

Ordonnement des expressions calendaires dans les textes

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

(1) LIPN – Université Paris XIII, CNRS UMR 7030, 99 Bd Jean-Baptiste
Clément 93240 Villetaneuse

schwer@lipn.univ-paris13.fr Adresse Electronique1

(2) Lalic – Université Paris-Sorbonne, 28 rue Serpente 75006 Paris
[lectronique2](#)

(3) MoDyCo – Université Paris X, CNRS, 200 avenue de la République
92001 Nanterre Cedex

cvaguer@free.fr

jminel@u-paris10.fr

Commenté [V1]: ajouter

Mis en forme : Police par défaut

Mis en forme : Police par défaut

Résumé Ici, un résumé en français (max. 150 mots).

Abstract Here an abstract in English (max. 150 words).

Mots-clés : une liste de mots-clés en français

Keywords: Là, une liste de mots-clefs en anglais

1 Introduction

Les expressions temporelles considérées ici sont des adverbiaux de localisation temporelle, dits aussi de datation (Guimier, 1996). Ce type d'expressions 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 les plus simples (*demain, dans trois jours, vers la fin de la semaine*, comme dans Maurel, 1993) jusqu'aux plus complexes (*deux jours après l'arrivée triomphante du Tour de France*, comme dans Vazov, 2001). Nous nous limitons ici à l'analyse et au repérage d'expressions temporelles qui font référence explicitement à des unités de temps¹, qu'elles soient relatives aux divisions courantes des calendriers ('an', 'semaine', etc. comme dans *l'an prochain, pendant la semaine*, etc.) ou propres à une communauté ('unité de temps relative à la durée séparant deux conférences TALN' comme dans *au cours du prochain TALN*). Nous qualifions ces expressions (Battistelli et al., 2006) d'expressions calendaires (désormais EC). Dans cette conceptualisation, une expression temporelle référant à un événement unique (*pendant l'Occupation, pendant la seconde guerre mondiale*), même si elle est positionnable sur un calendrier n'est pas considérée comme calendaire et est exclue de notre champ d'analyse. Pour une expression comme *pendant la seconde guerre mondiale*, 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.

Commenté [V2]: Changer l'exemple

Dans le présent article, nous nous intéressons aux mécanismes d'ordonnement de ces EC. Cela suppose d'aborder la question de leur représentation tout comme celle de leur annotation (sémantique). Cette question a été plus particulièrement posée dans le cadre de systèmes visant à ordonner des événements en référence à leur ancrage calendaire (Schilder et Habel, 2001) ou dans le cadre de systèmes de questions-réponses (Pustejovsky et al., 2005, Harabagiu et al., 2005). Dans ce dernier cas, la sémantique de ces expressions est envisagée en fonction du type de questions auxquelles elles permettent de répondre : à la question « quand ? » correspondent d'une part des EC² dites de date, à la question « combien de temps ? » (respectivement « de fois ») correspondent des EC dites de durée (respectivement de fréquence). Nous nous proposons ici de revenir sur la sémantique aspectuelle de ces expressions (partie 2) sur laquelle nous nous appuyons ensuite pour spécifier les modes de calcul des valeurs calendaires nécessaires pour ordonner ces expressions (partie 3). Nous terminons par la présentation de l'algorithme qui met en œuvre ces calculs et son intégration dans une plate-forme de navigation textuelle (partie 4).

Commenté [CV3]: Référence à ajouter en biblio

¹ Nous excluons donc les expressions localisatrices dans le temps qui correspondent à des subordonnées temporelles telle que *Quand X est arrivé, ...*.

² En réalité, il est abusif de parler ici d'EC puisque les expressions temporelles considérées dans ces approches ne relèvent pas de la définition que nous en donnons, au sens où il n'est pas proposé d'intégrer la notion de « grain propre à un corpus ».

