



**HAL**  
open science

## Informer sans s'engager : variations de prise en charge énonciative dans les sujets d'actualité

Marie Chagnoux

► **To cite this version:**

Marie Chagnoux. Informer sans s'engager : variations de prise en charge énonciative dans les sujets d'actualité. 2008. hal-00336855

**HAL Id: hal-00336855**

**<https://hal.science/hal-00336855v1>**

Submitted on 5 Nov 2008

**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.

Informers sans s'engager : variations de prise en charge énonciative dans les sujets d'actualité

Marie Chagnoux

IRIT  
118 route de Narbonne  
31062 Toulouse-Cedex 4  
France  
[marie.chagnoux@irit.fr](mailto:marie.chagnoux@irit.fr)

## Informer sans s'engager : variations de prise en charge énonciative dans les sujets d'actualité

Chagnoux Marie

IRIT

marie.chagnoux@irit.fr

<http://marie.chagnoux.free.fr>

La possibilité de faire varier la prise en charge énonciative au sein d'un même discours permet aux auteurs de sujets d'actualité d'informer sans s'engager sur la véracité de leurs propos, voir même de faire figurer dans un même article des informations contradictoires. Nous proposons dans cet article un cadre méthodologique pour repérer et représenter la dynamique textuelle liée à la notion de prise en charge énonciative sous forme de graphe. La profondeur du graphe associé à un texte rend alors compte de la complexité des mécanismes de variations de prise en charge. L'objectif de cette approche, directement implémentable, est de faire émerger automatiquement d'un corpus les textes sujets à controverses.

représentation discursive, énonciation, citation, graphe

fr

# 1. Informer sans s'engager

La révolution médiatique opérée par Internet a renforcé ces dernières années un certain nombre de phénomènes qui modifient sensiblement le caractère supposé objectif de l'information délivrée sur le support électronique par opposition aux supports plus traditionnels. À titre d'exemples, citons trois phénomènes distincts qui illustrent cette évolution. Le premier est corrélé à l'essor des sites collaboratifs. L'encyclopédie *Wikipedia*<sup>1</sup> propose des articles écrits à plusieurs mains : l'information délivrée n'est alors plus le point de vue d'un auteur sur un objet mais la synthèse d'une pluralité de points de vue. Le second réside dans l'extrême rapidité avec laquelle une information est disponible. La mise en ligne immédiate des données de l'AFP ne permet pas toujours d'en vérifier la véracité et il revient alors au journaliste de préciser le caractère incertain de l'information qu'il est en train de relater. Ainsi dans les heures qui suivent un accident d'avion, le bilan des morts et des blessés est toujours donné au conditionnel et ce n'est que plus tard que le bilan définitif sera exprimé au mode indicatif comme le montrent les extraits [1] et [2] tirés du site de *l'Express*<sup>2</sup>.

[1] L'accident d'avion à Phuket **aurait fait** 61 morts et 40 disparus.<sup>3</sup>

[2] L'accident d'un avion de ligne thaïlandais avec 130 personnes à bord, dimanche sur l'aéroport de Phuket (sud), **a fait** 89 morts, selon un dernier bilan.<sup>4</sup>

Le troisième est lié à l'internationalisation du support : l'information délivrée peut être lue par tous et partout dans le monde. Elle est donc accessible à des communautés de culture, de sensibilité et d'opinions différentes et le choix du lexique trahit souvent l'appartenance de l'auteur à une communauté : le même individu peut à la fois être vu comme un terroriste en Russie et comme un résistant en Tchétchénie. Ainsi, à sa mort, Chamil Bassaev fut qualifié de « terroriste numéro un » de la Russie<sup>5</sup> par un quotidien français tandis qu'un autre le désignait plus honorablement comme le « chef de guerre tchétchène »<sup>6</sup>. Tous ces facteurs incitent de plus en plus les auteurs d'articles en ligne à déléguer la responsabilité d'une information à un tiers ou à insister sur le caractère plausible ou incertain de l'information donnée. Dans un cas comme dans l'autre, la présence d'indices linguistiques permet de repérer les mécanismes utilisés par les auteurs pour se distancier d'un énoncé qu'ils ne prennent plus vraiment en charge.

La réticence des journalistes à s'engager sur la véracité des informations qu'ils délivrent a été particulièrement perceptible durant l'été 2006 autour du conflit armé israélo-palestinien où des sources d'information contradictoires s'affrontaient. Nous avons à cette époque collecté un corpus varié contenant à la fois des textes pro-israéliens, des textes pro-libanais et des textes visant *a priori* à l'objectivité. D'une part, l'intérêt de ce corpus était de montrer comment un même événement, la capture de deux soldats israéliens, pouvait être narré par deux communautés antagonistes, en l'occurrence la communauté pro-israélienne et la communauté pro-libanaise. D'autre part, il s'agissait d'analyser comment les médias arrivaient à concilier les deux points de vue sans prendre parti. Ce corpus d'une vingtaine de textes nous a permis de mettre en oeuvre le cadre méthodologique général proposé dans [Chagnoux 2006] pour rendre compte des variations de prise en charge énonciative et montrer qu'analyser la structure

---

<sup>1</sup> <http://fr.wikipedia.org>.

<sup>2</sup> Bien que nous opposions indicatif et conditionnel dans cet exemple, à l'instar, de [D. Leeman-Bouix 1994], nous pensons qu'il est possible de considérer le conditionnel comme un temps de l'indicatif.

<sup>3</sup> Site de *l'Express*, <http://www.Lexpress.fr>, le dimanche 16 septembre 2007, mis à jour à 14:17.

<sup>4</sup> Site de *l'Express*, <http://www.Lexpress.fr>, le lundi 17 septembre 2007, mis à jour à 14:34.

<sup>5</sup> *Libération*, 1 juillet 2006.

<sup>6</sup> *Le Monde*, 10 juillet 2006.

discursive fait émerger des éléments de nature typologique qui permettent de repérer les sujets d'actualité polémiques.

