



HAL
open science

Representing and Visualizing Calendar Expressions in Texts

Delphine Battistelli, Javier Couto, Sylviane R. Schwer, Jean-Luc Minel

► **To cite this version:**

Delphine Battistelli, Javier Couto, Sylviane R. Schwer, Jean-Luc Minel. Representing and Visualizing Calendar Expressions in Texts. SYMPOSIUM ON SEMANTICS IN SYSTEMS FOR TEXT PROCESSING, Sep 2008, Venice, Italy. pp.1-7, 2008. halshs-00321379

HAL Id: halshs-00321379

<https://shs.hal.science/halshs-00321379v1>

Submitted on 30 Oct 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Representing and visualizing calendar expressions in texts

Delphine Battistelli¹, Javier Couto², Jean-Luc Minel³, Sylviane Schwer⁴

¹ LaLic, EA, Université Paris-Sorbonne,
75 006 Paris, France
Delphine.Battistelli@paris-sorbonne.fr

² INCO, FING, UdelaR, 565 Herrera y Reissig,
11300 Montevideo, Uruguay
jcouto@fing.edu.uy

³ MoDyCo, UMR 7114, CNRS-Université ParisX,
92 001 Nanterre, France
Jean-Luc.Minel@u-paris10.fr

⁴ LIPN, UMR 7030 CNRS-Université ParisXIII,
93430 Villetaneuse, France
Sylviane.Schwer@lipn.univ-paris13.fr

Abstract. In this paper, we address the study of temporal expressions that refer directly to text units concerning common calendar divisions, that we name "calendar expressions" (CE). First, we propose to model these expressions by defining operator algebra. Then, based on our model, a calendar view is built up by a plug-in designed in the platform NaviTexte.

1 Introduction

Taking into account temporality expressed in texts appears as fundamental not only in a perspective of global processing of documents but also in the analysis of the structure of a document. Since the theory of discursive frames [1], a lot of works has put the emphasis on the importance of the temporal expressions as modes of discursive organization. Nevertheless, the analysis of temporality within texts has been studied by considering principally verbal time and temporal adverbials.

Our approach is focused on temporal expressions, which can be put on a calendar. We name them 'calendar expressions' (CE for short). Our goal is to propose a system of representation, which is able to perform calculus on these CE, but especially, provides several ways to refer to calendar system. Our hypothesis is that contrary to most current studies, it is not necessary to systematically compute all calendar references, but only to make these calculations possible within the representation system.

In this paper, we propose a formal description of linguistic expressions, in French written texts, with regard to location in calendars, in order to propose a set of annotations which will be used to navigate temporally in an annotated document.

The remainder of this paper is organized as follows. In the next section, we describe several approaches of identification of calendar expressions. The third section describes an algebra operator and the fourth a software application, which exploits functional representation, built with such operators and plugged in NaviTexte platform. At last, conclusions and future works are presented.

2 Background and Related Work

So far, the process of annotation is split in two subtasks. One task identifies CE and the other one fixes them on a “temporal line”, using a set of values relying on ISO 8601 standard format [2,3,4] in order to provide some portability.

The automatic processing of the temporality relying on annotated texts is tackled from two levels of analysis and representation. The first one aims at fixing temporal expressions in the calendar system, the second one computes the temporal scheduling of events in the texts. Historically, it is the second one – yet, the more complex – which has been studied at first [5, 6]. The first level of analysis, regarded as the most feasible and necessary in the context of the development of Q/A systems [7] or summarizing multi-documents [8], became the most studied. From our point of view, these two issues are closely related and both refer at the same fundamental issue, the one that concerns interaction of different modes expressing the temporality in the language (grammatical tenses, adverbials locutions, connectors and typographical clues).

Nowadays, in the field of temporality processing, automatic identification and annotation tasks of EC are the most developed, mainly because identifying and annotating expressions which, contain calendar unit is considered – a priori – as a trivial task (see below on this point). A lot of works have been undertaken to solve this issue, TIMEML [9] being the most known. TIMEML is a proposal of a metalanguage to annotate events and their temporal relations in texts. It is made up of two schemes of annotations, TIDES TIMEX2 [10] and Sheffield STAG [3] put forward from the analysis of temporal adverbials.

