Annotation of Events and Temporal Expressions in French Texts

André Bittar

Université Paris Diderot — ALPAGE

30 rue du Château des Rentiers

75013 Paris

France

`andre.bittar@linguist.jussieu.fr`

Abstract

We present two modules for the recognition and annotation of temporal expressions and events in French texts according to the TimeML specification language. The Temporal Expression Tagger we have developed is based on a large coverage cascade of finite state transducers and our Event Tagger on a set of simple heuristics applied over local context in a chunked text. We present results of a preliminary evaluation and compare them with those obtained by a similar system.

1 Introduction

TimeML (Pustejovsky et al., 2003) is a specification language for the annotation and normalization of temporal information in natural language texts. The annotation scheme allows for the annotation of events (`<EVENT>`), the tagging of temporal expressions and the normalization of their values (`<TIMEX3>`), as well as the annotation of temporal, aspectual and subordinating relations which may exist among them (`<TLINK>`, `<ALINK>` and `<SLINK>`, respectively). The linguistic markers of these relations can also be marked up (`<SIGNAL>`). A set of resources, including automatic and manual annotation tools and several reference corpora have been developed around the TimeML language. Evita (Saurí et al., 2005) is an application for automatically recognizing and annotating events in texts, based primarily on symbolic methods and linguistic data (input is a chunked text), although with some integration of statistical data. Its creators report precision of 74.03% and recall of 87.31% for an overall F-score of 80.12% for the task of event identification. GUTime (Mani and Wilson, 2000) annotates temporal expressions according to the TimeML schema and normalizes their values. The system achieves F-scores of 85%

and 82% for identification and normalization of temporal expressions, respectively. Further information is available on the TimeML project website[1].

2 A System for TimeML Annotation in French

(Parent et al., 2008) provide the description and evaluation of a system for the TimeML annotation of events and temporal expressions in French texts. The processing of temporal expressions is carried out on a text having undergone a part-of-speech tagging, morphological analysis and shallow syntactic analysis. The system functions by application of a cascade of 90 rules applied over 20 levels. Contrary to the Evita system developed for English, the event detection module relies on a full dependency parse as input for the event recognition task. The authors claim an advantage over chunker-based approaches with respect to the annotation of markable adjectives due to the fact that the dependency relation between copula verb and predicative adjective is available. The authors provide evaluation results according to grammatical category over a development corpus, made up of 35 biographical texts and 22 sports articles, and an evaluation ("unseen") corpus, consisting of an unspecified number of news articles from the website of the École Polytechnique de Montréal. The evaluation results, by grammatical category and global figures, are given in Table 1.

Cat	Development corpus			Evaluation corpus		
	Prec	Rec	F-sc	Prec	Rec	F-sc
Noun	61.5	40.0	48.4	54.7	53.7	54.2
Verb	94.1	97.3	95.7	65.6	90.9	76.2
Adj	66.7	77.8	71.8	N/A	N/A	N/A
Global	86.8	80.6	83.5	62.5	77.7	69.3

Table 1: Evaluation results according to corpora

The system performs best on the annotation of

---

event verbs and encounters the most difficulties in the annotation of event nominals. Adjectives are relatively well processed over the development corpus, but no adjectives were annotated by the human annotator in the evaluation corpus, so no results were calculated. As for the annotation of temporal expressions, precision is 83% and recall 79%, for an F-score of 81% over an evaluation corpus containing 544 human-annotated temporal expressions and an F-score of 50% for the normalization of values. These figures are comparable to those cited for GUTime for English.

3 Annotation Modules

In this section, we describe an annotation system, similar to that of (Parent et al., 2008) described above, although based on a rich cascade of finite state transducers and a shallow syntactic analysis, as opposed to a full dependency parse. The system is made up principally of two modules, the first tagging temporal expressions (section 3.1), the second identifying and annotating event expressions (section 3.2).

3.1 Temporal Expression Tagger

This module carries out the tagging and normalization of temporal expressions. The module consists of a large-coverage set of finiste state transducers developed with the Unitex[2] corpus processor. The transducers in this large-coverage grammar, applied to raw text, recognize patterns of dates, times, duration and frequency and tag expressions with the appropriately typed <TIMEX3>. A transducer matching expressions **not** to be marked up was also created. This transducer tags with the label <GARBAGE> expressions such as phone numbers, which could otherwise match numerical dates. The ambiguous word *été* (*been*/*summer*), when preceded by an adverb or the auxiliary verb *avoir* is tagged as <GARBAGE>, as it has its verb rather than noun reading in this context. Other expressions tagged as <GARBAGE> include the common expression *les 35 heures* (*the French 35 hour week*) and names of streets containing a date, such as *la place du 13 Mai*, etc. The normalization script, written in Perl, calculates the standard values of temporal expressions, including underspecified deictic expressions, and

---

[2]Unitex is a graphical corpus processing program, available for download under GNU General Public Licence at http://www-igm.univ-mlv.fr/~unitex/

removes annotations on items marked <GARBAGE>. The script consists of a set of substitution functions for each type of temporal expression tagged by the transducers. Each function converts the content of the expression into a TimeML standard value and inserts it in the `value` attribute of each <TIMEX3> tag. This module is available for download at http://www.linguist.univ-paris-diderot.fr/~abittar. This approach differs from that of (Parent et al., 2008) in that it relies almost entirely on lexical processing.