Pour présenter ces travaux, nous exposerons dans un premier temps le cadre méthodologique utilisé pour modéliser les variations de prise en charge énonciative. Dans un second temps, nous montrerons comment ce cadre est mis en oeuvre pour associer une représentation graphique à la structure discursive. Enfin, la dernière partie présentera l'exploitation des graphes obtenus ainsi que les différentes phases d'implantation qui automatise le processus de construction des graphes.

## 2. Repérer et modéliser les variations de prise en charge énonciatives

### 2.1. Les variations de prise en charge énonciative

Comme le définit [Culioli 1980:184], le concept de prise en charge repose sur le fait de « dire ce qu'on croit (être vrai) ». Ainsi, tout énoncé suppose la prise en charge implicite des propos. Dans les exemples [3], en énonçant des faits, les énonciateurs assurent implicitement de la véracité de leurs propos.

[3a] Au sud, à Rafah, le point de passage est toujours fermé par l'armée d'occupation, une femme et un enfant y sont morts hier.<sup>7</sup>

[3b] 8 Israéliens ont été tués jeudi 3 août, dans l'après-midi, suite à une salve de roquettes tirées par le Hezbollah sur la Galilée occidentale.<sup>8</sup>

Cependant, parce que les langues permettent de dire explicitement ce que l'on croit être plus ou moins vrai, tout énonciateur dispose de différents marqueurs grammaticaux, syntaxiques ou lexicaux pour prendre plus ou moins en charge un énoncé. De l'effacement énonciatif étudié par [Rabatel 2006] aux variations journalistiques proposées par [Rosier 1993], il existe une gradualité de la prise en charge. Les extraits du corpus de l'exemple [4] illustrent la diversité des mécanismes langagiers par laquelle un énonciateur peut exposer des faits ou des opinions sans en assumer complètement la prise en charge. Ces mécanismes peuvent être de deux types : les phénomènes citationnels attribuent la prise en charge à un énonciateur extérieur alors que les phénomènes de modalisation du discours instaurent une distanciation entre l'énonciateur et ses propos.

[4a] **Selon** le ministre canadien, il y **aurait** huit ressortissants canadiens parmi les victimes.

[4b] **Selon** un général, des commandos terrestres sont entrés sur le territoire libanais.

[4c] Le 30 juillet, Human Rights Watch **considère qu'**Israël a commis des crimes de guerre et le juge responsable des victimes du bombardement de Cana[...]

[4d] « Les deux prisonniers israéliens ont été transférés dans un lieu sûr », **a déclaré** le Hezbollah.

[4e] La police libanaise **affirme que** l'aviation israélienne a entièrement détruit un lieu de prière et de rassemblement chiite à Budaï, près de Baalbeck, dans l'est du Liban.

[4f] Le 31 juillet, un patrouilleur israélien Sa'ar 4 **aurait été atteint** au large de Tyr par un missile inconnu du Hezbollah, bien que l'armée israélienne ait démenti l'information.

[4g] Cela **pourrait** également ajouter des dégâts terrestres à la catastrophe écologique.

---

<sup>7</sup> "L'armée israélienne pénètre sur le territoire libanais, les avions survolent Beyrouth.", <http://www.france-palestine.org/article4161.html>

<sup>8</sup> "L'offensive du Hezbollah – Chronologie des événements", Site de l'ambassade d'Israël en France, <http://paris1.mfa.gov.il/mfm/web/main/missionhome.asp?MissionID=31&>.

### 2.1.1. Les phénomènes citationnels

Pour marquer sa distance avec les situations exprimées, l'énonciateur peut avoir recours aux mécanismes de polyphonie énonciative<sup>9</sup> qui instaurent une rupture énonciative marquée. Ainsi, dans les exemples [4a-e], l'énonciateur convoque un tiers, respectivement « le ministre canadien », « un général », « Human Rights Watch », « le Hezbollah » ou « la police libanaise ». Dès lors, ce n'est pas le journaliste qui assume qu'« Israël a commis des crimes » mais un énonciateur second, l'organisme Human Rights Watch. Ces exemples montrent comment le recours à la citation permet de présenter des faits sans en assumer la portée. Classiquement, la citation est abordée en tant que discours rapporté<sup>10</sup> : on distingue ainsi le discours rapporté, le discours rapporté indirect et l'assertion contextualisée.

Le discours direct se définit comme la reproduction fidèle de l'énoncé d'un tiers qui « dissocie deux systèmes énonciatifs » [Maingueneau 1986 : 97] comme dans l'exemple [4d] qui fait coexister deux énoncés parfaitement distincts : un énoncé cité, *Les deux prisonniers israéliens ont été transférés dans un lieu sûr*, et un énoncé citant, *a déclaré le Hezbollah*. Les deux énoncés sont articulés par des marqueurs typographiques (ici les guillemets) et lexicaux (ici un verbe introducteur). Tous les deux sont syntaxiquement indépendants.

Le discours indirect se définit comme « la traduction de l'énonciation citée » [*ibid*] comme dans les exemples [4c] et [4e]. L'énoncé est transposé pour s'intégrer au reste du texte sur le plan morphologique comme sur le plan syntaxique. Il est introduit par un verbe qui permet souvent à l'énonciateur de porter un jugement sur la situation comme dans l'exemple [5] où le verbe introducteur *croit savoir* infirme la véracité de l'information.

[5] Sept soldats israéliens auraient également été tués dans de violentes attaques lancées par le Hezbollah, **croit savoir** la chaîne d'information qatarie Al-Jazira

Enfin l'assertion contextualisée se définit comme l'expression du point de vue d'un tiers [Guentchéva 2000], comme dans les exemples [4a] et [4b]. Les énoncés ne sont assumés que dans le contexte déterminé par le tiers : *selon le ministre canadien* ou *selon un général*. Dans les faits, l'hétérogénéité énonciative est bien plus complexe que cette rapide typologie pourrait le laisser supposer : de très nombreux travaux linguistiques proposent des descriptions détaillées de tous les mécanismes citationnels comme la synthèse détaillée proposée par [Rosier 1999] ou l'analyse fine des différents types de discours et de leurs emboîtements de [Maingueneau 1986]. Si nous avons utilisé ces analyses pour la modélisation et la reconnaissance des citations, nous ne les présentons pas ici de manière plus exhaustive.

