



HAL
open science

Représentation algébrique des expressions calendaires et vue calendaire d'un texte

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

► **To cite this version:**

Delphine Battistelli, Javier Couto, Jean-Luc Minel, Sylviane R. Schwer. Représentation algébrique des expressions calendaires et vue calendaire d'un texte. TALN 08, Jun 2008, Avignon, France. pp.1-10. halshs-00275138

HAL Id: halshs-00275138

<https://shs.hal.science/halshs-00275138>

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.

Représentation algébrique des expressions calendaires et vue calendaire d'un texte

Battistelli D. (1), Couto J. (2), Minel J.-L. (3), Schwer S.R. (4)

(1) Lalic – Université Paris-Sorbonne – 28 rue Serpente 75006 Paris
delphine.battistelli@paris-sorbonne.fr

(2) INCO – Universidad de la Republica – 565 Herrera y Reissig 11300 Montevideo –
Uruguay
jcouto@fing.edu.uy

(3) MoDyCo – Université Paris X, CNRS UMR 7114 – 200 avenue de la République
92001 Nanterre Cedex
jminel@u-paris10.fr

(4) LIPN – Université Paris XIII, CNRS UMR 7030 – 99 Bd Jean-Baptiste Clément
93240 Villetaneuse
schwer@lipn.univ-paris13.fr

Résumé Cet article aborde l'étude des expressions temporelles qui font référence directement à des unités de temps relatives aux divisions courantes des calendriers, que nous qualifions d'expressions calendaires (EC). Nous proposons une modélisation de ces expressions en définissant une algèbre d'opérateurs qui sont liés aux classes de marqueurs linguistiques qui apparaissent dans les EC. A partir de notre modélisation, une vue calendaire est construite dans la plate-forme de visualisation et navigation textuelle NaviTexte, visant le support à la lecture de textes. Enfin, nous concluons sur les perspectives offertes par le développement d'une première application de navigation temporelle.

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" (EC). We propose to model these expressions by defining an operator algebra, the operators being related to different linguistics marker classes that occur in the EC. Based on our model, a calendar view is built up into the text visualization and navigation framework NaviTexte, aiming the support of text reading. Finally, we discuss the perspectives offered by the development of a first temporal navigation application.

Mots-clés : expressions temporelles calendaires, modélisation algébrique, visualisation

Keywords: temporal calendar expressions, algebraic modelization, vizualisation

1 Introduction

La prise en compte de la temporalité exprimée dans les textes apparaît comme fondamentale non seulement dans une perspective de traitement global de documents mais également dans l'analyse de la structure même d'un document, tout particulièrement quand ce dernier est long. Les expressions temporelles considérées ici sont des adverbiaux de localisation temporelle, dits aussi de datation (Guimier, 1996). Ce type d'expression a fait l'objet de nombreux travaux, tant sur le plan de leur analyse en tant que modes d'organisation des discours (Terran, 2002 ; Le Draoulec et Péry-Woodley, 2005) que sur le plan de leur repérage effectif dans les textes, depuis celles du type « demain », « dans trois jours », « vers la fin de la semaine » (comme dans Maurel, 1989) jusqu'à celles du type « lors d'une messe » (comme dans Vazov, 2001). Nous nous limitons ici à l'analyse et au repérage d'expressions temporelles qui font référence directement à des unités de temps relatives aux divisions courantes des calendriers ('an', 'semaine', etc.). Nous qualifions ces expressions d'*expressions calendaires* (désormais EC). Notons que dans notre conceptualisation (Battistelli et al., 2007) – même si nous n'en présentons pas d'analyse ici – une expression comme « au cours de TALN 2008 » constitue aussi une EC. En effet, même si 'TALN 2008' fait référence à un événement unique, il entre dans une série qui permet de générer une chronologie et donc de donner naissance à une unité calendaire. En revanche, une expression temporelle comme « pendant la seconde guerre mondiale », référant à un événement unique positionnable sur un calendrier, est exclue de notre définition d'une EC : l'événement 'guerre mondiale', s'il peut se reproduire (et être affecté d'un numéro), n'est en effet pas vécu comme une occurrence d'un événement type dont on pourrait prévoir sur une longue période de temps l'itération périodique (comme l'ont été par exemple les crues du Nil sur lesquelles s'est construit le calendrier agraire égyptien).