An evaluation was carried out on a subset of the corpus used to evaluate the similar module described in section 2. Our corpus consists of the 45 news articles from the Agence France Press used in the training and test sets described by (Parent et al., 2008). Figures for the evaluation are given in Table 2. The column labeled "Loose" represents the number of matches which cover an incomplete span of the expression, for example *un mois* (*one month*) instead of *un mois et demi* (*a month and a half*). The column "Strict" is for exact matches. The "Value" column represents the correctly normalized values for the temporal expressions detected, calculated over strict matches.

	Human	Found	Loose	Strict	Value
Number	592	575	508	484	317
Precision	-	-	85.8	84.2	55.0
Recall	-	-	88.4	81.8	44.9
F-score	-	-	87.1	83.0	49.4

Table 2: Evaluation results for the Temporal Expression Tagger

These figures are much in line with those of the system described in (Parent et al., 2008). Performance is slightly lower on loose matches (F-score 87.1 versus 91.0), but we achieve better results on strict matches (F-score 83.0 versus 81.0). This could be explained by the fact that we did not develop our grammar on the same type of source text, but shows that the grammar has a good coverage of the variants of each type of expression.

Sources of noise include age values tagged as durations, e.g. *M. Dupont, 58 ans* (*Mr. Dupont, 58 years old*) (11 errors), and numerical values taken to be years, e.g. *l'astéroïde 2001 UU92* (*Asteroid 2001 UU92*) (8 errors). Silence occurs mostly on coordinated date expressions or sequences, e.g. *les 4, 5 et 6 février* (*the 4th, 5th and 6th of February*) (11 errors) or expressions taking a

"vague" normalized value, e.g. *dans le passé* (*in the past*) (15 errors).

Results for the normalization of values for temporal expressions are practically identical to the other system for French. The majority of errors produced by our system (97 out of 167) are due to the fact that our normalization script does not yet fully deal with underspecified weekday expressions, such as *jeudi soir* (*Thursday evening*). In the hand-annotated corpus these expressions are fully resolved, with year, month and day values specified, e.g. `2002-01-15TEV`, whereas we provide a correct, but not completely resolved value, which specifies the day of the week, e.g. `2002-WXX-4TEV`. Excluding this difference in processing boosts precision to 73.6 and recall to 60.1 (F-score 66.85) for the normalization of values. We are currently working on fully normalizing these values.

3.2 Event Tagger

This module tags event expressions with the `<EVENT>` tag and classifies the events according to the ontology defined for TimeML. It also detects negative polarity contexts, as well as any aspectual or modal properties of certain verbal constructions. Input is a text having undergone part-of-speech tagging, an inflectional morphological analysis and shallow syntactic analysis, carried out by Macaon, a modular processing pipeline for French[3]. The Event tagger consists of several levels of processing - a layer of lexical processing, basically a lexical lookup for nouns and verb classes, and a layer of contextual processing consisting in the application of heuristics for detecting and eliminating event candidates and classifying them. This module relies on certain lexical resources. For the detection of event nominals, a lexicon containing nouns with at least one event interpretation is used. Many of the entries in this lexicon are ambiguous as they may also have a non-event interpretation. For example, *repas* (*meal*) has an object interpretation as well as an event reading. This highlights the need for disambiguation of nominals. The noun lexicon is based on the VerbAction lexicon (Hathout et al., 2002) which provided 9 200 unique deverbal noun lemmas. We further enriched the lexicon through semi-automated

search engine queries, such as *X a eu lieu* (*X took place*) and *lors du/de la/des X* (*during the X*), where X is likely to be an event nominal. An initial application of this method yielded 769 unique noun lemmas which were not in VerbAction - mostly rare or non-deverbal nouns, such as *anticoagulothérapie* (*anticoagulation therapy*) and *anniversaire* (*birthday*). The noun lexicon is of comparable size to that used in Evita.

We created by hand a verb lexicon which is used to perform classification of verbal events. It contains 200 lemmas for verbs in 6 of the 7 TimeML event classes[4]. Verbs were initially added to the lexicon by translating those proposed in the TimeML classifcation for English. The list of verbs was enriched by querying the dictionary of synonyms at the Université de Caen[5]. The lexicon is small for the time being and will need to be increased to ensure better coverage for classification. Like the noun lexicon, the lexicon of verbs contains ambiguities as certain verbs may belong to different classes or may not have an event reading in certain contexts. For example, the verb *expliquer* (*to explain*) belongs to the class `REPORTING` when it introduces a complementizer phrase in *que* (*that*) headed by an event (*Max a expliqué qu'il avait commis une erreur - Maca explained that he had made a mistake*). This is the class attributed by the lexicon. However, when it has a human subject and an event in object position (*Le manager a expliqué le renouvellemetn de l'équipe - the manager explained the renewal of the team*), it must be annotated with the class `I_ACTION`. Finally, if this verb has events in both subject and object position (*Le réchauffement climatique explique la fonte des calottes glacières - global warming explains the melting of the ice caps*), it is to be annotated with the class `CAUSE`. The system is thus confronted with the non-trivial problem of word sense disambiguation to identify the correct readings of nouns and verbs in the text. Initially, we tackle this problem for verbs with a number of heuristics, applied to local chunk context, for each of the TimeML verb classes in the lexicon. A total of 16 heuristics are used for choosing candidates for markup with the `<EVENT>` tag and 30 heuristics for classifying the events and determining values for the `aspect`, `modality` and `polarity` attributes. For example, in the case of the verb *expliquer* given above, the

