-XX-XXTNI" temporalFunction="true"
functionInDocument="NONE">noite</TIMEX3> em que o voo TWA800 <EVENT eid="e2"
class="OCCURRENCE" stem="cair" aspect="NONE" tense="PPI" polarity="POS"
pos="VERB">caiu</EVENT>.</s>
<TLINK lid="l1" relType="BEFORE" eventID="e2" relatedToTime="t53"/>
<TLINK lid="l2" relType="OVERLAP" eventID="e2" relatedToTime="t54"/>
```

Fig. 2. Sample of the Portuguese training data, corresponding to the fragment:
Em Washington, hoje, a Federal Aviation Administration publicou gravações do controlo de tráfego aéreo da noite em que o voo TWA800 caiu.

It must be noted that: (i) the Portuguese data are an adaptation of the English data used in the first TempEval, (ii) the results in Table 1 refer to

TempEval-2, (iii) the English data of TempEval and TempEval-2 are not identical, although there is a large overlap between them. For the data of the first TempEval there are unfortunately no published results that we know of concerning the identification and normalization of temporal expressions and event terms, as TempEval focused only on temporal relations. It is thus important to note that our results are not fully comparable to the results for English (and they are even less comparable to the results for Spanish, as they are based on completely different data).

In any case, these data allow for the training and evaluation of temporal processing systems for Portuguese. LX-TimeAnalyzer is a first attempt at full temporal processing of this language. Figure 3 shows the general architecture of LX-TimeAnalyzer.

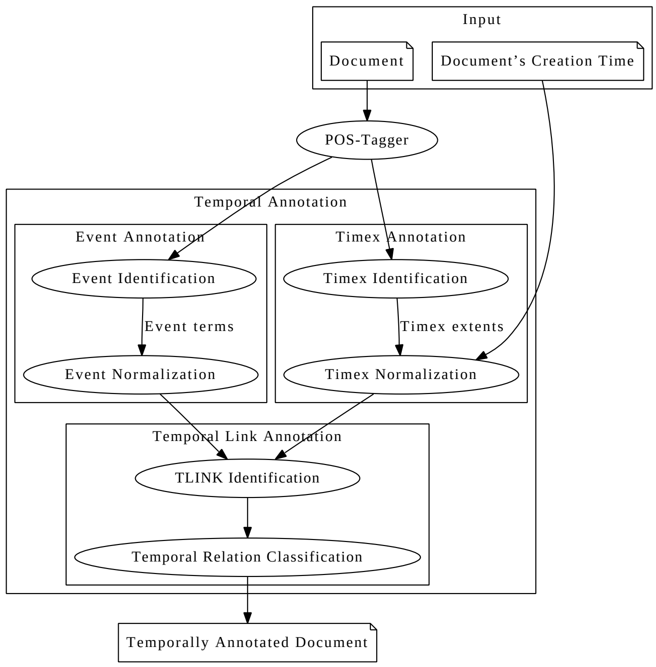

Fig. 3. Architecture of LX-TimeAnalyzer

In Table 2 we include information about the performance of our system, evaluating each subtask that was evaluated in TempEval-2 (with the exception of temporal relation classification, which is reported elsewhere). As already noted, the figures are not quite comparable to those of TempEval-2, since the data and the languages are different.

In the following sections, the more interesting components of the system are described.

Task	Score
Temporal expressions	
Extents	0.85
`type`	0.91
`value`	0.81
Events	
Extents	0.72
`class`	0.74
`tense`	0.95
`aspect`	0.96
`polarity`	0.99

Table 2. Evaluation on the test data. The evaluation measures used were the f-measure for the problems of identifying the extents of event and time expressions and accuracy for the tasks dealing with the attributes.

3.1 Part-of-Speech Tagging

The document to be processed is initially tagged with a part-of-speech tagger and morphological analyzer [2]. This tool annotates each word with, *inter alia*, its part-of-speech category (noun, verb, etc.), its lemma (i.e. its dictionary form), and a tag with information about inflection. This information is then used in the subsequent phases of processing.

3.2 Event Identification

A simple solution to identifying event terms in text is to classify each word as to whether it denotes an event or not. This strategy is not very efficient, since some very frequent words cannot possibly denote events (e.g. determiners, conjunctions etc.). Figure 4 shows the distribution of parts-of-speech for event terms, according to the training data. 92% of all event terms are verbs or nouns. Nevertheless, we followed this simple approach.