La question de l'annotation automatique des EC a été plus particulièrement posée dans le cadre de systèmes visant à ordonnancer des événements selon leur ancrage calendaire (Schilder et Habel, 2001) ou dans le cadre de systèmes de questions-réponses (Pustejovsky et al., 2005 ; Harabagiu et Bejan, 2005). Il est à noter à cet effet qu'une proposition de métalangage standard telle que TimeML, dédiée à l'annotation des événements dans des textes et de leurs relations temporelles (Pustejovsky et al., 2003), a été initialement mise en place dans le cadre du workshop TERQAS (2002), qui lui-même prenait place dans le contexte des systèmes de questions/réponses. Il intègre principalement deux schémas d'annotation, TIDES TIMEX2 (Ferro et al., 2003) et Sheffield STAG (Setzer et al., 2000), proposés à partir essentiellement de l'analyse d'adverbiaux temporels que sont les EC¹.

Dans cet article, nous nous proposons de décrire formellement ces EC – pour le français – en distinguant explicitement différentes classes de marqueurs linguistiques qui apparaissent dans les EC selon les types d'opérations qu'elles mettent en œuvre. Cette approche nous amène à proposer des critères fins pour d'une part l'annotation automatique de ces expressions et d'autre part la navigation temporelle dans un texte ou un corpus de textes. Nous nous distinguons des approches précédemment citées sur au moins deux points cruciaux : (i) notre but n'est pas de relier une EC ramenée à sa valeur – selon la norme ISO – à un événement dans un texte mais de relier qualitativement les EC d'un texte entre elles, c'est-à-dire établir leurs relations de positionnement relatif (l'ensemble de ces relations correspond à ce que nous appelons le calendrier propre au texte) ; (ii) nous modélisons la sémantique d'une EC comme

¹ Les EC sont généralement désignées sous le terme de « locutions adverbiales de temps » (voir par exemple pour le français (Maurel, 1989) qui qualifie ces expressions d'« adverbes de date »).

une expression algébrique². Dans la section 2, nous présentons notre approche formelle fondée sur une algèbre d'opérateurs et, dans la section 3, un projet d'application qui consiste à aider à la lecture d'un texte biographique. Nous concluons sur les perspectives offertes par le développement d'une première application de navigation temporelle.

2 Une algèbre des expressions calendaires

Nous postulons que l'EC utilisée pour référer un lieu calendaire peut être décrite par une suite d'opérations unaires (un seul argument) qui permettent d'accéder à ce lieu calendaire. Il s'agit donc de considérer une EC comme la composition d'un certain nombre d'opérateurs qui chacun donne un bout de chemin selon l'ordre suivant : à partir d'une *base calendaire* (BC) prise comme argument, qui peut être donnée immédiatement (21 juin 2008 dans « *le 21 juin 2008* ») ou médiatement (le jour courant dans « *demain* ») généralement introduite par un déterminant (le, ...) et/ou une préposition (en, à, ...) qui opère comme un pointeur, un second type d'opérateur exprime la portion « utile » de cette base (tout, le début, le milieu, la fin, une zone floue autour). Cette portion utile partitionne la ligne du temps en trois régions (cf. Fig. 1) : la demi-droite antérieure (A), la portion (U) elle-même et la demi-droite postérieure (P) – cette demi-droite étant bornée par l'instant présent (ou le maintenant du texte). Un troisième type d'opérateur permet d'accéder au lieu décrit par l'EC : soit par déplacement quantitatif (ex. : « *trois semaines plus tard* »), soit par zoom (ex. : « *l'automne de cette année-là* ») soit en prenant l'une des trois régions découpées (ex. : « *jusqu'en mai 2006* »).

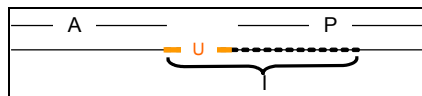


Figure 1 : Partition de la ligne du temps pour une EC unaire

Notre objectif est la définition d'une expression algébrique pour chaque EC. Une expression algébrique est dérivée de la décomposition d'une EC en sous-expressions chacune associée à un opérateur ou, pour la plus à droite, à un opérande, appelé base calendaire (BC). Analysons par exemple l'EC « *depuis la fin des années 1980* ». Notre démarche, constructive, commence par considérer « *années 1980* » comme point de départ de la décomposition. Cette expression correspond à BC, qui porte le grain et la valeur permettant, puisqu'elle entretient des relations avec d'autres BC, de référencer celle-ci dans le calendrier. À cette BC s'applique un opérateur de *pointage* correspondant à l'article « *les* », application qui renvoie l'expression « *les années 1980* »³ qui détermine la partie I de Fig. 1 et qui est une EC. Ensuite, la portion utile U de BC est obtenue par l'opérateur exprimé par « *la fin de* » ; enfin, la région dépeinte par l'EC est la partie P union U de l'opérateur « *depuis* ». Cette analyse est strictement qualitative comme l'exprime le schéma de Fig. 2. Un opérateur dit de *déplacement* correspond à la composante lexicale « *la fin de* » et un opérateur dit de *régionalisation* correspond à la préposition « *depuis* ». Ainsi, l'EC « *depuis la fin des années 1980* » est représentée par l'expression fonctionnelle suivante : *depuis*(*la fin de*(*les*(*années*

