

Représenter la dynamique énonciative et modale de textes

Delphine BATTISTELLI¹, Marie CHAGNOUX²

¹ Lalic – Université Paris IV-Sorbonne, 28 rue Serpente 75006 Paris

² France Télécom – Div. R&D, TECH/EASY/Langues Naturelles,

2 avenue Pierre Marzin, 22307 Lannion Cedex

Delphine.Battistelli@paris4.sorbonne.fr, Marie.Chagnoux@free.fr

Résumé. Nous proposons d'exposer ici une méthodologie d'analyse et de représentation d'une des composantes de la structuration des textes, celle liée à la notion de prise en charge énonciative. Nous mettons l'accent sur la structure hiérarchisée des segments textuels qui en résulte ; nous la représentons d'une part sous forme d'arbre et d'autre part sous forme de graphe. Ce dernier permet d'appréhender la dynamique énonciative et modale de textes comme un cheminement qui s'opère entre différents niveaux de discours dans un texte au fur et à mesure de sa lecture syntagmatique.

Abstract. We propose a methodological framework for analyzing and representing the concept of commitment, which is one of the features characterizing textual structure. We emphasize the hierarchical structure of textual segments commitment conveys to. We represent it first as a tree and then as a graph. The latter enables us to access the modal and enunciative textual dynamics, as it shows the path followed through different discursive levels during the syntagmatic reading of a text.

Mots-clés : linguistique textuelle, énonciation, représentation sémantique.

Keywords: textual linguistics, enunciation, semantical representation.

1 Introduction

Notre approche vise à mettre l'accent sur la dynamique d'interprétation d'un texte appréhendée au travers des mécanismes de prise en charge énonciative qui instruisent une structuration de type hiérarchique. Nous proposons d'en donner deux types de représentation : la première met en évidence la structure énonciative et modale du texte, que nous représentons sous la forme d'un arbre ; la deuxième met en évidence la dynamique énonciative et modale du texte, que nous représentons sous la forme d'un graphe. Ces représentations vont toutes deux être convoquées pour deux types d'utilisations que nous visons : la première consiste seulement à rendre possible la construction automatique et la visualisation d'arbres et de graphes à l'écran ; la seconde consiste à utiliser ces représentations comme représentations internes d'un texte pour proposer des parcours entre segments dans le cadre d'une plateforme de navigation textuelle. Nous proposons dans cet article d'explicitier notre démarche. Dans la section 2, nous la situons par rapport aux approches existantes. La section 3 présente les étapes préliminaires de segmentation et d'annotation automatiques qui permettent de construire les représentations selon des modalités détaillées dans la section 4.

2 La prise en compte du contexte énonciatif et modal dans les systèmes de recherche d'information

Les systèmes actuels de recherche d'information tels que les résumeurs automatiques ou les systèmes de questions-réponses ne sélectionnent pas des documents entiers mais des *segments* correspondant à une requête ou à un profil à l'intérieur de ceux-ci. L'extraction de ces segments brise la continuité référentielle à l'œuvre dans un texte (Battistelli et Minel, 2006) : privés de leur contexte, les segments peuvent alors être difficiles à interpréter, en particulier quant à leur ancrage temporel, énonciatif et modal.

2.1 Exemples

L'absence de prise en compte du contexte énonciatif et modal peut entraîner des imprécisions, des contresens ou des conflits comme l'illustrent les trois exemples du tableau 1. Si l'on considère le segment *l'appel à un cessez-le-feu urgent et non pas immédiat constitue un échec*, on infère que l'auteur de l'article dont est extrait ce segment considère que l'appel à un cessez-le-feu est un échec. Or, si on le réinscrit dans sa continuité référentielle de l'extrait 1 du tableau 1, il apparaît que l'auteur ne prend en charge que le principe de citation et qu'un énonciateur second, la Maison-Blanche, réfute cette proposition.