The classifier features we employed are the following:

- **Features about the last characters of the lemma**
  A Boolean attribute represents whether the lemma ends in one of several suffixes from a hand-crafted list. This list includes suffixes such as *mento*. The motivation is that this information may be useful especially to separate eventive nouns from non-eventive nouns. There are additional attributes that include information about the last two characters of the lemma and the last three characters of the lemma; they are intended to capture suffixes not covered by the list of suffixes.
- **The part-of-speech and the inflection tag assigned by the tagger.**
  As shown in Figure 4, information about part-of-speech can rule out many words in a document. The inflection tag may also be relevant. For instance, even though singular forms are more common than plural forms for both

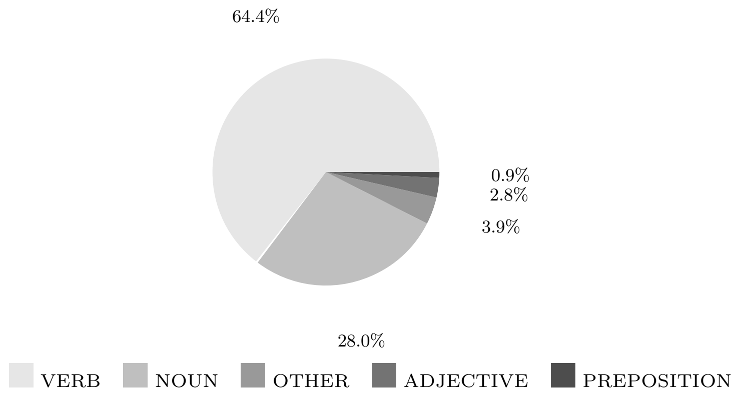

Fig. 4. Part-of-speech distribution of event terms

eventive and non-eventive nouns, this difference is sharper in the case of eventive nouns (since these denote multiple or repeated events), as shown in Table 3.

– **The part-of-speech and the inflection tag of: the preceding word token, the following word token, the preceding word token bigram, the following word token bigram**.
  These attributes are used in order to capture some contextual information.
– **Whether the preceding token was classified as an event**
  The intuition is that adjacent event terms are infrequent.

Nouns	Eventive Count (%)	Non-eventive Count (%)
Singular	1,433 (74)	6,101 (59)
Plural	491 (26)	4,177 (41)
Total	1,924	10,278

Table 3. Distribution of singular and plural forms of eventive vs. non-eventive nouns in the training data

Training a decision tree with these attributes (we used Weka's [16] implementation of the C4.5 algorithm) on the training data results in a classifier with an f-measure of 0.72 evaluated on the test data. This is somewhat worse than the best systems of TempEval-2 for both English (0.83) and Spanish (0.88). These systems followed a similar approach to ours, but they used additional classifier attributes based on the output of a syntactic parser (we also tried this, and it did not improve the results) and WordNet (which we were unable to experiment with). We believe that the information taken from WordNet is probably the major cause of the differences, as the structure of WordNet can be used to determine event terms (e.g. there is a synset for *event* "something that happens at a given place and time").

3.3 Event Normalization

This step is concerned with the annotation of the several attributes appropriate for `<EVENT>` elements., as described in Section 2.

The values of many of the attributes of `<EVENT>` elements are already provided by the morphological analyzer: `stem`, `tense` and `pos`. Three attributes are not, however: `aspect`, `polarity` and `class`.

A principled annotation of the `polarity` attribute (which encodes whether the event occurs in a positive or negative context) requires syntactic parsing. Nevertheless, we tried to simply check whether one of the three preceding words is a negative word—*não* "not", *nunca* "never", *ninguém* "nobody", *nada* "nothing", *nenhum/nenhuma/nenhuns/nenhumas* "no, none", *nenhures* "nowhere"—and there is no other event intervening between this n-word and the event that is being annotated. On the test data, the accuracy of this simple heuristic is 0.99, which is identical to the best score in TempEval-2 for English (0.99) and better than the one for Spanish (0.92).

In the Portuguese data, the attribute `aspect` only encodes whether the verb form is part of a progressive construction. This attribute is also computed symbolically, and the implementation simply checks for gerund forms (e.g. *fazendo*) or constructions involving an infinite verb form immediately preceded by the preposition *a* (*a fazer*). On the evaluation data, its accuracy is 0.96.