² Dans cet article, nous nous limitons à la description d'EC ne faisant pas intervenir de quantification comme dans « *chaque mardi* » et ne correspondant pas à une séquence comme dans « *lundi et mardi* ». Nous considérons par ailleurs ici des EC qui ne nécessitent pas de faire appel pour leur calcul à des marques temporelles autres qu'adverbiales (comme les temps verbaux ou les adjectifs).

³ Cette expression n'est pas lexicalisée dans l'EC à cause du phénomène de contraction.

1980))). Tab. 1 montre les expressions obtenues en appliquant la même démarche à d'autres EC.

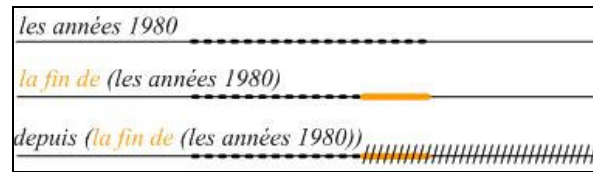


Figure 2 : Analyse qualitative d'une EC

EC	Modélisation fonctionnelle
<i>mai 1974</i>	(mai 1974)
<i>en 1945</i>	en ((1945))
<i>pendant l'été 2001</i>	pendant(l'(été 2001))
<i>au mois d'août</i>	à (le (mois d'août))
<i>à l'aube de l'année 2001</i>	à l'aube de (l' (année 2001))
<i>à partir de 1958</i>	à partir de ((1958))
<i>à la veille de Noël 1994</i>	à (la veille de ((Noël 1994)))
<i>trois jours avant Noël 1994</i>	(trois jours avant ((Noël 1994)))
<i>jusqu'en juillet 1962</i>	jusqu'en((juillet 1962))
<i>le 21 juin</i>	le (21 juin)
<i>au milieu de la fin des années 1990</i>	à (le milieu de (la fin de (les (années 1990))))

Tableau 1. Exemples de modélisation fonctionnelle d'EC

Les exemples de Tab. 1 montrent que certains opérateurs sont implicites : les EC « *trois jours avant Noël 1994* » et « *mai 1974* » ne contiennent pas d'expressions explicites associées au pointage. Les expressions fonctionnelles associées font apparaître cet opérateur par un parenthésage systématique de leur BC, qui les transforme en des EC. L'EC « *depuis la fin des années 1980* » correspond à une transformation successive des trois EC suivantes : EC1= « les années 1980 », EC2 =« la fin des années 1980 » et EC3= « depuis la fin des années 1980 ». Une parenthèse ouvrante non précédée d'une expression d'opérateur correspond à l'expression d'un *opérateur implicite*. Pour ne pas surcharger l'écriture, on ne représente que l'opérateur implicite de pointage, car c'est le seul opérateur non interne aux EC.

Il ressort des exemples étudiés (tirés de notre corpus de biographies et de textes accessibles sur le web) que les opérateurs relèvent de trois types différents : (i) les *opérateurs de pointage*, qui associent à la BC une zone calendaire (I) de la Fig. 1 ; (ii) des *opérateurs* dits de *focalisation/déplacement*, qui sont de deux sortes : méréologiques pour la focalisation (tout, début, milieu, fin) qui définissent la zone calendaire utile (U) de la Fig. 1 et qui peuvent s'itérer (ex. : « au milieu de la fin des années 1990 ») et (in/dé)crémentaux pour le déplacement (ex. : « trois jours avant Noël »), c'est-à-dire qui déportent la zone utile à gauche ou à droite de U ; (iii) des *opérateurs* dits de *régionalisation* qui définissent la partie de l'espace temporel issue de U captée par l'EC : on obtient soit U, soit A, soit P, soit U et A, soit U et P, soit A et P. Notons ensuite que les opérateurs de focalisation peuvent s'appliquer plusieurs fois, comme dans l'EC « *au milieu de la fin des années 1990* », de même que les opérateurs de régionalisation comme dans l'EC « *jusqu'à au-delà de 1940* ». Enfin, la vaste combinatoire potentielle des opérateurs est, dans la pratique, plus restreinte, au vu de différents exemples que nous avons étudiés dans notre corpus. Il existe un ordre entre les trois types d'opérateurs ; des expressions telles que « *au début de depuis les années 1980* » ou « *depuis jusqu'en 2005* » vont ainsi être considérées comme agrammaticales. Finalement, nous considérons donc qu'une EC est définie par une BC sur laquelle s'applique d'abord un seul opérateur de pointage, des opérateurs de déplacement s'appliquent ensuite, et enfin des opérateurs de régionalisation, ce que nous illustrons par l'expression suivante :