### 2.1.2. Les phénomènes de modalisation

Alors que les phénomènes citationnels portent sur les conditions d'énonciation (qui assume l'énoncé ?), les phénomènes de modalisation portent sur la vérité de l'énoncé (comment est assumé l'énoncé ?). En effet, pour se distancier des propos qu'il tient, un énonciateur peut également recourir aux modalités qu'elles soient de nature alétique, temporelle ou épistémique. Ainsi, il peut utiliser le mode conditionnel comme dans [4f] ou [4g] ou recourir à des verbes modaux comme *pouvoir* dans [4g].

Comme pour les phénomènes citationnels, notre présentation des phénomènes de modalisation est très succincte car elle vise essentiellement à résumer des connaissances nécessaires pour comprendre nos travaux dans le champ de la modalité, champ qui reste

---

<sup>9</sup> En linguistique, selon la tradition on parle de polyphonie, d'hétérogénéité énonciative [Authier 1981] ou de dialogisme, la distinction n'est pas pertinente ici et comme le signale [Rabatel 2006], les deux conceptions se rejoignent en ce qu'elles « mettent fin au mythe de l'unicité du sujet parlant ».

<sup>10</sup> Bien certains auteurs préfèrent réserver le terme de *citation* au discours rapporté direct (Tuamarla 2000), nous l'utilisons comme le terme générique permettant de référer à tout discours rapportant les propos d'un tiers.

encore à définir comme le souligne [Martin 2005]. Le traitement des phénomènes modaux que nous avons proposé dans notre analyse s'inspire largement de la typologie proposée par [Le Querler, 1996] et de l'approche plus applicative de [Gosselin 2006]

Nous souhaitons montrer qu'à travers des phénomènes citationnels ou de modalisation, un énonciateur peut présenter des faits et des opinions en les prenant plus, moins ou même pas du tout en charge. Dans tous les cas, on observe un mécanisme de rupture dans la prise en charge énonciative qui distingue le segment cité ou modalisé du reste de l'énoncé. Dans le cas de la citation, la rupture énonciative repose sur un changement d'énonciateur comme dans les exemples [3a-f] alors que dans le cas de la modalisation, la rupture énonciative repose sur un changement de statut de l'énoncé : l'énonciateur introduit un commentaire réflexif sur son propre énoncé comme dans l'exemple [4g]. La comparaison des exemples [4a] et [4b] montre que les deux types de mécanismes peuvent s'associer, factorisant ainsi le nombre de variations de degré de prise en charge. L'information contenue dans [4b] est moins certaine que dans [4a]. À nouveau, nous insistons sur le fait que les phénomènes linguistiques que nous citons ici relèvent de problématiques largement débattues et de traitements plus complexes, mais cette présentation succincte permet de rappeler rapidement au lecteur les principes sur lesquels se fonde l'analyse.

## 2.2. Ruptures énonciatives et cohérence discursive

Ainsi, tous les exemples de [4] ont en commun le fait de situer le propos sur un référentiel<sup>11</sup> autre que celui de l'énonciateur, que ce soit sur le référentiel d'un tiers ou sur un référentiel modalisé de l'énonciateur lui-même. Les marqueurs de surface permettent au lecteur de passer sans ambiguïté d'un référentiel à l'autre en articulant les référentiels entre eux. L'exemple [6] témoigne de la possible coexistence dans un même texte de plusieurs référentiels de nature énonciative.

[6] À propos du lieu de l'enlèvement des deux soldats, les versions diffèrent. Les Israéliens indiquent qu'ils ont été capturés près de la ferme collective de Zarit en territoire israélien 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 en territoire libanais donc proche de la frontière libano-israélienne où une unité israélienne avait pénétré le matin même.<sup>12</sup>

On peut ici relever trois référentiels énonciatifs distincts :

- Le référentiel de l'énonciateur sur lequel sont validés trois faits - la divergence des versions de la capture et les deux déclarations antagonistes ;
- le référentiel des Israéliens introduit par « Les Israéliens indiquent qu' » et sur lequel est validé le fait que la capture a eu lieu près de la frontière libanaise (et donc en territoire israélien) ;
- le référentiel de la police libanaise introduit par « la police libanaise soutient que » et sur lequel est validé le fait que la capture a eu lieu en territoire libanais.

Seule la prise en compte de la structure discursive balisée par les marqueurs de surface permet de comprendre comment deux informations contradictoires peuvent figurer dans un même texte sans contrarier la cohérence de ce dernier.

Afin de rendre compte de la structure discursive, notre objectif est de modéliser les ruptures référentielles qui articulent entre eux différents segments textuels en repérant (i) les marqueurs de cohésion ou de rupture qui structurent le discours et (ii) les segments cohérents qui forment des blocs homogènes à l'intérieur d'un même référentiel. À partir de l'identification des segments et des articulations, le texte peut être considéré non comme une

---

<sup>11</sup> Nous utilisons ici le terme de référentiel tel qu'il est utilisé en physique : en physique, un référentiel est un système de coordonnées spatiales et temporelles lié à un observateur. Un même événement sera différent selon l'emplacement où se trouve cet observateur et selon les repères qu'il a.

<sup>12</sup> « L'armée israélienne investit le Liban sud », [http://www.rfi.fr/actufr/articles/079/article\\_45021.asp](http://www.rfi.fr/actufr/articles/079/article_45021.asp).

suite linéaire de propositions mais comme une organisation structurée et hiérarchisée de segments textuels.