2 Annotation référentielle et catégorisation aspectuelle des EC

Le repérage et l'annotation référentielle des EC ont été réalisés à l'aide d'Unitex (fondé sur l'utilisation de grammaires locales – Maurel, 1988 et 1993, Constant, 2004, Schilder et Habel, 2005) reconnaissant les unités calendaires et les modificateurs aspectuels qui leurs sont associés (prépositions, déterminants, ...). L'annotation de ces EC se déploie donc selon deux paradigmes : un « paradigme référentiel » qui permet de considérer les EC comme étant en relation médiata ou immédiate avec un calendrier ; un « paradigme aspectuel » qui permet de définir qualitativement les types de zones temporelles du calendrier auxquelles les EC sont associées.

Seul le paradigme référentiel est convoqué dans le cadre du repérage des EC et renvoie à la distinction entre EC absolues et EC relatives (anaphoriques ou déictiques). Un premier travail (Christova et De Vismes, 2006), réalisé à partir d'un corpus de biographies³, a permis de dégager des patrons d'extraction d'EC afin de répondre à l'objectif visé alors : la génération d'un système calendaire propre aux textes étudiés en vue d'une application telle que la navigation textuelle à l'intérieur de biographies. Une seconde étude (Vaguer, 2006) a permis d'affiner les caractéristiques syntaxiques et sémantiques des EC : le corpus retenu a été le même que celui précédemment cité, soit un corpus de biographies d'environ 16 500 mots, dans lequel ont été identifiées plus de 450 expressions temporelles dont 388 catégorisées comme EC absolues, relatives anaphoriques textuelles et relative à valeur déictique (cf. Tableau 1). Ces dernières nécessitent pour être interprétées la connaissance d'une date (celle de rédaction de l'article, de l'instant d'énonciation). Quant aux EC relatives à valeur anaphorique, elles ont nécessairement un antécédent dans le texte même.

EC absolues	78%	<i>le 14 décembre 1936, entre 1986 et 1988, à dater de 1958, du printemps 1991 à décembre 1994, de 1963 à novembre 1970, été 2003, dans les années 1950, à l'aube des années 1980, à la veille de Noël 1994, un soir de septembre 1995...</i>
EC relatives anaphoriques textuelles	14%	<i>la même année, deux ans plus tard, pendant trois ans, à l'automne, dès le début de l'année, après plus de cinq ans, cela fait trente ans, l'année suivante, le lendemain, ce jour-là...</i>
EC relatives à valeur déictique	8%	<i>après le 29 mai, dimanche 15 janvier, jeudi dernier, le 15 septembre dernier, en avril, en cette deuxième semaine de janvier...</i>

Tableau 1 : Identification et répartition des EC dans le corpus

Les transducteurs ainsi développés ont permis le repérage d'EC qui ne sont pas prises en charge par les grammaires locales développées dans (Constant, 2006) comme, par exemple, des dates entre parenthèses [(1952), (1964-1970), (été 2003)], des syntagmes prépositionnels [à la veille de Noël 1994, après le 29 mai, avant le 21 avril, jusqu'en 1978], des syntagmes nominaux [ce jour-là, cette année(-là), la même année, le lendemain, cinq ans plus tard, juin 1940, un jour de septembre 1997], des EC mixtes [du printemps 1991 à décembre 1994].

³ Différents corpus ont été considérés : le Corpus 1 contenait de très courtes biographies (3 à 14 lignes) dans lesquelles la vie de la personne concernée était décrite de manière totalement chronologique grâce à des unités explicites comme « en 1949 », « le 3 janvier », etc. Le Corpus 2 était constitué de trois articles de même type mais plus longs (30 à 50 lignes), contenant quelques dates anaphoriques. Le Corpus 3 était restreint à la seule biographie de De Villepin, d'environ 320 lignes (8 pages parues dans *Le Monde* du 4/10/2005), rédigée dans un ordre peu chronologique, avec de « nombreux retours en arrière et un emploi fréquent d'EC relatives » (Christova et De Vismes, *op.cit.* : 25). D'autres biographies sont venues compléter cet ensemble de textes : celle de F. Mitterrand (*Le Monde*, 1/01/2006), celle de M. Bachelet (*Le Monde*, 14/01/2006), celle de J. Rau (*Le Monde*, 28/01/2006) et celle de J. Chirac (*Le Monde*, 8/05/2005).