1. La Maison-Blanche s'est, elle, montrée satisfaite des résultats de la réunion de Rome et a réfuté que <i>l'appel à un cessez-le-feu urgent et non pas immédiat constitue un échec</i> .
2. 1. Le Hezbollah a capturé deux soldats israéliens qui patrouillaient dans des jeeps blindées <i>sur le territoire libanais</i> à la frontière avec Israël.
2. 2. Le mercredi 12 juillet, 8 soldats de Tsahal ont été tués et 2 ont été kidnappés, au cours d'une attaque du Hezbollah <i>à l'intérieur du territoire israélien</i> , près de la frontière israélo-libanaise.
3. A propos du lieu de l'enlèvement, les versions diffèrent. Les Israéliens affirment qu'ils ont été capturés près de la ferme collective de Zarit <i>en territoire israélien</i> tout près de la frontière libanaise. De son côté, la police libanaise soutient que la capture s'est produite dans la région de Aïta al-Chaab <i>en territoire libanais</i> donc, proche de la frontière libano-sraélienne où une unité israélienne avait pénétré le matin même.

Tableau 1 : Extraits de textes portant sur le conflit entre Israël et le Liban durant l'été 2006

L'exemple 2 du même tableau montre que l'extraction de segments isolés de leur contexte énonciatif peut conduire à des conflits. Ainsi les exemples 2.1 et 2.2 sont contradictoires. Les deux extraits réfèrent à la même situation : la capture de deux soldats au Moyen-Orient le 12 juillet 2006. Or, selon l'énonciateur, le lieu de la capture est différent : pour l'association France Palestine, la capture a eu lieu sur le territoire libanais ; pour l'Ambassade d'Israël en France, elle s'est déroulée à l'intérieur du territoire israélien. La reconnaissance de la source peut paraître triviale dans le cas d'application inter-documents puisqu'il suffit d'annoter le texte pour le doter de cette information comme cela est proposé par exemple pour le Web sémantique. Cependant l'examen de l'exemple 3 du tableau 1 montre que le même conflit peut apparaître au sein d'un seul et même texte.

2.2 Annotation vs. Représentation

Le texte est la trace d'au moins un acte d'énonciation, cet acte est celui accompli par l'énonciateur principal qui prend en charge l'ensemble du discours. Cependant, certains segments de son discours ne sont pas complètement assumés par cet énonciateur, soit qu'un énonciateur second est convoqué au terme d'une citation, soit que ces segments sont assujettis à un certain degré de plausibilité ou d'intentionnalité. Il devient alors crucial pour les

systèmes de disposer de cette information qui instruit le contexte énonciatif et modal d'un segment textuel, comme en témoignent les trois exemples du tableau 1. Actuellement, de nombreux travaux s'inscrivent dans ce paradigme et proposent des systèmes d'annotation de ce type d'information (Giguet et Lucas 2004 ; Harabagiu et al., 2004 ; Sauri et al., 2005 ; Wilson et Wiebe, 2005). Nous pensons pour notre part que la seule tâche d'annotation des ruptures de prises en charge énonciative n'est pas suffisante et qu'il est nécessaire de les organiser en une structure représentationnelle qui rend compte des liens de type hiérarchique entre les segments annotés. En ayant ainsi repéré et hiérarchisé les différentes prises en charge énonciatives, un traitement informatique est à même de rendre compte de la véritable attribution de tel ou tel segment à tel ou tel énonciateur, que ce soit dans le cadre d'un traitement inter-documents ou intra-document.

2.3 Linguistique textuelle et systèmes de navigation textuelle

Pour des systèmes de navigation textuelle, qui proposent une alternative aux systèmes classiques de résumé automatique et de questions-réponses en permettant une navigation à l'intérieur d'un document parmi des segments annotés inscrits dans leurs co-textes (Bilhaut et al., 2003 ; Crestan et al., 2004 ; Minel et Couto, 2004), disposer de ce type d'informations constitue un enjeu important. En outre, il est intéressant de relever que ces systèmes s'appuient sur les structures discursives identifiées (le plus souvent de type hiérarchique) pour proposer une navigation à l'intérieur des documents ; ce faisant, ils sont directement concernés par les travaux récents de la linguistique textuelle qui a pour objet l'analyse des modes d'organisation discursifs (Péry-Woodley, 2001). Parmi ces modes, la temporalité occupe une place privilégiée très largement étudiée : par l'analyse des indices calendaires en position initiale de phrases qui constituent des marques d'ouverture de cadres temporels (Charolles et Vigier, 2005 ; Le Draoulec et Péry-Woodley, 2005), par l'analyse des variations de prises en charge énonciative et modale de contenus propositionnels (Desclés et Guentchéva, 2000 ; Kronning, 2003), etc.