3 A Calendar Expressions Algebra

We postulate that CE used to refer a calendar area can be described by a succession of operators which give access at this calendar area. Thus, CE is a composition of operators, each one giving a piece of the processing in the following order: from a calendar base (CB) taken as an argument, usually introduced by a determinant and/or a preposition, which works like a pointer, a second kind of operator expresses the useful part of this base (all, the beginning, the middle, the end, a fuzzy area around). Consequently, the time line is partitioned in three areas (cf. fig. 1): the former half-line (A), the Useful portion (U) and the posterior half-line (P), this one being limited by

the present moment. A third kind of operator gives access at the area described by the CE: either by qualitative shifting (like in “trois semaines plus tard” (three weeks later)), or by zooming (like in “cette année là” (the autumn of this present year)), or selecting one of the three portioned areas (like in “jusqu’en avril 2006” (until April 2006)).



Fig. 1. Partition of the time line for a unary CE

Our approach consists in modelling CE by defining algebra of operators. Moreover, there is an order in the application of operators. First, only one pointing operator *OpPointing*, followed by one or more shifting operators *OpShifting* (+) and finally only one zoning operator *OpZoning*. Consequently, the representation of CE has the following generic form: *OpZoning (OpShifting+(OpPointing (CB)))*.

Each operator is characterized by its arity (the number of its arguments) and type. With regard to the arity, in this paper we put the emphasis only on unary operators.

Three types of operators have been defined, pointing, shifting and zoning. The pointing operator is trivial but the two others need some refinements.

Several kinds of shifting time may be expressed. For instance, in the expression “au début de mai 2005” (at the beginning of may 2005), the shifting is localised inside the BC (mai 2005), whereas in the expression “trois semaines avant mai 2005 “ (three weeks before may 2005) it is outside of the BC. Consequently, six sub-operators have been identified [16].

Zoning operators refers at zones of the calendar, relatively at the BC, like for instance in the expression “avant le 21 avril 1997” (before the 21 of april 1997). Depending on prepositions, we have defined six kinds of zoning [16].

As mentioned before, it is necessary to use several N-ary operators to represent some CE. For instance, a binary operator is used for representing an expression like “entre fin mai 2005 and avril 2006” (between the end of may 2005 and april 2006). The study of these operators is under investigation.

3 Software Application

Many applications, which exploit temporal expressions in texts in particular in information extraction, have been designed [11]. Our application is plugged in the textual navigation workstation NaviTexte [12,13], in order to combine a traditional linear reading with a chronological one. With this intention, we have undertaken the construction of a computerized aided reading of biographies. Consequently, we have addressed two issues. First, identifying temporal expressions and ordering chronologically text segments in which they are included. Second, building calendar views of the text and navigating through these views.

From the linguistic study presented above, we have defined a set of annotations which are used to automatically annotate biographic texts. This process is carried out by transducers which put XML¹ annotations through the processed text. These annotations describe on the one hand, the granularity of CE, and on the other hand, the kind of identified operator. From these annotations, an automatic ordering relying on values of CE is carried out. It must be emphasized, however, that the computation is adequate only when disjointed CE can be linearly ordered. But, to process no disjointed CE, like “En juin 2007 (...) en été 2007” (in June 2007 (...) in summer 2007) a more powerful formalism is needed. We are working on formalism relying on S-Languages [14] to solve this point.

A new kind of view, a calendar one, has been built in the NaviTexte platform. This view is built from texts which contain CE annotated as described above. An example is shown in figure 3. Conceptually, a calendar view is a graph coordinated with a two-dimensional grid. In the left part of the view, lexical chains of various occurrences of CE in the text are displayed. By default, those are ordered according to their order of appearance in the text, but it is possible to display a chronological order, by using options offered in the panel located in bottom of the view. Nodes in the graph represent these lexical chains. The visual representation of a CE depends of the functional representation computed as described in section 3. A simple CE, with only a pointing operator like in “l’année 2008” (the year 2008) is always visualised like a white ellipse. An operator of focalising/Shifting like “la fin de” (the end of) selects an area of the ellipse and blackens it. Finally, a zoning operator like “avant” (before) is visualised by a bold line displaying the area that is referred to.