Nous avons identifié les attributs suivants pour caractériser les EC du point de vue référentiel (cf. Tableau 2). Dans les transducteurs, cette annotation a été effectuée à l'aide de variables (qui viennent délimiter ce qui doit être annoté) et permet de reconnaître aussi bien des EC simples (*le 10 avril 1943*) que complexes (*i.e.* des EC mixtes qui contiennent des grains différents telle que *du printemps 1991 à décembre 1994*).

Attribut	Description	Valeurs possibles
Grain	Unité temporelle qui relève de la division du calendrier en différentes unités hiérarchisées	Année, Mois, Semaine, Jour, Saison...
Relation_Calendrier	L'EC est-elle en relation médiante ou immédiate avec un calendrier ?	Absolu ou Relatif
Lieu_Enonciation	De quel type est l'EC ? Quelles sont les règles linguistiques convoquées ?	Déictique ou Anaphorique
Referent_Textuel	L'antécédent est-il présent dans le texte ?	Oui ou Non

Tableau 2 : Liste des attributs retenus pour annoter les EC du point de vue référentiel

L'interprétation référentielle des EC ne suffit pas pour représenter d'une part leur sémantique, d'autre part le parcours opéré dans le calendrier (il faut en effet pour cela être en mesure de d'associer aux EC une zone temporelle dans le calendrier). Ce n'est donc pas la valeur de l'unité calendaire (*i.e.* le grain) qui va nous intéresser (tout grain, quel qu'il soit, conduit nécessairement à considérer une zone temporelle) mais celles de marqueurs tels que les déterminants, les locutions prépositionnelles, les connecteurs, les quantificateurs, *etc.* qui permettent de spécifier la zone temporelle désignée : *depuis jeudi matin* indique la zone inchoative tandis que *jusqu'à jeudi matin* la zone terminative de la zone temporelle. Nous parlons à ce propos de « paradigme aspectuel » qui initie une distinction entre trois types de marqueurs selon qu'une borne initiale ou finale est explicitement désignée (après le 20 juin, avant le 20 juin)...

Commenté [V4]: Paragraphe à rédiger par Delphine pour remplacer le Tableau 3.

L'interprétation des modifieurs d'EC est un élément important dans l'ordonnement des EC entre elles.

(cf. Tableau 3)⁴.

marqueurs d'intervalle	Inchoatif / terminatif	à dater de, à partir de, après, depuis, dès, cela fait / jusqu'en
	global	de...à..., entre...et...
	séquent	pendant, dans, durant
marqueurs de zones ⁵	interne ou externe	un matin de, un soir de, un jour de, au soir de, à la veille de, au début de, le début de, début, à la fin de, à l'aube de
marqueurs de points	Déterminants, prépositions, autres marqueurs	Ø, le, la...même, ce...là Ø, en, avant, après

⁴ Le classement obtenu n'a pas été établi de façon intuitive mais repose sur des paraphrases réglées. Ainsi les marqueurs d'intervalle sont identifiés à l'aide des marqueurs hyperonymes 'depuis', 'jusqu'en' ou 'au cours de' selon le type d'intervalle sollicité (*cela fait trente ans ≈ depuis trente ans*) ; les marqueurs de zone le sont au moyen de 'au début de' ; enfin les marqueurs de point, les plus difficilement interprétables, sont identifiés par les hyperonymes 'en' (*il y a trois ans, trois ans auparavant / plus tôt ≈ en 2004*) et 'dans' (*deux ans plus tard / après ≈ dans deux ans*).

⁵ La valeur des marqueurs de zone dépendra du grain de l'unité calendaire.