La méthodologie de la linguistique textuelle consiste à établir le lien entre structures discursives et marques formelles organisées en systèmes. Elle se distingue en cela de celle mise en œuvre dans le cadre de modèles du discours comme ceux de la RST (Mann et Thompson 1988) ou de (Hobbs 1990) qui insistent sur l'indépendance du type de structures qui les intéresse par rapport à toute marque linguistique¹. Nous situant pour notre part dans une démarche d'analyse empirique des modes de structuration discursive – ici, énonciative et modale –, nous nous inscrivons dans les principes méthodologiques de la linguistique textuelle et donc aussi dans ceux qui fondent selon nous le développement de systèmes de navigation textuelle qui s'appuient directement sur les structures discursives mises en évidence. Nous exposons brièvement dans la section suivante l'étape qui précède celle de la construction de la représentation hiérarchique du texte : la tâche d'annotation. Elle consiste (i) d'une part à spécifier le type d'unités textuelles à annoter et donc à procéder à une segmentation du texte ; (ii) et d'autre part à préciser la typologie des annotations sémantiques à associer à ces unités.

¹ Il est à noter que l'approche de la SDRT (Asher, 1993) est à cet égard quelque peu différente. Il existe en effet depuis plusieurs années une volonté de lier le modèle aux instanciations langagières des relations de discours.

3 Segmentation et annotation des segments textuels

Nous présentons dans cette section les principes qui régissent l'annotation et en gouvernent l'automatisation. Au niveau discursif, le texte est considéré comme une suite de propositions en relations de rupture ou de continuité. Ces relations qui conduisent à considérer des segments homogènes regroupant plusieurs propositions sont définies lors de la construction de la représentation. Au niveau de l'annotation, le texte est segmenté en propositions par le système PROPOS (Wonsever, 2004) avant d'être annoté. Les critères d'annotation sont fondés sur les types de décrochage de prise en charge et sur la typologie des référentiels proposée par (Desclés, 1995) et précisée par (Chagnoux, 2006). Ces critères permettent d'inscrire les propositions sur différents référentiels :

- Le référentiel énonciatif *RE*, sur lequel le contenu propositionnel est complètement assumé par un énonciateur *E* ; il peut être global (noté *REG*) quand il s'agit de l'énonciateur premier ou local (noté *REL*) quand il s'agit d'un discours rapporté, direct ou indirect, ou médiatisé ;
- Le référentiel possible *RP*, sur lequel le contenu propositionnel est considéré comme possible éventuel (dans ce cas, le référentiel est noté *RPE*) ou possible contre-factuel (ici, le référentiel est noté *RPC*) ;
- Le référentiel mental noté *RM*, où s'inscrivent des propositions introduites par *il pense que*, *il croit que*, etc., à rapprocher des méta-représentations décrites par (Recanati, 2000).

L'objectif de cet article étant davantage de s'intéresser aux mécanismes de construction des structures hiérarchiques qui résultent de l'identification de ces différents référentiels dans les textes, nous ne présentons pas ici les analyses linguistiques détaillées sur lesquelles se fonde cette typologie. On pourra se reporter à (Guentchéva, 1996), (Desclés et Guentchéva, 2000) ou (Chagnoux, 2006) pour une présentation des analyses de discours directs, indirects, indirects libres et de la médiation.

L'identification et l'annotation des segments s'appuient sur la présence de marqueurs linguistiques de différentes natures : temps et modes des verbes conjugués, déictiques, constructions syntaxiques, lexèmes, adverbess, etc. Nous insistons sur le fait que ces catégories fonctionnent en systèmes et que c'est l'expression de patrons complexes qui rend compte des opérations de rupture ou au contraire de cohésion. Une base de marqueurs et une base de règles heuristiques ont été implémentées dans Chronotext² afin d'annoter les référentiels. Le tableau 2 détaille la segmentation et l'annotation d'un extrait court : l'extrait 1 est segmenté selon le format proposé en 2 par PROPOS (résumé en 3 pour une meilleure lisibilité) et annoté comme dans 4 par Chronotext.