### 2.2.1. Identifier les ruptures de cohérence discursive

D'autres cadres théoriques ont déjà abordé la question de la structuration de surface des textes. Le postulat commun est que la structure discursive s'élabore dans des espaces de validation qui s'ouvrent, se ferment et éventuellement s'emboîtent et à l'intérieur desquels le principe de cohérence est respecté. Une première approche vise à rechercher localement les segments sémantiquement cohérents. Ainsi [Charolles 1997] cherche à repérer des cadres de discours : chaque cadre de discours correspond à un segment sémantiquement homogène selon des critères thématiques [Porhiel 2001], temporels [Terran 2002], spatio-temporels [Charolles *et al.* 2005], etc. Il est également possible de citer les travaux sur le médiatif menés par [Schrepfer 2006]. Comme les référentiels, les cadres de discours peuvent également être considérés comme des équivalents discursifs des espaces mentaux de [Fauconnier 1985]. Une autre approche consiste à représenter globalement la structure textuelle de manière hiérarchisée comme dans [Mann et Thompson 1988], [Hobbs 1990] [Cornish 2006] ou [Asher 1993].

Ces différentes perspectives ont pour objectif de dégager des blocs homogènes et/ou de rendre compte des éventuelles relations hiérarchiques entre ces blocs. L'originalité de notre approche par rapport à ces travaux est qu'elle vise à dépasser les notions de ruptures et de phénomènes locaux pour s'intéresser à la dynamique qui permet de passer dans les textes d'un bloc à l'autre.

### 2.2.2. Modéliser la structure discursive en reconstruisant la dynamique textuelle

Nous proposons en effet de modéliser le cheminement de lecture entre les différents référentiels d'un texte. Notre représentation de la structure discursive se base la construction dynamique de l'organisation textuelle à partir de la lecture syntagmatique. Pour représenter et visualiser la structure discursive en référentiels, nous avons proposé dans [Battistelli *et al.* 2006] de modéliser la dynamique discursive sous forme de graphe. Cette représentation considère le texte comme un ensemble de segments en relation d'inclusion ou de succession. Elle permet de visualiser le cheminement qui s'opère entre les différents référentiels au fur et à mesure de la lecture syntagmatique en insistant sur les opérations de rupture repérées par les différents indices linguistiques.<sup>13</sup>

Cette méthodologie présente (i) l'avantage de pouvoir représenter les segments discursifs homogènes d'un texte ainsi que les relations hiérarchiques qu'ils entretiennent entre eux; (ii) l'avantage d'offrir une visualisation du parcours nécessairement dynamique de lecture; (iii) l'avantage de proposer un mécanisme de fermeture systématique des référentiels ouverts; (iv) l'avantage de pouvoir s'appliquer à tout texte, quel que soit son type<sup>14</sup>.

## 3. Cadre méthodologique

Dans cette partie, nous décrivons le cadre théorique utilisé, les éléments linguistiques sur lesquels il s'appuie ainsi que les principes méthodologiques de construction des graphes.

---

<sup>13</sup> Dans [Battistelli *et al.*, 2006], nous avons proposé une modélisation des référentiels temporels. Nous reprenons et explicitons ici le même formalisme mais appliqué au traitement de la prise en charge énonciative.

<sup>14</sup> Que cette méthodologie soit applicable à tout type de textes ne signifie aucunement qu'elle est pertinente pour tout type de texte : certains textes présentent davantage de variations de prise en charge que d'autres, la méthodologie vise justement à établir une typologie en ce sens.



## 3.1 Les référentiels énonciatifs et modaux

La distinction de différents référentiels énonciatifs a été proposée par [Desclés 1980] pour rendre compte de distinctions aspectuelles entre des valeurs d'état, de processus et d'événements attribuées à des propositions. Notre analyse s'appuie ce modèle, mais s'en distingue car nous proposons une représentation alternative qui ignore les phénomènes aspectuels temporels fondamentaux pour rester à un niveau de surface et nous focaliser uniquement sur les mécanismes d'ouverture et de fermeture des référentiels énonciatifs. Nous présenterons donc la manière dont les textes articulent les différents référentiels sans considérer les intervalles de validation traditionnellement associés aux différentes propositions.

Parmi la typologie des référentiels proposée par [Desclés 1995] et précisée par [Chagnoux 2006], nous avons retenu trois types de référentiels pour rendre compte de la structure discursive<sup>15</sup> :

- les référentiels énonciatifs (noté RE) ;
- les référentiels du possible (noté RP) ;
- les référentiels mentaux (noté RM).

### 3.1.1. Les référentiels énonciatifs

Les référentiels énonciatifs se définissent comme les référentiels à l'intérieur desquels le contenu est directement assumé par un énonciateur E. Tout texte suppose l'ouverture d'un référentiel énonciatif global, noté RE<sub>G</sub>, dont les coordonnées sont celles de l'énonciateur principal. Par la suite, d'autres référentiels énonciatifs pourront s'ouvrir mais il s'agira de référentiels énonciatifs locaux, noté RE<sub>L</sub> dont les coordonnées sont liées à l'énonciateur du discours rapporté. Ainsi, l'exemple [5] articulait trois référentiels énonciatifs distincts :

- RE<sub>G</sub> : le référentiel de l'énonciateur;
- RE<sub>L1</sub> : le référentiel des Israéliens;
- RE<sub>L2</sub> : le référentiel de la police libanaise.

Par conséquent, quel que soit le mécanisme de construction, tous les discours rapportés, directs, indirects ou contextualisés convoquent un référentiel énonciatif local comme l'illustrent les exemples de [7]<sup>16</sup> où les énoncés en gras appartiennent à RE<sub>L</sub>.

[7a] Six villageois, dont trois enfants, ont été tués et six blessés dans le bombardement d'Aïtaroun, un village du Sud-Liban. A Aabba, près de Nabatiyé, à 70 km au sud-est de Beyrouth, des chasseurs bombardiers ont détruit plusieurs maisons. Selon le ministre canadien des Affaires étrangères, **il y aurait huit ressortissants canadiens parmi les victimes.**

[7b] « **Toutes les victimes sont des employés de la compagnie nationale ferroviaire** », a précisé un pompier intervenant sur les lieux du drame.

[6c] Selon le président du Parlement libanais, Nabih Berri, « **il existe une opportunité pour obtenir un cessez-le-feu** » via une tierce partie.

[7c] A défaut d'un tel cessez-le-feu dans les 24 heures, « **toute la région du Proche-Orient sera en danger** », a prévenu le président pro-syrien.