OpRégionalisation⁺ (*Opfocalisation/déplacement*⁺ (*OpPointage* (BC)))

Notons que l'algèbre est bien définie. L'opérateur de pointage, le plus interne, transforme une BC en une EC et tous les autres opérateurs prennent en argument une EC et fournissent en résultat une EC.

2.1 Analyse des opérateurs

Un opérateur désigne un ensemble d'instances lexicales possibles liées par une relation d'équivalence. Le critère de regroupement est lié à l'effet qu'un opérateur produit sur une EC. Ainsi, nous considérons que les expressions « à l'aube des années 1980 » et « au début des années 1980 » désignent des zones utiles identiques (les expressions « à l'aube de » et « au début de » sont donc considérées comme deux instances d'un même opérateur). De même, « pendant » et « en » désignent la même région calendaire, même si elles ne désignent pas la même relation entre la zone calendaire et le procès concerné. Rappelons que notre but est de situer relativement les unes aux autres les EC, et non les procès qu'elles contribuent à situer temporellement.

2.1.1 L'opérateur de pointage

Cet opérateur transforme une BC en une EC, que nous appelons EC élémentaire.

2.1.2 Les opérateurs de focalisation/déplacement

Le premier opérateur de Tab. 2 est celui qui ne modifie pas l'EC élémentaire, c'est donc l'opérateur d'identité. Les trois opérateurs suivants focalisent sur une partie de l'EC élémentaire. Nous avons adopté une partition en trois parties : début, milieu et fin, comme c'est généralement le cas (voir par exemple (Maurel, 1989) qui les encode comme trois parties successives et de même taille). Notre approche uniquement qualitative de représentation laisse à l'utilisateur le soin de définir les proportions respectives de chaque partie. Les deux derniers opérateurs correspondent à des déplacements avant ou arrière sur l'axe temporel relativement à l'EC élémentaire. Ces opérateurs de déplacement sont nécessairement paramétrés par une unité et un ordinal. Par exemple, l'EC « trois jours avant Noël » correspondra à l'expression *DéplacementAvant*(unité, 3)(*OpPointage*(Noël)).

Nom de l'opérateur	Exemples d'EC
<i>IdFocalDeplac</i> (<i>IdFD</i>)	en 1945 ; pendant l'été de 2001 ; au mois d'août
<i>FocalDébut</i> (<i>Fd</i>)	à l'aube des années 1980 ; au début de mai 2005
<i>FocalMilieu</i> (<i>Fm</i>)	au milieu de l'année 2006
<i>FocalFin</i> (<i>Ff</i>)	à la fin d'août 1979
<i>DéplacementAvant</i> (unité, n) (<i>Dep</i> (u, -n))	trois jours avant Noël
<i>DéplacementAprès</i> (unité, n) (<i>Dep</i> (u, +n))	cinq semaines après Pâques.

Tableau 2. Opérateurs de focalisation/déplacement

2.1.3 Les opérateurs de régionalisation

Un opérateur de régionalisation opère sur *Opfocalisation*⁺/*déplacement*⁺ (*OpPointage*(BC)) correspondant à la zone calendaire utile U qui devient le point d'ancrage servant à définir la région calendaire décrite par l'EC. A partir de cette zone six régions sont définissables, donnant lieu à six opérateurs (cf. Tab. 3). L'absence de marqueur linguistique de régionalisation spécifie l'opérateur *RegionId*. Un opérateur de régionalisation floue a été prévu pour décrire des expressions comme « vers le 21 avril 1997 » ou « peu avant le 21 avril 1997 » ; il est visualisé à l'aide d'une ellipse en pointillés. Signalons que, à l'instar des opérateurs de pointage et de Focalisation/Déplacement, les opérateurs de régionalisation sont

unaires, à l'exception de l'opérateur de régionalisation binaire associé à des EC comme « entre mai et août 2006 », « depuis l'an 2006 et jusqu'à l'an 2007 » ou « de l'été 2007 à mai 2008 » ou encore des EC dites booléennes comme « le 21 juin ou/et le 15 août ». Les deux EC arguments définissent alors une partition de la droite temporelle en cinq zones],U1[, U1,]U1,U2[, U2,]U2,-[qui peut générer *a priori* autant de régionalisations que la combinatoire le permet (cf. Fig. 3). Par exemple, *entre* est associé à la région comprise entre U1 et U2, bornes comprises ou non, selon les opérateurs de régionalisations unaires appliqués aux EC arguments.