Les annotations de 4 sont obtenues par les règles présentées dans le tableau 3 et qui peuvent être glossées comme suit : nous partons du postulat que tout texte s'inscrit dans un référentiel énonciatif global par défaut. Tant qu'aucun marqueur ne permet de les assigner à un autre référentiel, toutes les propositions relèvent de ce référentiel énonciatif global. Ici, c'est le cas de P1.

² Chronotext, présenté en détail dans (Chagnoux 2006) permet la reconnaissance des différents référentiels grâce à une base de 76 règles et de 708 marqueurs organisé dans 119 classes implémentés dans *Semantext*, plateforme de collecte et d'organisation de connaissances linguistiques pour l'annotation des textes.

1. Ses cadres affirment qu'ils ont les moyens de procéder à une escalade de grande envergure si Tsahal poursuit ses raids sur le sol libanais.
2. <PHRASE TEXTE><PROP COD=pli1 NIVEAU=1 INDEP VERBE = affirment>Ses cadres affirment </PROP><PROP COD=prl1 NIVEAU=2 INSERT IP=mot (que) VERBE = ont>qu'ils ont les moyens de procéder à une escalade de grande envergure ><PROP COD=gf21 NIVEAU=3 INDEP VERBE = poursuit>si Tsahal poursuit ses raids sur le sol libanais. </PROP></PROP></PHRASE_TEXTE
3. [P1 : Ses cadres affirment] [P2 : qu'ils ont les moyens de procéder à une escalade de grande envergure] [P3 : si Tsahal poursuit ses raids sur le sol libanais.]
4. <REG> Ses cadres affirment </REG> <REL> qu'ils ont les moyens de procéder à une escalade de grande envergure </REL> <RPE> si Tsahal poursuit ses raids sur le sol libanais.< /RPE>

Tableau 2 : Segmentation et annotation d'un extrait de texte

En revanche, la présence du patron *&ouverture_citation* + *que*³ permet d'identifier automatiquement que P2 relève d'un référentiel énonciatif local. L'annotation de P3 résulte de l'introduction de la proposition courante par *si* associée à la règle de concordance des temps. . Des règles de récursivité définissent des conditions de propagation qui permettent à l'annotation sémantique d'une proposition d'être associée à la proposition suivante afin d'annoter l'ensemble des propositions du texte.

1. NOT (\$proposition\$ contains (/ \$Referentiel-possible/ OR / \$Referentiel-mental/ OR /Referentiel-enonciatif2/) AND (\$proposition\$ contains(/&deictique/ OR /&interjection/)) AND NOT (\$proposition\$contains (/&naturalise/ OR /&PS/))
2. ((\$proposition justbefore \$proposition\$)contains /&ouverture-citation/) AND (\$proposition\$ contains/ &que/)
3. (\$proposition\$ starts-with /&si/) AND (\$proposition\$ contains /&present/) AND (\$proposition justbefore \$proposition\$ contains /&present/)

Tableau 3 : Exemples de règles d'annotation des segments

4 Représentation de la sémantique énonciative et modale d'un texte

La tâche de construction de la représentation oblige à spécifier les relations entre les unités distinguées lors de l'étape précédente. Le fait de considérer une structure textuelle hiérarchisée est classique en analyse du discours (Mann et Thompson, 1988 ; Hobbs, 1990, Asher 1993). Comme le souligne (Cornish 2006), le « scripteur » réalise linguistiquement une organisation qui passe par la constitution de segments, la hiérarchisation de ces segments et leur mise en relation. Autrement dit, un texte est vu comme un ensemble de segments qui entrent en relation d'inclusion ou de succession⁴ et il peut donc être représenté à l'aide d'une structure d'arbre. Nous proposons d'exposer ici les principes de construction de la représentation « statique » sous forme d'arbre de la structure de l'extrait (section 4.1) ; puis d'exposer ceux relatifs à la construction de la représentation sous forme de graphe de la « dynamique » du texte (section 4.2). Cette dernière permet de visualiser le cheminement qui s'opère entre différents niveaux de discours dans un texte au fur et à mesure de sa lecture

³ Dans *Chronotext*, *&ouverture_citation* est le nom de la classe contenant plus de 90 verbes susceptibles d'introduire les propos d'un tiers comme *s'exclamer, dire, hurler*, etc.