---

[3]Macaon is freely available for download at http://pageperso.lif.univ-mrs.fr/~alexis.nasr/macaon/.

[4]As the class `OCCURRENCE` is the default class, it has no entries in the lexicon

[5]http://www.crisco.unicaen.fr/cgi-bin/cherches.cgi

heuristics include a search for the complementizer *que* in the chunk after the verb and a search for an event nominal chunk directly to the left of the verb chunk (approximation of subject position). Further heuristics are used to eliminate verbs and nouns which do not have an event reading. For example, event nominal chunks which do not have a determiner, such as in *prisonier de guerre* (*prisoner of war*), are not considered as candidates as they do not denote event instances, but rather event types, and cannot be attributed a specific temporal localisation. A set of heuristics is used to detect predicative adjectives, like in *Jean était malade* (*Jean was sick*), which are potential candidates for markup with the `<EVENT>` tag. For example, if the preceding verb is a copula, the adjective is flagged as a markable.

To evaluate our event tagger we used a corpus of 30 hand-annotated news articles from the newspaper *Le Monde*. The corpus was split into a development set of 20 documents (11 224 tokens, 1 187 `EVENT` tags) and a test set of 10 documents (5 916 tokens, 583 `EVENT` tags). Overall, the corpus contains 1 205 verbal, 471 nominal, 62 adjectival and 18 prepositional phrase `EVENT` tags.

	Development corpus			Evaluation corpus		
Category	Prec	Rec	F-sc	Prec	Rec	F-sc
Noun	50.2	94.5	72.4	54.0	95.1	74.5
Verb	87.7	92.3	90.0	86.5	91.1	88.8
Adjective	60.0	72.4	66.2	46.0	82.1	64.1

Table 3: Evaluation results for the Event Tagger

The results shown in Table 3 are fairly homogenous over both the development and test sets. The detection of event verbs performs slightly lower than that of the other system for French, although the evaluations were carried out on different corpora. For nominals, our system makes a vast improvement on the performance of the other system described in this paper (an F-score of 74.5 versus 54.2 over the respective test sets). The large-coverage lexicon of event nominals allows for a good recall, although precision remains low as more disambiguation is required to filter out nominals with non-event readings. Performance on adjectival events is lower than the other system, although not as bad as might have been expected. This is likely due to the difference in depth of syntactic analysis available to each system.

4 Conclusion

We have presented a comparative evaluation of two systems for the TimeML annotation of events and temporal expressions in French texts. Results show that a lexical approach to annotating temporal expressions performs generally just as well as an approach based on a shallow syntactic analysis. For event detection, the benefits of a full dependency parse are apparent, especially for the detection of markable adjectives, although comparable performance can be obtained with a chunked text as input. The benefits of a large-coverage lexicon for identifying event nominals are evident, although without effective disambiguation techniques precision remains very low. This is one point which requires particular attention and more elaborate guidelines for the annotation of event nominals would be of great value. Figures from the evaluation give a rough indication of performance across systems, however, a validated reference corpus for French is yet to be developed in order to give more meaningful comparisons. These are issues we are currently addressing.

References

Nabil Hathout, Fiammetta Namer, and Georgette Dal. 2002. An Experimental Constructional Database: The MorTAL Project. In Paul Boucher, editor, *Many Morphologies*, pages 178–209. Cascadilla, Somerville, Mass., USA.

Inderjeet Mani and George Wilson. 2000. Processing of news. In *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics*, pages 69–76, Hong Kong, October. Association for Computational Linguistics.

Gabriel Parent, Michel Gagnon, and Philippe Muller. 2008. Annotation d'expressions temporelles et d'événements en français. In *Actes de TALN 2008*, Avignon, France, June.

James Pustejovsky, José Casta no, Robert Ingria, Roser Saurí, Robert Gaizauskas, Andrea Setzer, and Graham Katz. 2003. TimeML: Robust Specification of Event and Temporal Expressions in Text. In *Proceedings of IWCS-5, Fifth International Workshop on Computational Semantics*.

Roser Saurí, Robert Knippen, Marc Verhagen, and James Pustejovsky. 2005. Evita: A Robust Event Recognizer for QA Systems. In *Proceedings of HLT/EMNLP 2005*, pages 700–707.