The most interesting problem of event normalization is determining the value of the `class` attribute of `<EVENT>` elements. It is also the hardest, with the best system for English showing a 0.79 accuracy in TempEval-2 and the best one for Spanish showing only 0.66. This attribute of `EVENT` elements encodes information about aspectual type (see Section 2), which is sensitive to both lexical and contextual (i.e. syntactic) information. For this attribute, a specific classifier was trained (also a decision tree, with Weka). This classifier takes advantage of a very minimal set of features:

- **The lemma of the event term being classified**
  This type of information is highly lexicalized, so it is expected that the lemma of the word token can be quite informative.
- **Contextual features**
  These attributes encode the part-of-speech of the previous word and that of the next word, and the following bigram of inflection tags.

We experimented with more features, similar to the ones used for event detection, but they did not improve the results. We obtained a result with 0.74 accuracy.

3.4 Temporal Expression Identification

In order to identify temporal expressions, we trained a classifier that, to each word in the text, assigns one of three labels: `B` (begin), `I` (inside), `O` (outside).

We once again used a decision tree classifier and the features employed were:

- **Features about the current token**
  This includes the token's part-of-speech and its inflection tag. Additionally, there is an attribute that checks whether the current token's lemma is part of a list of temporal adverbs. This is specially useful for the `B(egin)` class, which is the one with the highest error rate.
- **Features about the previous token and features about the following one**
  Once again these features are taken from the morphological analyzer and encode part-of-speech and inflection tag.
- **The classification for the previous token**
  This is relevant because tokens classified as `I(nside)` cannot directly follow tokens classified as `O(utside)`.
- **Whether there is white space before the current token and the previous one**
  The reason behind this attribute is to treat punctuation and special symbols in a special manner (they are tokenized separately; e.g. a time expression of the form `XXXX-XX-XX` is tokenized into five word tokens).
- **Whether (i) the current token's lemma was seen in the training data at the beginning of a temporal expression, or (ii) it was seen inside a temporal expression, or (iii) the bigram of lemmas formed by the current token's lemma and the next one's was seen inside a temporal expression**
  Instead of using an attribute encoding the lemma directly, we used a series of Boolean attributes capturing distinctions that are expected to help classification.

As shown in Table 2, this component shows an f-measure of 0.85 for the `B(egin)` and `I(nside)` classes.

3.5 Temporal Expression Normalization

This problems consists in identifying the value of the `TIMEX3` attributes `type` and `value`.

LX-TimeAnalyzer solves this problem symbolically. The normalization rules take as input the following parameters:

- The word tokens composing the temporal expression, and their morphological annotation
- The document's creation time
- An anchor. This is another temporal expression that is often required for normalization. For instance, an expression like *the following day* can only be normalized if its anchor is known. We use the previous temporal expression that occurs in the same text and that is not a duration. This simple heuristic is similar to previous approaches found in the literature.

- The broad tense (present, past, or future) of the closest verb in the sentence where it occurs, with the distance being measured in number of word tokens from either boundary of the time expression. For example, all past tenses are treated as *past*. This is used for instance in order to decide whether a time expression like *February* refers to the previous or the following month of February (relative to the document's creation time).

These rules are not implemented in a dedicated format, they are simply implemented by a Java method. It takes approximately 1600 lines of code and is recursive: e.g. when normalizing an expression like *terça de manhã* "Tuesday morning", the expression *terça* "Tuesday" is normalized first, and then its normalized `value` is changed by appending `TMO` (with `T` being the time separator and `MO` the way to represent the vague expression "morning"); its `type` is also changed from `DATE` to `TIME`. The same method fills in both the `value` and the `type` attributes of `TIMEX3` elements.

This implementation was conducted by looking at the examples in the training data, and additionally to a small set (c. 5000 words) of news reports taken from on-line newspapers.

3.6 Temporal Relation Classification

Besides annotated temporal expressions and event terms, the data sets of the two TempEval challenges also have annotated temporal relations between these elements. The major tasks of these two challenges were in fact related to determining the type of these temporal relations: given an ordered pair of temporal elements (two events or one event and one time, date or duration), the participating systems had to classify the type of the temporal relation holding between them as `OVERLAP`, `BEFORE` or `AFTER`. LX-TimeAnalyzer includes one component to classify these temporal relations, following a machine-learning approach. Details on the classifiers developed for this task and integrated in LX-TimeAnalyzer are presented in [4].