⁴ Ces relations peuvent éventuellement faire l'objet d'une typologie sémantique comme c'est le cas pour les relations dites de discours (narration, élaboration, ...) proposées par exemple par (Mann et Thompson 1998) ; il reste qu'elles renvoient toutes à une structuration arborescente du texte.

syntagmatique. Ces deux représentations sont construites à partir des segments annotés du texte présenté dans le tableau 4.

Hier après-midi, quarante-huit heures après l'ouverture du front libano-israélien, le Hezbollah qui, selon son chef Hassan Nasrallah, disposerait de plus de 10 000 roquettes et missiles pointés sur Israël, avait déjà frappé 165 fois le territoire de l'Etat hébreu. Et ce ne pourrait être qu'un début. Ses cadres affirment qu'ils ont les moyens de procéder à une escalade de grande envergure si Tsahal poursuit ses raids sur le sol libanais.

Tableau 4 : *Incipit* de l'article "Les chiites fous du Parti de Dieu", *Libération*, (15/072006)

4.1 Représenter la structure énonciative et modale sous forme d'arbre

Pour faire émerger la structure textuelle, l'annotation n'est plus seulement faite au niveau propositionnel par des balises XML mais se réalise au niveau segmental via un modèle de représentation du texte défini par (Couto 2006) et spécifié dans une DTD. Ce modèle permet de définir des Unités Textuelles (UT) à partir de propositions connexes mais surtout de traiter le problème de la discontinuité par une opération de Séquence qui rétablit la cohésion entre des segments discontinus dans le texte. A chaque UT est associé un attribut Nature qui définit la nature du segment. Chaque type de référentiel convoque des attributs différents qui permettent d'encoder toutes les propriétés qui lui sont associées. La figure 1 présente la mise au format du texte dont est extrait l'exemple 1 du tableau 2. La Tête contient les opérations de Séquence et le Corps contient les différentes propositions du texte.

```
<?xml version="1.0" encoding="iso-8859-1"?>
<!DOCTYPE Texte SYSTEM "DocumentNaviTexte.dtd">
<Texte>
  <Tete>
    <Sequence Type="Regroupement" Nro="1" >
      <UTP Type=" Segment_referentiel " Nro="1"/>
      <UTP Type=" Segment_referentiel " Nro="4"/>
    </Sequence>
  </Tete>
  <Corps>
    <UT Type=" Segment_referentiel " Nro="1">
      <Attribut Nom="Nature">Enonciatif_global</Attribut>
      <UT Type=" Proposition " Nro="1">
        <Chaine>Hier après-midi, quarante huit heures après l'ouverture du front libano-israélien, le
        Hezbollah avait frappé 165 fois le territoire de l'Etat hébreu</Chaine></UT>
      <UT Type=" Segment_referentiel " Nro="2">
        <Attribut Nom="Nature">Enonciatif_local
      </Attribut>
      <UT Type=" Proposition " Nro="2">
        <Chaine> qui, selon son chef Hassan Nasrallah, disposerait de plus de 10 000 roquettes et
        missiles pointés sur Israël, </Chaine></UT></UT>
      <UT Type=" Segment_referentiel " Nro="3">
        <Attribut Nom="Nature">Possible_eventuel</Attribut>
      <UT Type=" Proposition " Nro="3">
        <Chaine> Et ce pourrait n'être qu'un début </Chaine></UT></UT>
      <UT Type=" Segment_referentiel " Nro="4">
        <Attribut Nom="Nature">Enonciatif_global</Attribut>
      <UT Type=" Proposition " Nro="4">
        <Chaine> Ses cadres affirment</Chaine></UT></UT>
      <UT Type=" Segment_referentiel " Nro="5">
        <Attribut Nom="Nature">Enonciatif_local</Attribut>
      <UT Type=" Proposition " Nro="5">
        <Chaine> qu'ils ont les moyens de procéder à une escalade de grande envergure
      </Chaine></UT></UT>
      <UT Type=" Segment_referentiel " Nro="6">
        <Attribut Nom="Nature">Enonciatif_local</Attribut>
      <UT Type=" Proposition " Nro="6">
        <Chaine> si Tsahal poursuit ses raids sur le sol libanais </Chaine></UT></UT></UT></UT>
      </Corps>
    </Texte>
```

Figure 1. Texte annoté en conformité avec la DTD définie dans (Couto 2006)

Représenter la dynamique énonciative et modale de textes

A partir de cette organisation des segments, un arbre est généré via Graphviz⁵. La figure 2 présente le graphe associé à cet extrait. La racine de l'arbre correspond à *REG* ; les nœuds correspondent aux différents référentiels sur lesquels s'inscrivent les propositions du texte.