Ordonnement des EC repérées dans les textes

		il y a, dernier, auparavant, plus tôt, plus tard, suivante, avant, après
--	--	--

Tableau 3 : Modificateurs aspectuels dans les EC

3 Ordonnement chronologique des blocs textuels

Une des applications visées de notre travail est de proposer une lecture calendaire d'un texte en exploitant les fonctionnalités offertes par la plate-forme de navigation textuelle NaviTexte (Couto, 2006 ; Couto et Minel, 2006). Il convient donc de pouvoir ordonner chronologiquement les blocs textuels indexés par des EC. En effet, comme souligné dans (Le Draoulec et Péry-Woodley, 2005), s'il est indéniable que les dates jouent un rôle structurant majeur, elles ne remplissent pas forcément un rôle d'index, mais fonctionnent plutôt comme des jalons, des moments clés sélectionnés comme tels par l'auteur.

Cet ordonnancement suppose de traduire une représentation symbolique dans une représentation numérique de façon minimale sur laquelle sera appliqué un algorithme d'ordonnancement. Il s'agit donc de situer relativement les unes par rapport aux autres (donc ordonnancer) les zones temporelles décrites par les EC. Pour cela, nous devons en premier lieu calculer ces les-zones temporelles pour toutes les EC retenues (absolues et relatives), puis les ordonner ces zones. A partir de ce calcul, nous générons un S-langage (Schwer, 2007Schwer [??]) qui décrit l'ordonnancement des segments.

Commenté [CV5]: A compléter

3.2.13.1.1 De la double nature des périodes décrites par les occurrences des granularités

Par nature, chaque EC est associée à une unité calendaire (un « grain ») et possède une valeur aspectuelle. Une EC réfère *in fine* à une période de la droite métaphorique du temps découpée partiellement ou totalement en segments contigus et « numérotés », correspondant chacun à une occurrence de la granularité choisie, nommée *granule*. Chaque segment noté *g* est, pour la granularité correspondante, atomique mais néanmoins pourvu d'une extension non nulle, c'est-à-dire décomposable qualitativement en trois types de parties propres, comme tout segment : une partie contenant son extrémité gauche mais pas son extrémité droite, une partie contenant son extrémité droite, noté *[g]*, mais pas son extrémité gauche (*g*), et une partie contenant aucune de ses extrémités, noté (*g*), comme le montre la figure 1 (selon Schwer 2002).

Commenté [CV6]: Ajouté par CV

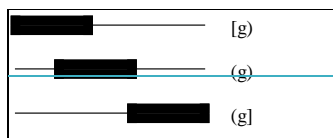
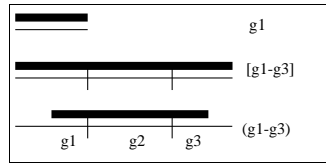


Figure 1 : Description symbolique d'une partie d'une occurrence de granularité



Figure_2 : Description symbolique d'un intervalle de granules

3.2.23.1.2 De l'expression qualitative des EC⁶

Chaque EC peut ainsi être représentée comme une ou partie ou intervalle de granule(s). Nombre d'EC (1-4) correspondent à ce genre de représentation. Ce sont ces représentations qualitatives (bien que numériques) qui vont étiqueter les segments textuels ; l'ordonnement de ces segments repose sur l'ordonnement de ces représentations.

EC	Granularité	Valeur aspectuelle	Représentation symbolique
Au début des années 80	Décennie	inchoatif	[198-)
Début 94	An	inchoatif	[1994)
A la fin de l'année 2006	An	terminatif	(2006]
Dans le mois de janvier 2007	Mois	sécant	(01-2007]
Mozart (1756-1791)	An	global	(1756—1791)
En 2002	ANEn-2002	global	2002

Commenté [CV8]: Cette colonne est-elle à compléter ?

Tableau 4 : Représentation symbolique des expressions calendaires

3.2.33.1.3 Ordonnement des EC qualitatives

Les expressions introduites ci-dessus suffisent à ordonner les EC repérées à chaque fois qu'elles correspondent à des occurrences de granularité disjointes, ce qui évite les conversions. Par exemple, *au début des années 80* et *fin 2006* correspondent à des segments temporels disjoints, car [198-) et (2006] correspondent à des décennies disjointes. Aussi, une simple comparaison arithmétique entre 198- et 200- suffit pour affirmer que *au début des années 80* précède chronologiquement *fin 2006*. En particulier, l'ordonnement des EC ponctuelles correspondant à une même granularité ne pose aucun problème. Toujours à granularité égale, ordonner des intervalles d'intersection vide ne pose pas de problème.