[7d] Peretz a en revanche réaffirmé que **l'armée israélienne n'avait aucune intention de se « réinstaller »** au Liban. « **Nous n'avons aucune intention de nous enliser dans le borbier libanais** », a-t-il ajouté.

---

<sup>15</sup> L'objectif de cet article est de décrire les mécanismes de construction de la structure référentielle, par conséquent, nous ne précisons pas ici les analyses détaillées sur lesquelles se fondent cette typologie. Pour le référentiel énonciatif, on pourra consulter [Guentchéva 1996], [Desclés et Guentchéva 2000] et [Chagnoux 2006] et pour le référentiel des possibles, [Vinzerich et Desclés 2007] et [Kronning 2003].

<sup>16</sup> Extraits tirés de « Israël frappe lourdement le Sud-Liban », *Le Figaro*, <http://www.lefigaro.fr>, édition en ligne du 15/10/07.

### 3.1.2. Les référentiels possibles

Les référentiels possibles se définissent comme des référentiels à l'intérieur desquels le contenu est considéré comme éventuel, noté RP<sub>E</sub>, ou contre-factuel, noté RP<sub>C</sub>. Le référentiel possible éventuel marque la création d'un nouveau référentiel où les situations sont envisagées comme possibles ou probables, comme dans les exemples [8]

[8a] Le porte-parole de la diplomatie européenne, Javier Solana, est arrivé dimanche à Beyrouth. **Il devrait s'entretenir avec le premier ministre libanais.**<sup>17</sup>

[8b] **Le 31 juillet, un patrouilleur israélien Sa'ar 4 aurait été atteint au large de Tyr par un missile inconnu du Hezbollah**, bien que l'armée israélienne ait démenti l'information.<sup>18</sup>

Le référentiel possible contrefactuel se distingue du référentiel possible éventuel par la possibilité de confronter les situations qu'il décrit avec celles qui se sont déjà réalisées dans un autre référentiel comme dans l'exemple [9]

[9] Dan Gillerman, ambassadeur d'Israël à l'ONU, devant le Conseil de Sécurité (30 juillet) : « Ces personnes, y compris les femmes et les enfants, qui ont été tuées dans ce tragique accident, peuvent avoir été tuées par les tirs israéliens, mais ils sont victimes du Hezbollah. Ils sont victimes du terrorisme. **S'il n'y avait pas de Hezbollah, ceci ne se serait jamais produit.** »<sup>19</sup>

Outre l'importance lexicale des verbes modaux (pouvoir, devoir, etc.) et de la construction syntaxique, il faut noter le rôle prépondérant des temps et des modes dans le repérage de ces référentiels. Ainsi, par exemple, parmi les possibles éventuels, il est possible de distinguer les possibles réels et les possibles potentiels. Ces distinctions sont marquées linguistiquement par l'emploi des temps verbaux. Si la proposition conditionnelle est au présent, le possible relève d'une projection de l'énonciateur. Si la proposition conditionnelle est à l'imparfait, le possible est plus strictement potentiel.

### 3.1.3. Les référentiels mentaux

Les référentiels mentaux se définissent comme des référentiels à l'intérieur desquels le contenu a un statut comparable à celui des méta-représentations décrites par [Recanati 2000] : il exprime un fait en même temps qu'une représentation de ce fait. Ainsi, ils sont le plus souvent convoqués par des propositions comme *il pense que, il croit que, il imagine que, etc.*

Une fois proposée cette typologie, nous avons défini la nature des segments de textes cohérents en terme de prise en charge énonciative. Ainsi, l'extrait [7a]<sup>20</sup> contient 3 segments qui sont identifiés en [7b] sur : RE<sub>G</sub> un référentiel énonciatif global où E correspond à l'auteur de l'article, RP<sub>E</sub> un référentiel possible éventuel et RE<sub>L</sub> un référentiel énonciatif local où E correspond au « général Amir Eshel ».

[7a] L'opération au Liban serait de longue durée. C'est ce que déclare le chef d'état-major adjoint de l'armée de l'air israélienne, le général Amir Eshel. Il prévient l'opinion internationale et intérieure que : « Il ne faut pas voir cela comme une question de jours. Cela va durer longtemps ».

[7b] RE<sub>G</sub> { RP<sub>E</sub> { L'opération au Liban serait de longue durée }. C'est ce que déclare le chef d'état-major adjoint de l'armée de l'air israélienne, le général Amir Eshel. Il prévient l'opinion internationale et intérieure que : « RE<sub>L</sub> { Il ne faut pas voir cela comme une question de jours. Cela va durer longtemps } ».

<sup>17</sup> *ibid*

<sup>18</sup> "Conflit israélo-libanais de 2006.", <http://fr.wikipedia.org>

<sup>19</sup> "L'offensive du Hezbollah – Chronologie des événements.",

<http://paris1.mfa.gov.il/mfm/web/main/missionhome.asp?MissionID=31&>

<sup>20</sup> « Israël en guerre contre le Hezbollah », <http://www.liberation.fr/actualite/monde/193193.FR.php>

## 3.2. Cohésion et incohésion : les marqueurs linguistiques

Comme nous l'avons indiqué en introduction, dans une perspective d'analyse de surface, l'ouverture et la fermeture des référentiels se traduisent linguistiquement par la présence de marqueurs linguistiques : des marqueurs strictement grammaticaux comme les temps verbaux dans les exemples [4f] et [4g] ou les groupes adverbiaux comme dans les exemples [4a] ou [4b], des marqueurs syntaxico-sémantiques comme dans l'exemple [4e] et des marqueurs typographiques comme dans [4d].

Ces marqueurs linguistiques fonctionnent sur des principes de cohésion ou de rupture. Les marqueurs de cohésion assurent l'homogénéité discursive en liant ensemble les segments au sein d'un même bloc cohérent. Ainsi dans l'exemple [8], la conjonction de coordination « et » permet d'intégrer la troisième proposition dans le même segment que la seconde – soit le référentiel énonciatif local associé au Cheikh Nasrallah et non à celui associé à l'auteur de l'article.