Nom de l'opérateur	Exemples d'EC
RégionId (RId)	le 21 avril 1997
RégionAvant (Rav)	avant le 21 avril 1997
RégionJusque (Rj)	jusqu'au 21 avril 1997
RégionDepuis (Rd)	depuis le 21 avril 1997
RégionAprès (Rap)	après le 21 avril 1997
RégionHors (Rh)	sauf le 21 avril 1997

Tableau 3. Classes d'opérateurs de régionalisation

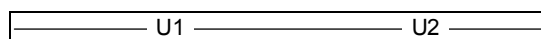


Figure 3 : Partition de la ligne du temps pour une EC binaire

2.2 Exemples d'analyse algébrique d'expressions calendaires

Nous reprenons ci-après les expressions de Tab. 1 en en donnant une analyse algébrique.

Exemples d'EC	Modélisation algébrique
mai 1974	RId(IdFD(mai 1974))
en 1945	RId(IdFD(1945))
pendant l'été 2001	RId(IdFD(été 2001))
au mois d'août	RId(IdFD(mois d'août))
à l'aube de l'année 2001	RId(Fd(année 2001))
à partir de 1958	Rd(IdFD(1958))
à la veille de Noël 1994	Rid(Dep(jour,-1)(Noël 1994))
trois jours avant Noël 1994	Rid(Dep(jour,-3)(Noël 1994))
jusqu'en juillet 1962	Rj(IdFD(juillet 1962))
le 21 juin	RId (IdFD(21 juin))
au milieu de la fin des années 1990	RId(Fm(Ff(années 1990)))

Tableau 4. Exemples de modélisation algébrique d'EC

3 Construction de la vue calendaire d'un texte

L'implémentation informatique de notre approche se fonde sur la plate-forme logicielle de navigation textuelle NaviTexte (Couto, 2006 ; Couto et Minel, 2007). Nous avons tout d'abord défini une manière d'encoder l'algèbre dans les textes, puis un nouveau type de vue, nommé *vue calendaire*, a été conçu et développé. Nous présentons ces deux aspects ici.

3.1 Repérage et encodage des expressions calendaires

Afin que les EC puissent être informatiquement exploitées, il est nécessaire de repérer et de représenter préalablement celles-ci sous un format traitable. La représentation informatique utilisée est celle de NaviTexte, dont nous rappelons brièvement les principes qui s'inspirent du standard TEI. La représentation du texte, décrite dans un format XML, se divise en deux

parties : le *Corps*, où les unités textuelles significatives pour la tâche sont délimitées, et la *Tête*, où s'expriment les relations non hiérarchiques entre ces mêmes unités. Dans le *Corps*, l'élément de base du modèle est l'*Unité Textuelle* (UT) typée, ce qui permet d'incorporer de nouveaux éléments textuels de manière simple. Dans le cadre de notre travail, les EC sont annotées dans le *Corps* comme des UT de type « expression calendaire ». Par ailleurs, un objet *Séquence* qui représente l'ordonnancement des UT des EC est déclaré dans la *Tête*.