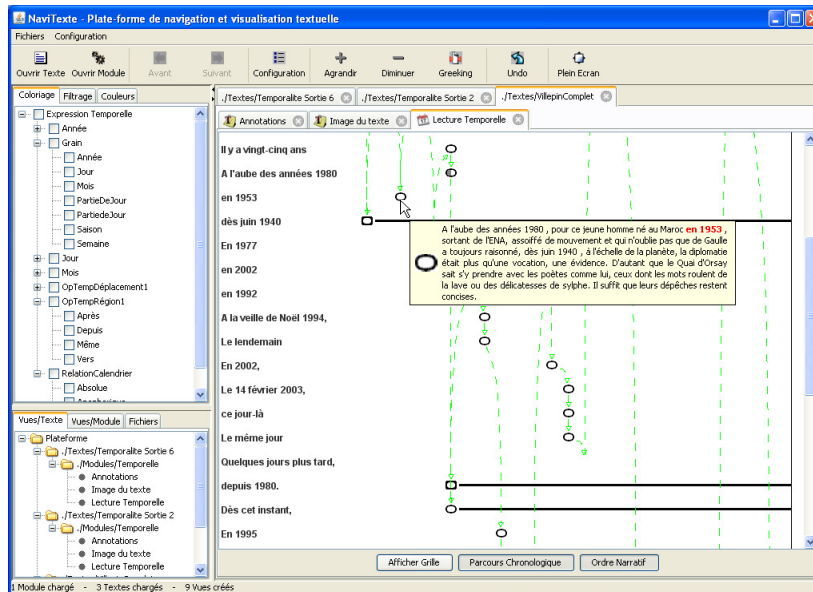


Fig. 3. Example of calendar view in NaviTexte

¹ A DTD is defined in [12].

4 Conclusion

In this work, we proposed algebra of operators to analyse calendar expressions and build a functional representation of these expressions. Then, we described an implementation of this approach in the plat-form NaviTexte and we shown how the functional representation is used to visualise a calendar view of a text. In future work, we will rely on a methodology presented in [15] in order to take into account several temporal axis, and thus of several calendar structures, which are expressed in texts by different levels of enunciations, like citations.

References

1. Charolles M.: L'encadrement du discours - Univers, champs, domaines et espace, Cahiers de recherche linguistique, LANDISCO, vol. 6 Université Nancy 2, (1997) 1-73.
2. Mani I., Wilson G.: Robust Temporal Processing of News, Proceedings 38ème ACL, (2000) 69-76
3. Setzer A., Gaizauskas R.: Annotating Events and Temporal Information in Newswire Texts, Proceedings 2nd LREC, (2000) 64-66
4. Filatova E., Hovy E.: Assigning Time-Stamps to Event-Clauses, Workshop on Temporal and Spatial Information Processing, ACL'2001, (2001) 88-95
5. Song F., Cohen R.: Tense Interpretation in the Context of Narrative, 9th AAI, (1991) 131-136
6. Hitzeman J., Moens M., Grover C.: Algorithms for Analyzing the Temporal Structure of Discourse, EACL'95, (1995) 253-260
7. TERQAS, Time and Event Recognition for Questions Answering Systems, an ARDA Workshop on Advanced Question Answering Technology, 2002, <http://www.timeml.org/terqas/>
8. Barzilay R., Elhadad N., McKeown K.: Sentence Ordering in Multidocument Summarization, HLT'01, (2001)
9. Pustejovsky J., Castano J., Ingria R., Sauri R., Gaizauskas R., Setzer A., Katz G.: TimeML : Robust Specification of Event and Temporal Expressions in Text, IWCS-5 Fifth International Workshop on Computational Semantics, (2003)
10. Ferro L., Gerber L., Mani I., Sundheim B., Wilson G.: TIDES Standard for the Annotation of Temporal Expressions, <http://www.mitre.org/work/tech-papers/tech-papers-04/ferro-tides/>, (2003)
11. Pazienza M.T.: Information Extraction, toward scalable, adaptable systems, Springer-Verlag, New York, (1999)
12. Couto J.: Modélisation des connaissances pour une navigation textuelle assistée. La plateforme logicielle NaviTexte. PhD, Université Paris-Sorbonne (2006)
13. Couto J., Minel J.-L.: NaviTexte, a Text Navigation Tool , Lecture Notes in Artificial Intelligence 4733, Springer-Verlag (2007) 251-259.
14. Schwer S. R. S-arrangements avec répétitions. Comptes Rendus de l'Académie des Sciences de Paris, Série I 334 (2002) 261-266.
15. Battistelli D., Chagnoux M. Représenter la dynamique énonciative et modale de textes, in actes TALN'07 (Traitement automatique du langage naturel (2007) 13-23.
16. Battistelli D., Minel J.-L., Schwer S. Représentation des expressions calendaires dans les textes: vers une application à la lecture assistée de biographies . *Revue TAL Vol 47, n°3.* (2007), p. 11-37.