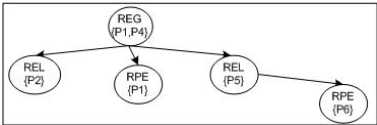


Figure 2 : Arbre de la structure énonciative et modale de l'extrait 2

4.2 Représenter la dynamique énonciative et modale sous forme de graphe

Le phénomène des différentes prises en charge énonciative à l'œuvre dans un texte peut être rendu par une structure de *graphe*. Dans ce cas, on met l'accent sur la *dynamique* énonciative interne du texte comme l'illustre le tableau 5 où se construit progressivement le graphe associé au texte du tableau 3.

Toute énonciation ouvre un référentiel énonciatif global (REG) : on construit ici REG, premier nœud du graphe sur lequel on situe P ₁ .	<div>REG {P₁}</div> <div>1</div>
Bien que situé sur REG, P ₁ contient des indices de rupture énonciative (<i>selon</i> suivi d'un emploi du conditionnel) qui ouvrent un référentiel énonciatif local REL.	<div>REG {P₁}</div> <div>1</div> <div>REL 2</div> <div>1 → 2 ("selon" + COND)</div>
C'est sur ce référentiel que s'inscrit la proposition P ₂ . La fin de la proposition relative permet de fermer le référentiel REL et de revenir sur REG. P ₃ contient un indice de rupture modale (<i>pouvoir</i>) qui ouvre un référentiel du possible éventuel RPE où situer P ₃ . Il se referme grâce à la présence du signe de ponctuation ".	<div>REG {P₁}</div> <div>REL {P₂}</div> <div>RPE {P₃}</div> <div>1</div> <div>2</div> <div>3</div> <div>1 → 2 ("selon" + COND)</div> <div>2 → 1 (relative)</div> <div>1 → 3 ("pouvoir")</div> <div>3 → 1 (".")</div>
P ₄ contient un indice de rupture énonciative (<i>affirmer que</i>) qui ouvre un référentiel énonciatif local REL où se situe P ₅ .	<div>REG {P₁, P₄}</div> <div>REL {P₂}</div> <div>RPE {P₃}</div> <div>REL {P₅}</div> <div>1</div> <div>2</div> <div>3</div> <div>4</div> <div>1 → 2 ("selon" + COND)</div> <div>2 → 1 (relative)</div> <div>1 → 3 ("pouvoir")</div> <div>3 → 1 (".")</div> <div>1 → 4 ("affirmer" + que)</div>

⁵ GraphViz est un logiciel libre qui permet de générer automatiquement des arbres et des graphes.

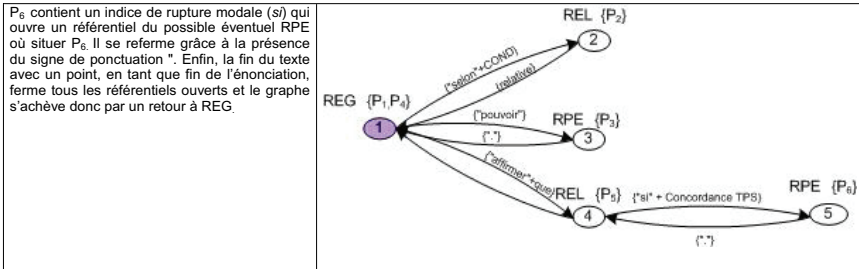


Tableau 5 : Etapes de construction du graphe de la dynamique énonciative et modale du texte

Les nœuds correspondent toujours aux référentiels identifiés. Les arcs indiquent les parcours entre les différents niveaux de discours et sont étiquetés avec les marqueurs qui font passer d'un niveau à un autre. Il faut souligner que cette structure encode la structure arborescente sous-jacente puisque les bi-arcs entre nœuds correspondent exactement aux liens de l'arbre. Lors de la construction de la structure textuelle, la principale difficulté réside dans l'identification automatique de la fermeture des segments repérés. Cette difficulté se pose également dans les autres modèles qui rendent compte de la structure textuelle (Charolles et Vigier 2005), diverses stratégies sont mises en œuvre pour résoudre ce point comme par exemple dans cet extrait le retour systématique au premier nœud du graphe en fin de texte.