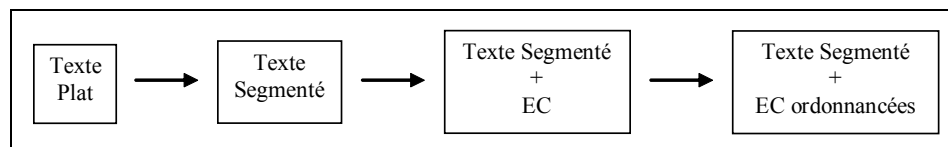


Figure 4 : Etapes de traitement

Le traitement se décompose en trois étapes (cf. Fig. 4) qui, à partir d'un texte non formaté, construisent un texte traitable par NaviTexte où les EC ont été repérées, encodées et ordonnancées. Le premier module segmente le texte en sections, paragraphes et phrases, repérant également les titres. Le deuxième, fondé sur des modèles d'automates à états finis, est chargé de repérer les différentes instances des EC et de créer les UT correspondantes avec toutes les annotations nécessaires. Ces annotations encodent notre algèbre. À titre d'exemple, nous donnons dans Fig. 5 les annotations de l'EC « *avant le début de la fin de l'année 2008* ». Les annotations sont définies en relation avec l'ensemble des informations mettant en particulier en évidence les types de cheminement opérés dans le calendrier selon la modélisation proposée (cf. annotations *OpTempRégion_n*, *OpTempDéplacement_n*). Enfin, un troisième module calcule l'ordonnancement des EC.

```

<UT Type="Expression Calendaire" Nro="7">
  <Annotation Nom="Grain">Année</Annotation>
  <Annotation Nom="Année">2008</Annotation>
  <Annotation Nom="RelationCalendrier">Absolue</Annotation>
  <Annotation Nom="OpTempRégion1">Avant</Annotation>
  <Annotation Nom="OpTempDéplacement1">FocalFin</Annotation>
  <Annotation Nom="OpTempDéplacement2">FocalDébut</Annotation>
  <Chaine>avant le début de la fin de l'année 2008</Chaine>
</UT>
  
```

Figure 5 : Annotation de l'EC « avant le début de la fin de l'année 2008 »

3.2 Vue calendaire d'un texte

On appelle *calendrier (propre) du texte* l'ensemble des éléments calendaires qui interviennent dans le texte. Au contraire du travail présenté dans (Battistelli et al., 2007), où nous nous étions restreints aux EC cadratives, nous représentons actuellement toutes les EC. Un nouveau type de vue a été créé dans la plate-forme NaviTexte. Cette vue calendaire est construite à partir d'un texte comportant des EC annotées comme décrit à la section 3.1. Un exemple est donné dans Fig. 6 ; il correspond à l'analyse d'un texte biographique tiré du journal *Le Monde* du 6 octobre 2005 et intitulé « Villepin Pile et face ». La vue est conçue comme un graphe coordonné, visuellement, à une grille bidimensionnelle. Dans la partie gauche de la vue nous trouvons les chaînes lexicales des différentes instances d'EC du texte. Par défaut, celles-ci sont ordonnées selon l'ordre d'apparition dans le texte, mais il est possible de passer à un ordonnancement chronologique, en utilisant les options offertes dans le panneau situé en bas de la vue. Ces chaînes lexicales sont représentées par des sommets dans le graphe. La représentation graphique de toute EC est placée sur l'axe horizontal, en fonction de sa position dans l'ordonnancement temporel qui a été préalablement calculé et

encodé comme une séquence dans le texte (cf. section 3.1). Notons que la relation entre l'axe horizontal et la séquence dans le texte n'est pas biunivoque, différentes EC pouvant référer à la même position sur l'axe horizontal. Pour chaque EC, d'autres sommets sont créés dans le graphe, placés sur la même ligne horizontale. Le premier correspond à la représentation graphique de celle-ci. Ensuite, deux autres sommets sont créés, correspondant, l'un à l'*origo* du référentiel propre au texte (dans l'exemple de Fig. 6, le 6 octobre 2005), l'autre au futur. Ces deux sommets peuvent coïncider si toutes les EC sont antérieures à l'*origo*. Une fois les sommets créés, des arcs peuvent l'être. Lorsqu'il s'agit d'une EC déictique (ex. « *le 21 juin dernier* »), un arc dirigé entre sa représentation graphique et le sommet représentant l'*origo* est créé. Lorsque l'EC contient un opérateur de régionalisation (ex. : « *depuis 1980* »), un arc non dirigé est créé, dépendant de l'opérateur.