[8] Comme l'indiquait le Cheikh Nasrallah, dirigeant du Hezbollah, la date de l'attaque de la patrouille israélienne n'avait pas été programmée à l'avance et c'est un concours de circonstances favorables mais imprévues qui l'a provoqué

Il faut préciser que les marqueurs de cohésion sont des connecteurs, des adverbes, des temps verbaux concordant en système (dans le cas d'un système d'alternance d'imparfait et de passé simple par exemple) et qu'un même marqueur comme « et » peut être à la fois marqueur de cohésion comme dans l'exemple [8] et marqueur de rupture selon les éléments contextuels.

À l'inverse des marqueurs de cohésion, les marqueurs de rupture indiquent les transitions entre référentiels distincts. Il peut s'agir de marqueurs simples comme les indices typographiques tels que les guillemets, suivis ou non de deux points, ou d'ensembles de marqueurs comme dans l'exemple [9] qui mêle éléments lexicaux, syntaxiques et typographiques.

[9] « Nos avions, nos chars et notre artillerie opèrent à l'intérieur du territoire libanais », a déclaré un porte-parole militaire israélien, mercredi 12 juillet.<sup>21</sup>

Un autre élément important de l'analyse est que l'absence de certains marqueurs constitue un indice : il ne s'agit donc pas simplement de repérer des marqueurs mais plutôt de repérer des configurations de marqueurs permettant de conclure ou non à l'intégration de segments dans un bloc. Nous utilisons ces marqueurs de surface comme des indices de la structure profonde du texte : c'est à partir de leur seule présence ou absence qu'est reconstruite la structure sous-jacente au discours.

## 3.3. Les propriétés du graphe discursif

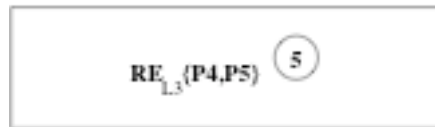
À partir de l'identification des marqueurs discursifs et des référentiels, il est possible de reconstruire dynamiquement le graphe discursif qui représente le texte.

### 3.3.1. Les nœuds du graphe

Dans ce graphe, les nœuds correspondent aux référentiels identifiés. Les nœuds sont numérotés pour préciser le chemin à l'intérieur du graphe (il n'existe qu'un seul et unique chemin) et étiquetés sous la forme  $RT_{sn}$  où  $T$  désigne le type de référentiel, (énonciatif, possible ou mental),  $s$  désigne le statut du référentiel (global ou local pour l'énonciatif, éventuel ou contre-factuel pour le possible) et  $n$  le numéro associé au référentiel. L'étiquette du référentiel est suivie des différentes propositions qui s'inscrivent dans ce référentiel. Ainsi, par exemple  $RE_{L3}\{P_4, P_5\}$  désigne dans la figure 1 le cinquième référentiel du texte qui est le troisième énonciatif local et précise qu'il contient les propositions  $P_4$  et  $P_5$ .

---

<sup>21</sup> "Conflit israélo-libanais de 2006.", <http://fr.wikipedia.org>

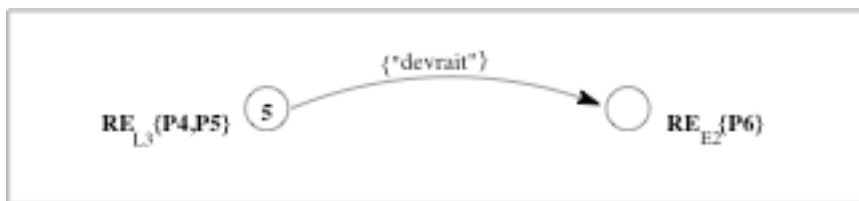


**Figure 1 : exemple de nœud du graphe.**

Le premier nœud du graphe occupe un statut particulier du point de vue théorique puisqu'il correspond toujours au référentiel énonciatif global. Même si aucun marqueur linguistique ne l'introduit explicitement, il est toujours présent et c'est à partir de ce premier nœud que se développe l'organisation textuelle<sup>22</sup>. De plus, il est également le nœud terminal puisque le chemin de lecture s'achève à son niveau. Afin de permettre une identification rapide et aisée de ce nœud primordial, il est toujours représenté en noir dans les graphes.

### 3.3.2. Les arcs du graphe

Dans le graphe, les arcs correspondent aux transitions entre référentiels et sont labellisés avec les marqueurs discursifs de rupture. Ils sont tous orientés et sont de deux types selon qu'ils portent une information de type linguistique ou non. Ils sont étiquetés par les marqueurs linguistiques indicateurs de rupture. Lorsqu'un arc ne porte pas ce type d'information, il indique un retour vers le référentiel immédiatement précédent et participe directement ou indirectement à la fermeture d'un ou plusieurs référentiels. Ainsi l'arc illustré par la figure 2 est un arc entre un le nœud  $RE_{L3}\{P_4, P_5\}$  et le nœud  $RE_{E2}\{P_6\}$ , l'arc est étiqueté avec le marqueur linguistique dont l'identification a permis de repérer la rupture, le verbe « devrait ».



**Figure 1 : exemple d'arc du graphe.**

### 3.3.3. Les règles de construction

Un ensemble de règles définit les propriétés du graphe et en gouverne la construction. Elles peuvent se résumer comme suit :

- le graphe ne propose qu'un seul et même chemin qui commence et finit toujours sur  $RE_G$ ;
- la prise en compte d'une nouvelle proposition répond à l'une des trois opérations suivantes : l'ouverture d'un nouveau référentiel, la fermeture du référentiel courant par un retour au niveau précédent ou la continuation du référentiel courant;
- quel que soit le référentiel sur lequel se situe la dernière proposition du texte, il est nécessaire de revenir à  $RE_G$  en fermant les uns après les autres tous les référentiels restés ouverts dans l'ordre dans lequel ils ont été ouverts. Ainsi, le point final d'un texte déclenchera la fermeture de tous les référentiels restés ouverts.