5 Conclusion

Dans cet article, nous avons proposé un modèle de représentation de la dynamique énonciative et modale ainsi que son automatisisation via la chaîne de traitement que synthétise la figure 3.

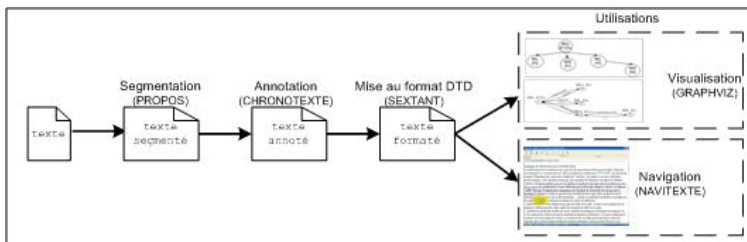


Figure 3: Chaîne de traitement pour l'automatisation de la génération de graphes

Cette chaîne est actuellement en cours de réalisation et est testée sur un corpus de 25 textes (5 textes journalistiques, 5 textes tirés de la Wikipedia, 5 incipit de roman, 5 textes scientifiques et 5 textes "entreprises"). Les tâches de segmentation et d'annotation sont opérationnelles et nous travaillons actuellement d'une part à la mise au format automatique dans la DTD choisie et d'autre part à la génération automatique de l'arbre et du graphe via Graphviz. Il nous faut encore résoudre un certain nombre de difficultés : lors de l'annotation automatique des propositions, il est parfois difficile de distinguer automatiquement la nature de tous les référentiels. Ainsi par exemple, le repérage des assertions introduites par "selon" ou "pour" en position antéposée produit encore trop de bruits en identifiant des propositions comme "selon

l'appartenance sociale" ou "pour le développement". Les règles qui traitent les phénomènes modaux doivent encore être affinées pour permettre une meilleure annotation. Il faudrait également à terme pouvoir regrouper les segments discontinus pris en charge par un même énonciateur.

En ce qui concerne l'évaluation, comme l'explicite (Chaudiron, 2004), la difficulté qu'il y a à construire des protocoles d'évaluation est particulièrement accrue dans le domaine sémantique. A défaut de pouvoir évaluer la pertinence de la représentation sémantique proposée ici, nous pensons qu'illustrer son utilisation pour d'autres tâches permet de valider notre approche. Les premiers résultats que nous avons obtenus montrent d'une part qu'il nous est possible de traiter tout type de texte et, par là même, de proposer une typologie basée sur les mécanismes énonciatifs (Battistelli et Chagnoux, 2006). D'autre part, le format d'encodage des informations permet un certain nombre d'opérations de navigation à l'intérieur d'un texte dans une plate-forme comme celle de (Couto et Minel, 2006). Notre approche se distingue de la « simple » approche d'annotation des ruptures de prise en charge énonciative et modale puisque nous proposons de construire, à partir des textes annotés, des représentations des structures discursives. Notre modèle de représentation discursive prend non seulement en compte les ruptures de prises en charge énonciative mais les organise en une structure hiérarchique, ce qui constitue un enjeu fondamental pour les systèmes actuels de recherche d'information. Par ailleurs, ce projet prend toute sa légitimité dans le cadre de plateformes de navigation textuelle qui s'appuient directement sur ces structures pour proposer des parcours de navigation entre segments textuels.