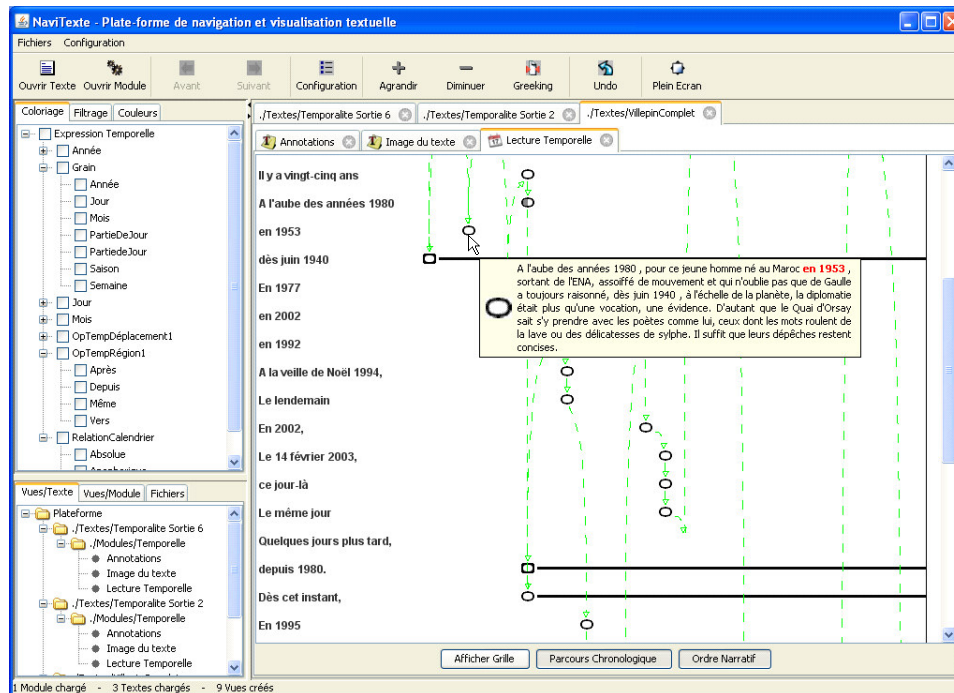


Figure 6 : Exemple de vue calendaire dans NaviTexte

L'avantage d'utiliser une structure de graphe est double. D'une part, en tant que structure de données, celle-ci modélise de manière naturelle les divers éléments participant à la vue calendaire d'un texte. D'autre part, en terme de visualisation, la gestion est simplifiée : lorsque l'utilisateur demande des changements (zooms de la vue, changements d'ordre des EC), il suffit de recalculer certaines positions des sommets, tandis que les arcs sont automatiquement modifiés. Nous avons utilisé le package Java nommé JGraphe pour réaliser l'implémentation informatique. Cela nous a permis en outre de combiner des aspects de modélisation de données (comme les arcs correspondant aux opérateurs de régionalisation) avec des aspects purement de visualisation (par exemple, dans l'interface de navigation développée, les arcs en vert qui montrent le parcours chronologique, ou les arcs servant comme grille horizontale, qui peuvent être activés et désactivés en utilisant les options affichées dans le panneau situé en bas de la carte). L'utilisation d'infobulles sert à contextualiser les EC par rapport au texte original. Dans Fig. 6, l'EC « *en 1953* » apparaît en rouge dans l'infobulle (partie de texte encadrée) qui montre la chaîne lexicale du paragraphe la contenant. Nous voudrions enfin attirer l'attention sur la manière dont la représentation graphique d'une EC est calculée par NaviTexte : elle procède d'une métamorphose relativement à l'application progressive des différents opérateurs (cf. Tab. 5). Prenons comme

autre exemple l'EC « *après le milieu de la fin des années 1990* », dont l'expression fonctionnelle est *après(le milieu de (la fin de(les (années 1990)))*). L'application de l'opérateur de pointage « *les* » crée une représentation initiale qui est, toujours, une ellipse vide. Ensuite, l'opérateur « *la fin de* » sélectionne une partie de cette ellipse vide, partie sur laquelle se réalise une autre sélection lorsqu'on applique l'opérateur « *le milieu de* ». Une fois les opérateurs de déplacement appliqués, un opérateur de régionalisation peut encore modifier l'image (s'il s'agit d'un opérateur inclusif comme *depuis* ou *jusque*) et susciter la création d'un arc montrant la région référée⁴.





<i>l'(année 2008)</i>	
<i>la fin de (l'(année 2008))</i>	
<i>le début de (la fin de (l'(année 2008)))</i>	
<i>avant (le début de (la fin de (l'(année 2008))))</i>	

Tableau 5. Métamorphose progressive dans la représentation graphique d'une EC

4 Conclusion

Nous avons présenté ici une algèbre d'opérateurs rendant compte de la sémantique d'expressions calendaires et à même de spécifier des opérations de représentation graphique de ces expressions. Comme exposé dans (Le Draoulec et Péry-Woodley, 2005), s'il est indéniable que les adverbiaux temporels jouent un rôle structurant, ils ne constituent pas nécessairement un outil d'indexation fiable de propositions ou de blocs de propositions. Aussi, nous les considérons seulement comme des « indices » de structuration temporelle ou encore comme autant de marques disposées dans le texte non explicitement reliées à des segments textuels donnés. Ce qui nous intéresse ici n'est pas d'analyser la portée sémantique de ces marques et de chercher alors à structurer le texte à partir de ces marques. Il s'agit plutôt de s'abstraire du texte pour en extraire son calendrier propre à partir duquel l'on cherche à modéliser le balayage effectué par le lecteur selon une dimension temporelle. Dans cette démarche, l'accent est donc mis sur la mise en perspective de la façon dont un lecteur appréhende le parcours opéré dans le calendrier au travers de ces marques. Nous travaillons par ailleurs à l'identification de sous-systèmes calendaires à partir de l'analyse des différentes prises en charge énonciative et modale dans les textes selon une méthodologie présentée dans (Battistelli et Chagnoux, 2007). Cette dernière rend compte des articulations complexes possibles entre plusieurs niveaux de discours à l'intérieur d'un même texte ; ils correspondent à une plus ou moins grande distance marquée par l'énonciateur principal dans les situations qu'il décrit (avec l'utilisation de marques de discours directs et indirects en particulier). Les segments textuels ainsi identifiés sont organisés de manière hiérarchique. Il s'agit alors de considérer que chaque ensemble d'EC possiblement présentes dans un segment forme un sous-système calendaire propre et que l'ensemble des EC d'un texte peut donc renvoyer à différents sous-systèmes calendaires propres à l'intérieur desquels – ou entre lesquels – un parcours de navigation peut être envisagé.