<sup>22</sup> Il correspond au référentiel énonciatif de [Desclés 1980]

## 4. Exemples de construction dynamique

Pour illustrer l'analyse et la méthodologie mises en œuvre, nous proposons dans cette partie de les appliquer à deux extraits textuels : un extrait court pour lequel nous détaillons les principes de construction et un article complet, non détaillé, qui illustre le résultat obtenu sur un texte long.

### 4.1. Application à un extrait court

À partir des éléments précédemment définis, le tableau 1 présente, étape par étape, la construction du graphe associé à l'extrait de texte [10a] segmenté selon le format présenté dans l'exemple [10b]

[10a] Depuis le début de la guerre, plus de 2070 roquettes sont tombées sur le nord d'Israël. Selon les estimations en Israël, la moitié des habitants de cette région, soit 800.000 personnes, l'ont fuie pour le centre et le sud du pays. Dans les villes qui ont été très gravement touchées, ce chiffre atteindrait 60 ou 70 %, par exemple à Nahariya et à Kiryat Shmona.<sup>23</sup>

[10b] P1 : Depuis le début de la guerre, plus de 2070 roquettes sont tombées sur le nord d'Israël.

P2 : Selon les estimations en Israël, la moitié des habitants de cette région, soit 800.000 personnes, l'ont fuie pour le centre et le sud du pays.

P3 : Dans les villes qui ont été très gravement touchées,

P4 : ce chiffre atteindrait 60 ou 70 %, par exemple à Nahariya et à Kiryat Shmona.

Tout énonciatif ouvre un référentiel énonciatif global quant bien même celui-ci n'est pas explicitement marqué dans le texte. On construit ici $RE_G$ , premier noeud du graphe	$RE_G$ ①
Puisqu'il n'y a pas d'indice de rupture, P1 est situé sur ce premier référentiel.	$RE_G(P1)$ ①
Des indices de rupture de citation contextualisée ("selon" suivi de "estimations") ouvre un nouveau référentiel $RE_{L1}$ sur lequel est situé P2.	$RE_G(P1)$ ① → $RE_{L1}(P2)$ ② {"Selon", "estimations"}
En l'absence d'indices de cohésion et la typographie permettent de refermer $RE_{L1}$ et de revenir sur $RE_G$ pour situer P3.	$RE_G(P1,P3)$ ① ← $RE_{L1}(P2)$ ② {"", ""}
Le mode de "atteindrait" est un indice qui permet d'ouvrir un référentiel $RP_{E1}$ sur lequel se situe P4. Le point final en tant que marqueur de fin d'énonciation ferme tous les référentiels ouverts et le graphe s'achève sur $RE_G$ .	$RE_G(P1,P3)$ ① → $RE_{L1}(P2)$ ② {"atteindrait"} ② → $RP_{E1}(P4)$ ③ {"", ""}

**Tableau 1 : construction dynamique du graphe associé à l'exemple [10a].**

### 4.2. Application à un article complet

La possibilité de matérialiser la structure de textes volumineux dans un espace restreint peut sembler anecdotique. Dans les faits, l'expérience montre que les autres systèmes de représentation – par exemple la représentation en référentiels temporels, en blocs de cadres de discours ou en blocs de la RDST – ne peuvent s'appliquer qu'à des textes courts qui excèdent

<sup>23</sup> "L'offensive du Hezbollah – Chronologie des événements.", <http://paris1.mfa.gov.il/mfm/web/main/missionhome.asp?MissionID=31&>

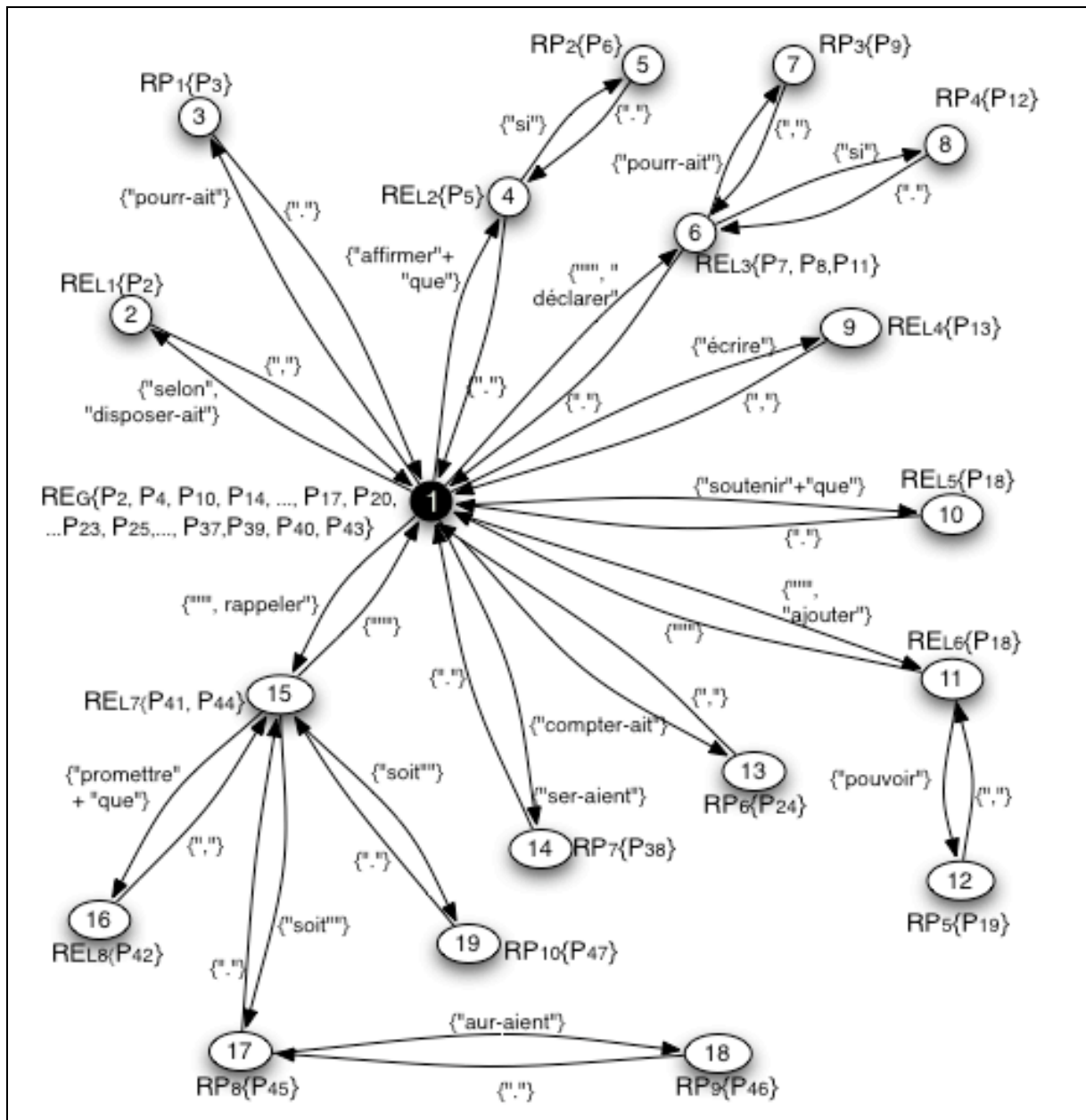
difficilement quelques lignes. Dès que le texte porte sur plusieurs pages, la représentation devient complexe et difficilement communicable.