En revanche, les intervalles possédant des extrémités toutes différentes mais se chevauchant comme (2002—2008) et (2000—2004) nécessitent de *colorer* chacun des intervalles pour les retrouver. On écrira par exemple (2000(2002—2004)2008). Plus difficile est, *a priori*, le positionnement des EC dont les extrémités coïncident. Il est alors nécessaire d'utiliser le calcul d'intervalles de granules (Schwer, 2002^{be}), qui peut conduire à des indéterminations. En effet, deux expressions qualitatives de type [g] et (g] ne correspondent pas nécessairement

⁶ Nous nous limitons ici à quelques expressions classiques à titre d'illustration. Par exemple, nous ne traitons pas « à l'aube de » qui nécessite l'accès à une partie finissant de la granule précédente, ni des expressions itératives de type « tous les lundis », « chaque mois » par manque de place.

Ordonnement des EC repérées dans les textes

à la même portion de granule. Il n'est donc pas possible de les situer relativement l'une à l'autre. Il convient alors de laisser l'indétermination, à ce niveau d'analyse. Nous empruntons alors la notation de *produit cartésien* des S-langages (Battistelli et al. 2004) pour exprimer cette indétermination et notons $(g] \otimes [g]$.

Pour situer deux EC symboliques de granularité différente, dont l'extension temporelle de l'une contient l'autre, il est nécessaire de représenter l'expression contenante en un intervalle, si elle ne l'est pas déjà, comme par exemple, 2007 et 06-02-2007. ~~Pour cela, on transforme le point 2007 en un intervalle contenant deux extrémités identiques soit [2007 – 2007].~~ Remarquons qu'il s'agit bien d'un intervalle du point de vue de la seconde expression, c'est à dire pour une granularité plus fine. Nous pouvons alors calculer – d'après la loi de formation du calendrier utilisé – si la seconde expression contient la première ou bien si elle est la première (resp. la dernière) granule de la première expression.

D'autre part, certaines EC, comme par exemple *Au début des années 80* ou *début 94*, laissent place à une certaine indétermination. Aussi avons-nous exprimé des règles de transcription qui convertissent des marqueurs lexicaux (« début », « aube », « fin », « quelques », « mention d'un grain », « avant » etc.) en intervalle dont l'empan dépend de la granularité exprimée lexicalement. Il convient de souligner que ces choix d'empan sont subjectifs et pourront être modifiés. Par exemple ces règles, si le calcul le nécessite, transforment [198-) en [1980-1983] et [01-1980—12-1983] en [01-01-1980—31-12-1983].

Nous allons illustrer les règles de transformation sur un texte (T1) dans lequel nous ne laissons figurer que les expressions temporelles repérées et leur S-Mot (réduit ici à leur identité) associé noté entre parenthèses. On remarque que l'ordre narratif ne correspond pas à l'ordre chronologique des périodes temporelles évoquées dans le texte.

(T1) *Au début des années 80 (S1)...En février 1980 (S2)]...En 2002 (S3)...En décembre 1982 (S4)... En 1983 (S5)...Le 14 janvier 2002 (S6)...En décembre 1983 (S7)...*

Nous pouvons tout d'abord représenter graphiquement (cf. figure 3) l'ordonnement des EC symboliques : chaque ligne horizontale correspond à une granularité, sur laquelle l'antériorité est représentée par la position « sur la gauche ». Les flèches représentent les foncteurs reliant deux granularités (Schwer 2002b), positionner en début, milieu ou fin. Elles sont épaisses quand l'expression qualitative requiert une décision pour le calcul, c'est-à-dire par la présence d'un des signes « (» ou «) » dans l'expression. Les nœuds marqués du graphe représente une EC – à lire en suivant les flèches.