Références

- ASHER N. (1993). *Reference to Abstract objects in Discourse*. Kluwer Academic Publishers.
- BATTISTELLI D., CHAGNOUX M. (2006). Vers une typologie de mécanismes discursifs. *Actes de la journée ATALA "Typologie de textes"*, Paris.
- BATTISTELLI D., MINEL J.-L. (2006). Les systèmes de résumé automatique : comment assurer une continuité référentielle dans la lecture des textes, in G. Sabah (Ed.), *Compréhension des langues et interaction*, 295-330.
- BATTISTELLI D., MINEL J.-L., SCHWER S. (2006). Représentation des expressions calendaires dans les textes : vers une application à la lecture assistée de biographies. *TAL* vol. 47/2.
- BILHAUT F., HO-DAC M., BORILLO A., CHARNOIS T., ENJALBERT P., LE DRAOULEC A., MATHET Y., MIGUET H., PÉRY-WOODLEY M.-P., SARDA L. (2003). Indexation discursive pour la navigation intradocumentaire : cadres temporels et spatiaux dans l'information géographique, *Actes TALN'03*, Batz-sur-Mer, 315-320.
- CHAGNOUX M. (2006). Temporalité et aspectualité dans les textes français. Modélisation sémantico-cognitive et traitement informatique. *Thèse de l'Université Paris-Sorbonne*.
- CHAROLLES M., VIGIER D. (2005). Les adverbiaux en position préverbale : portée cadrative et organisation des discours, *Langue Française*, 148, 9-30.
- CHAUDIRON, S. (2004). *L'évaluation des systèmes de traitement de l'information*. Lavoisier.
- 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.

CORNISH F. (2006). Relations de cohérence en discours : critères de reconnaissance, caractérisation et articulation cohésion-cohérence, Journée d'étude du CRISCO *Organisation des textes et cohérence des discours*, Université de Caen, texte publié en ligne en 2006

CRESTAN É., DE LOUPY C., MANIGOT L. (2004). Analyses sémantiques pour la navigation textuelle ; Actes *CIDE 7*, La Rochelle ; 22-25 juin 2004, 293-308.

DESCLES J.-P. (1995). Les référentiels temporels pour le temps linguistique. *Modèles linguistiques* 16, 9-36.

DECLÈS J.-P., GUENTCHEVA Z. (2000). Enonciateur, locuteur, médiateur. Erikson Ph et Monod-Becquelin A (éds). *Les rituels du dialogue*. Nanterre : Société d'ethnologie.

GUENTCHEVA Z. (éd.) (2000). *L'énonciation médiatisée*. Louvain-Paris : Peeters.

GIGUET E., LUCAS N. (2004). La détection automatique des citations et des locuteurs dans les textes informatifs. *Le Discours rapporté dans tous ses états : question de frontières*, Paris, L'Harmattan.

HARABAGIU M., MAIORANO S.J., MOSCHITTI A., BEJAN A.C. (2004). Intentions, Implicatures and Processing of Complex Questions. Actes *HLT-NAACL Workshop on Pragmatics of Q. A.*

HOBBS J.R. (1990). Chap. 5 : The coherence and structure of discourse, *Literature and Cognition*, Leland Stanford Junior University, Calif: *CSLI Lecture Notes* 21, 83-114.

KRONNING H. (2003). Modalité et évidentialité. Birkelund, M., Boysen, G. & Kjærsgaard, P. S. (éds). *Aspects de la Modalité, Linguistische Arbeiten* 469, 131-151.

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

MANN W.C., THOMPSON S.A. (1988). Rhetorical Structure Theory: toward a functional theory of text organization, *Text* 8(3), 243-281.

PÉRY-WOODLEY M.-P. (2001). Présentation du numéro : Cohérence et relations de discours à l'écrit, *Verbum*, XXIII, 1.

RECANATI F. (2000). *Oratio Obliqua, Oratio Recta : An Essay on Metarepresentation*. Cambridge : MIT Press.

SAURI R., VERHAGEN M., PUSTEJOVSKY J. (2005). Annotating and recognizing Event Modality in Text. Actes *FLAIRS '06*, Melbourne Beach, Florida.

WILSON T., WIEBE J. (2005). Annotating Attributions and Private States, *Proceedings of the ACL Workshop on Frontiers in Corpus Annotation II: Pie in the Sky*, Ann Arbor, Michigan.

WONSEVER D. (2004). Repérage automatique des propositions par exploration contextuelle. *Thèse de l'Université Paris-Sorbonne*.