4 Concluding Remarks

This paper describes a temporal tagger for Portuguese. Given a text document, it annotates it with information about events, dates and times, and temporal relations holding between these elements. It is the first full temporal tagger for Portuguese. It seems to perform in line with the state-of-the-art for other languages, although (i) the data used for evaluation are not fully comparable, and (ii) event detection is somewhat worse, but can possibly be improved by incorporating information similar to that in WordNet.

Full temporal information processing is fairly recent. Only in the TempEval-2 challenge, last year in 2010, were there systems capable of fully annotating raw text with temporal information (e.g. [12, 8]), similarly to LX-TimeAnalyzer. The languages handled then were English and Spanish.

References

1. Ahn, D., van Rantwijk, J., de Rijke, M.: A cascaded machine learning approach to interpreting temporal expressions. In: Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Proceedings of the Main Conference. pp. 420–427. Association for Computational Linguistics, Rochester, New York (April 2007)
2. Branco, A., Silva, J.: A suite of shallow processing tools for portuguese: LX-Suite. In: Proceedings of the 11^{th} Conference of the European Chapter of the Association for Computational Linguistics (EACL'06). Trento, Italy (2006)
3. Charniak, E.: A maximum-entropy-inspired parser. In: Proceedings of the 1^{st} North American chapter of the Association for Computational Linguistics conference. pp. 132–139 (2000)
4. Costa, F.: Processing Temporal Information in Unstructured Documents. Ph.D. thesis, Universidade de Lisboa, Lisbon (to appear)
5. Costa, F., Branco, A.: Temporal information processing of a new language: Fast porting with minimal resources. In: ACL2010—Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (2010)
6. Dowty, D.R.: Word Meaning and Montague Grammar: the Semantics of Verbs and Times in Generative Semantics and Montague's PTQ. Reidel, Dordrecht (1979)
7. Hagège, C., Baptista, J., Mamede, N.: Reconhecimento de entidades mencionadas com o XIP: Uma colaboração entre o INESC-L2f e a Xerox. In: Mota, C., Santos, D. (eds.) Desafios na avaliação conjunta do reconhecimento de entidades mencionadas: O Segundo HAREM., pp. 261–274. Linguateca (2008)
8. Llorens, H., Saquete, E., Navarro, B.: TIPSem (English and Spanish): Evaluating CRFs and semantic roles in TempEval-2. In: Erk, K., Strapparava, C. (eds.) SemEval 2010—5^{th} International Workshop on Semantic Evaluation—Proceedings of the Workshop. pp. 284–291. Uppsala University, Uppsala, Sweden (2010)
9. Mota, C., Santos, D. (eds.): Desafios na avaliação conjunta do reconhecimento de entidades mencionadas: O Segundo HAREM. Linguateca (2008)
10. Negri, M., Marseglia, L.: Recognition and normalization of time expressions: ITC-irst at TERN 2004. Tech. rep., Trento (2004)
11. Strassel, S., Przybocki, M., Peterson, K., Song, Z., Maeda, K.: Linguistic resources and evaluation techniques for evaluation of cross-document automatic content extraction. In: Calzolari, N., Choukri, K., Maegaard, B., Mariani, J., Odjik, J., Piperidis, S., Tapias, D. (eds.) Proceedings of the Sixth International Language Resources and Evaluation (LREC'08). European Language Resources Association (ELRA), Marrakech, Morocco (may 2008)
12. UzZaman, N., Allen, J.F.: TRIPS and TRIOS System for TempEval-2: Extracting temporal information from text. In: Erk, K., Strapparava, C. (eds.) SemEval 2010—5^{th} International Workshop on Semantic Evaluation—Proceedings of the Workshop. pp. 276–283. Uppsala University, Uppsala, Sweden (2010)
13. Vendler, Z.: Verbs and times. Linguistics in Philosophy pp. 97–121 (1967)
14. Verhagen, M., Gaizauskas, R., Schilder, F., Hepple, M., Pustejovsky, J.: SemEval-2007 Task 15: TempEval temporal relation identification. In: Proceedings of SemEval-2007 (2007)
15. Verhagen, M., Saurí, R., Caselli, T., Pustejovsky, J.: SemEval-2010 task 13: TempEval-2. In: Strapparava, C., Erk, K. (eds.) SemEval 2010—5^{th} International Workshop on Semantic Evaluation—Proceedings of the Workshop. pp. 51–62. Uppsala University, Uppsala, Sweden (2010)

16. Witten, I.H., Frank, E.: Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations. Morgan Kaufmann, San Francisco (2005), second edition