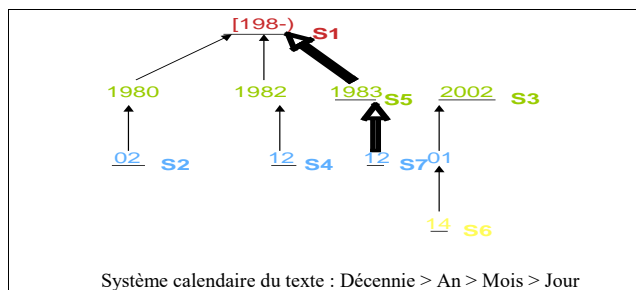


Figure 3 : Représentation graphique de l'ordonnement

3.3.2 Description d'un segment textuel en fonction de son EC associée

Nous avons vu qu'une EC possède, pour une granularité donnée, une valeur aspectuelle que nous avons qualifiée soit de ponctuelle (comme « le 21 juin 2003 » pour le grain Jour), soit de durative (comme « du 21 juin au 2 juillet 2006 ») soit d'inchoatif, terminatif ou sécant. Nous avons adopté le modèle des S-langages⁷ (Schwer 2007^{2a}) étiquetés pour décrire l'ordonnement des segments textuels. Ce modèle repose premièrement sur la définition d'un alphabet de base, en associant à chaque objet traité (ici un segment textuel S_i) une identité (ici son nom S_i). Deuxièmement, sur la représentation de chaque objet par un « mot » constitué d'un nombre déterminé (entier ou symbolique pour les expressions itératives) d'occurrences de son identité. Le segment S associé à une EC peut être représenté i) par le mot « S » de longueur un, étiqueté par l'EC qualitative ; ii) par le mot décoré « 'SS' » de longueur deux, la première occurrence étant étiquetée par la date du début de la période associée, la seconde occurrence par la date de la fin de la période.

Troisièmement, ce modèle repose sur la représentation d'une relation temporelle entre les objets comme S-mot. Il convient de noter que la représentation du segment textuel ne dépend pas directement de la nature aspectuelle de l'EC, mais des relations chronologiques qu'il entretient avec les autres segments. S'il est non contenant, il restera décrit par S, mais si son EC contient chronologiquement une autre EC (ou une partie) alors on substituera au mot « S » le mot « 'SS' » afin de pouvoir insérer un élément entre ses extrémités.

3.3.13.2.1 Algorithme d'ordonnement des blocs textuels

L'algorithme d'ordonnement, pour une chaîne de granularités, est décrit dans le tableau 5. Appliqué sur le texte T1, l'ordonnement chronologique des segments textuels de (T1) est alors donné par le S-mot : 'S1 S2 S4 'S5 { S7, S5', S1' } 'S3 S6 S3'. Plusieurs points sont à souligner à propos de l'algorithme : (i) Un minimum de calcul et de comparaison sont effectués pour réaliser l'ordonnement ; (ii) Le S-mot obtenu est une image linéarisée du système calendaire du texte, qui en respecte à la fois la dimension hiérarchique et l'ordre chronologique ; (iii) Contrairement à la théorie générale des S-langages, pour laquelle l'ordre des lettres à l'intérieur d'une S-lettre est sans importance, on constate ici que l'ordre d'écriture fait sens si l'on conserve le rapport hiérarchique entre les unités calendaires : cela explique l'ordre choisi à l'intérieur de la S-lettre { S7, S5', S1' } qui suit Mois < An < Décennie ; (iv) Il est possible, en conservant les étiquettes calendaires des segments d'insérer n'importe quel nouveau segment, à n'importe quel niveau de granularité, en modifiant légèrement l'algorithme (au lieu d'itérer par niveau de granularité, il est possible d'itérer en suivant les flèches du système calendaire du texte).

Données : une chaîne de granularités $g_1 > g_2 > \dots > g_n$ (g_1 la plus grossière) et pour chaque granularité g , la liste des segments textuels S_g étiquetés par une EC symbolique de granularité g .