Remerciements

Notre travail s'inscrit dans le cadre des financements de l'ANR (programme Blanc Conique). /NaviTexte/ est soutenu actuellement par le programme ECOS-Sud U05H01.

⁴ Nous avons trouvé une occurrence d'un tel type d'expressions, à savoir « depuis le milieu de la fin 1990 », dans le compte-rendu d'audience du Tribunal Pénal International pour l'ex-Yougoslavie, consulté le 15/11/2007 à l'adresse <http://www.un.org/ictytransf54/060203IT.htm> (p. 47899).

Références

- BATTISTELLI D., CHAGNOUX M. (2007). « Représenter la dynamique énonciative et modale de textes ». *Actes TALN'07*, 5-8 juin 2007, Toulouse, p. 23-32.
- BATTISTELLI D., MINEL J-L., SCHWER S. (2007). « Représentation des expressions calendaires dans les textes : vers une application à la lecture assistée de biographies ». *TAL Vol 47/3*, p. 11-37.
- COUTO J. (2006). Modélisation des connaissances pour une navigation textuelle assistée. La plate-forme logicielle NaviTexte. Thèse de doctorat, Université Paris-Sorbonne (Paris IV).
- COUTO J., MINEL J.-L. (2007). « NaviTexte, a Text Navigation Tool ». 10th Congress of Italian Association for Artificial Intelligence. *Lecture Notes in Computer Science 4733*, p. 251–259.
- FERRO L., GERBER L., MANI I., SUNDHEIM B., WILSON G. (2003). « TIDES Standard for the Annotation of Temporal Expressions », <http://www.mitre.org/work/tech-papers/tech-papers/04/ferro-tides/>.
- GUIMIER C. (1996). *Les adverbes du français. Le cas des adverbes en -ment*. Paris. Ophrys.
- HARABAGIU S., BEJAN A. (2005). « Question Answering Based on Temporal Inference ». *Actes AAAI-2005 Workshop on Inference for Textual Question Answering*.
- LE DRAOULEC A., PÉRY-WOODLEY M.-P. (2005). « Encadrement temporel et relations de discours ». *Langue Française* 148, p. 45-60.
- MAUREL D. (1989). Reconnaissance de séquences de mots par automates, adverbes de date du Français. Thèse de doctorat, Université Paris 7.
- PUSTEJOVSKY J., SAURI R., CASTAÑO J., RADEV D.R., GAIZAUSKAS R., SETZER A., SUNDHEIM B., KATZ G. (2004). « Representing Temporal and Event Knowledge for QA Systems ». Mark T. Maybury (ed.), *New Directions in Question Answering*. MIT Press, Cambridge.
- PUSTEJOVSKY J., CASTANO J., INGRIA R. , SAURI R., GAIZAUSKAS R., SETZER A., KATZ G. (2003). « TimeML : Robust Specification of Event and Temporal Expressions in Text ». *Actes IWCS-5 Fifth International Workshop on Computational Semantics*.
- SCHILDER F., HABEL CH. (2001). « From Temporal Expressions to Temporal Information : Semantic Tagging of News Messages. ». *Actes ACL'01, Workshop on temporal and spatial information processing*, p. 65 -72.
- SETZER A., GAIZAUSKAS R. (2000). « Annotating Events and Temporal Information in Newswire Texts ». *Actes 2^{ème} LREC*, p. 64-66.
- TERQAS (2002). Time and Event Recognition for Questions Answering Systems, an ARDA Workshop on Advanced Question Answering Technology. <http://www.timeml.org/terqas/>.
- TERRAN E. (2002). Le cadrage temporel en français. Thèse de doctorat, Université Paris 3.
- VAZOV N. (2001). « A System for Extraction of Temporal Expressions from French Texts ». *Actes TALN'2001*, p. 315-324.