Résultat : un S-langage écrit sur l'alphabet décoré des identités des segments textuels

⁷ Rappelons qu'à partir d'un ensemble fini, appelé alphabet, dont les éléments sont appelés lettres, on définit son S-alphabet comme l'ensemble de toutes ses parties non vides. Une S-lettre est un élément du S-alphabet. Un S-mot est une suite finie de S-lettres, un S-langage un ensemble de S-mots. Il est usuel de confondre une S-lettre réduite à une lettre avec cette lettre. Ainsi écrit-on $ac\{a,b\}$ le S-mot composé des quatre S-lettres $\{a\}$, $\{c\}$, $\{a,b\}$. L'opération $f \otimes g$ représente l'ensemble de toutes les situations possibles entre f et g , sans aucune contrainte.

Ordonnement des EC repérées dans les textes

<p>Traitement : Tant que la chaîne des granularités n'est pas vide, faire : Prendre la granularité la plus grossière de la chaîne et les segments associés, les éliminer des données. Ordonner les segments associés à cette granularité selon l'ordre chronologique de leurs EC symboliques ou calculées et représenter le S-langage associé, en étiquetant chaque occurrence de lettre par l'EC symbolique associée. Insérer le S-langage obtenu à l'intérieur du S-langage précédemment obtenu s'il existe. Fin de tant que Effacer les étiquettes des S-Lettres (optionnel)</p>
--

Tableau 5 : Algorithme d'ordonnement

4 Intégration dans NaviTexte

La navigation proposée s'appuie sur le logiciel NaviTexte. Dans celui-ci, la navigation est conceptualisée comme une opération reliant une UT *source* avec une UT *cible* (cf. tableau 6). La manière dont ces deux UT sont liées est fonction de quatre paramètres : i) la condition à vérifier par l'UT source ; ii) la condition à vérifier par l'UT cible ; iii) le type d'opération de navigation ; iv) la relation existant entre l'UT source et l'UT cible.

<pre>IF (condition UT_{source}) THEN : DO SELECT CRITERIA (Orientation, Ordre) WHERE {(condition UT_{cible}) AND (Relation (UT_{source}, UT_{cible})) ; : DO SHOW (Libellé de l'Opération) ;</pre>

Tableau 6 : Opération de navigation générique (Couto & Minel 2006)

Le type de l'opération permet de spécifier un ordre de recherche de la cible. Il est ainsi possible de spécifier que la recherche de la cible ne se fait pas en suivant l'ordre des UT dans le texte (ordre narratif), mais un ordre déclaré par l'intermédiaire d'un objet *Séquence*. C'est cette fonctionnalité que nous utilisons. Il convient donc de traduire le S-Mot calculé en un ou plusieurs objets proposés par le langage Sextant pour représenter un ordre de lecture. Deux cas sont à considérer.

Soit le S-Mot ne contient aucun ensemble (dénote par $\{\}$) auquel cas le S-Mot donne lieu à la construction d'objet de type *Séquence* (Couto 2006) ou chaque élément de la *Séquence* est un élément du S-Mot. Soit le S-Mot contient un ou plusieurs ensembles, ce qui signifie qu'il n'existe pas d'ordre total. Plusieurs objets textuels sont alors construits. D'une part, un objet *Séquence* qui représente l'ordre chronologique des segments textuels. Cette *Séquence* contient tous les éléments du S-Mot sans distinguer les débuts et les fins d'intervalles. Ainsi le S-Mot 'S1 S2 S4 'S5 { S7, S5', S1' } 'S3 S6 S3' produit l'objet *Séquence* S1 S2 S4 S5 S7, S5, S1 S3 S6 S3. En effet l'EC *En décembre 1983* (S7) désigne un intervalle qui est inclus dans l'intervalle désigné par *En 1983* (S5) et par conséquent une navigation à partir de cette expression doit proposer, comme période précédente mais aussi suivante le segment textuel indexé par *En 1983*. Les autres objets construits, de type *Ensemble*, contiennent les segments qui se chevauchent (dénote par les $\{\}$) et vont permettre de représenter la contemporanéité des segments textuels. Ainsi dans le cas du S-Mot calculé pour le texte T1, un objet *Ensemble* composé de S1 S2 S4 S5 S7 est produit en plus de l'objet *Séquence*.

Conclusion

Références

Commenté [CV10]: Références biblio complétées par CV

BATTISTELLI D., MINEL J.-L., PICARD E., SCHWER S. (2004). Temporalité linguistique et S-Langages. Actes *TALN 2004*, 33-38.

BATTISTELLI D., MINEL J.-L., SCHWER S. (2006). Représentation des expressions calendaires dans les textes : une application à la lecture assistée de biographies. *TAL* vol. X (article soumis).

CHRISTOVA V., DE VISMES O. (2005/2006). Repérage d'expressions calendaires et relations de dépendance. *Projet Annuel, Master 2 Professionnel ILGII*, Université Paris-Sorbonne, 182 p.

CONSTANT M. (2004). GRAAL, une bibliothèque de graphes : mode d'emploi. In C. Muller, J. Royauté et M. Silberstein (éds) *Cahiers de la MSH Ledoux 1 : INTEX pour la linguistique et le traitement automatique des langues*, Besançon, Presse Universitaire de Franche-Comté, 321-330.

COUTO J. (2006). *Modélisation des connaissances pour une navigation textuelle assistée. La plate-forme logicielle NaviTexte*. Thèse de l'Université Paris-Sorbonne.

COUTO J., MINEL J.-L. (2006). SEXTANT, un langage de modélisation des connaissances pour la navigation textuelle. Actes *ISDD'06*, Caen, 80-90.

GUIMIER C. (1996). *Les adverbes du français. Le cas des adverbes en '-ment'*. Paris : Orphys.

Harabagiu

Commenté [CV11]: A compléter

LE DRAOULEC A., PÉRY-WOODLEY M.-P. (2005). Encadrement temporel et relations de discours. *Langue française* 148, 45-60.

MAUREL D. (1988). Grammaire des dates. Etude préliminaire à leur traitement automatique. *Lingvisticae Investigationes* XII/1, 101-128.

MAUREL D. (1993). Reconnaissance automatique d'un groupe nominal prépositionnel. Exemple des adverbes de date. *Lexique* 11, 147-161.

PUSTEJOVSKY J., KNIPPEN R., LITTMAN J., SAURI R. (2005). Temporal and Event Information in Natural Language Text. *Language Resources and Evaluation* 39/2-3, 123-164.

SCHILDER F., HABEL C. (2001). From Temporal Expressions to Temporal Information : Semantic Tagging of News Messages. *Proceedings of ACL'01 workshop on temporal and spatial information processing*, 65-72.

SCHWER R. S. (2002a). S-arrangements avec répétitions. *Comptes Rendus de l'Académie des Sciences de Paris Série I* 334, 261-266.

SCHWER R. S. (2002b). Formalizing Calendars with the Category of Ordinals. *Applied Intelligence* 17 (3), 275-295.

Ordonnancement des EC repérées dans les textes

SCHWER R. S. (2002c). Reasoning with intervals on granules, *Journal of Universal Computer Science* 8 (8), 793-808.

[SCHWER R. S. \(2007\). Chronos, \(sous presse\)](#)

VAGUER C. (2006). Point sur les graphes édités pour le repérage et l'annotation des expressions temporelles calendaires dans un corpus de biographies. *Rapport Interne* du 9 janvier 2007 (36 p.).

VAZOV N. (2001a). A System for Extraction of Temporal Expressions from French Texts. *Actes de la conférence Traitement Automatique du Langage Naturel (TALN'2001)*, 313-322.

VAZOV N. (2001b). A System for Extraction of Temporal Expressions from French Texts Based on Semantic and Syntactic Constraints. *Proceedings of the ACL'2001 Workshop on Temporal and Spatial Information Processing*, 96-103.

Mis en forme : Français (France)