Afin de montrer la pertinence de cette représentation, nous proposons le graphe associé à l'article complet présenté dans l'exemple [11]

[11]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. « Haïfa, que nous n'avons pas encore pris pour cible, contrairement à ce que disent les Israéliens, pourrait figurer très prochainement parmi nos objectifs, déclare le député du parti chiïte Hussein Haj Hassan. Mais, nous sommes également prêts à étendre nos opérations à une échelle beaucoup plus vaste si les attaques contre nos civils perdurent. » La milice chiite a-t-elle pour autant les moyens d'atteindre Tel-Aviv comme l'a écrit le *Jane's Defence Weekly*, hebdomadaire britannique spécialisé dans les questions de défense ? Les responsables islamistes refusent d'infirmier ou de confirmer. Mais un observateur libanais qui connaît bien le Parti de Dieu soutient qu'il dispose de l'armement nécessaire à des tirs en profondeur. « Le Hezbollah peut aussi se tourner vers un autre plan et décider de frapper à l'extérieur de l'Etat hébreu », ajoute-t-il. Pour l'heure, les miliciens arrosent le nord d'Israël de leurs roquettes. Disséminés dans le sud du Liban, ils sont lourdement armés. Leur arsenal compterait, entre autres, missiles antichars, mortiers, lance-roquettes B7, une artillerie antiaérienne ou encore des unités d'artillerie. Un équipement fourni par Téhéran et qui transite par Damas. Depuis le début de son opération, Tsahal a veillé à empêcher tout transfert d'armes vers la frontière libanaise en pilonnant les principaux points de passages vers le sud. Mais l'autre force du Hezbollah, qui a vu le jour après l'invasion israélienne de 1982, réside dans sa capacité de mobilisation de la population. L'immense majorité des Libanais chiïtes sont acquis à la cause de la résistance islamique et vouent un véritable culte à Nasrallah, élu par ses pairs en 1992. Vendredi, le leader charismatique a échappé de peu au bombardement de son QG dans la banlieue de Beyrouth et promis à la télévision. Alors que le Parti de Dieu a intégré pour la première fois le gouvernement en juillet, il consacrait jusqu'ici l'essentiel de son énergie à la lutte contre l'occupation israélienne. Pour ce faire, il dispose de dizaines de milliers d'hommes en âge de se battre. Parmi eux, 5 000 miliciens, dont quelques centaines de commandos seraient particulièrement aguerris. Ses objectifs : libérer les fermes de Chebaa, un territoire de 25 km<sup>2</sup> au sud du pays, défendre le pays contre toute agression de Tsahal et obtenir le retour de cinq Libanais détenus en Israël. Des objectifs exclusivement nationaux, donc, tout comme l'est son discours. Malgré tout, ses affinités affichées avec l'Iran, la Syrie ou le Hamas, sèment le doute. « Lors des récentes réunions de la conférence de dialogue national [au Liban, ndlr], Nasrallah a promis aux leaders politiques qu'il veillerait à ce que l'été soit calme, rappelle l'éditorialiste du quotidien An-Nahar, Rosana Boumonsef. Or, nous assistons actuellement à une offensive israélienne sans précédent depuis 1982. Alors de deux choses l'une. Soit le Hezbollah a mal calculé les répercussions qu'aurait la capture de deux soldats israéliens. Soit, il a été poussé par l'Iran à venir en aide au Hamas, actuellement considérablement affaibli à Gaza. »<sup>24</sup>

---

24 « Israël en guerre contre le Hezbollah », <http://www.liberation.fr/actualite/monde/193578.FR.php>.



**Tableau 2 : Graphe discursif associé à l'exemple [11]**

## 5. Automatisation et exploitation des résultats

### 5.1. Automatisation

Le formalisme de notre approche permet de le faire correspondre avec des bases de règles et de marqueurs pour l'analyse des référentiels et un algorithme de construction de graphe. Il est donc directement implantable, ce qui permet à notre approche d'être mise à l'épreuve de corpus. Si à l'instar de [Giguet et Lucas 2004], il existe des systèmes de repérage automatique de citations, la particularité de ce système est de réinscrire les citations dans la structure discursive au lieu de faire de l'extraction et, à ce titre, se rapproche davantage des techniques de détection des introducteurs d'univers de [Bilhaut 2006]

Nous avons présenté dans [Battistelli & Chagnoux 2007] l'implantation informatique qui permet d'associer automatiquement un graphe discursif à un texte brut. Nous la résumons ici

afin de montrer quelles sont actuellement les difficultés linguistiques qui limitent ce traitement automatique. L'implantation a été réalisée au sein du système *Chronotexte*, dont les principes ont été présentés dans [Chagnoux 2006]. Ce système utilise une base d'environ 800 marqueurs linguistiques, organisés en 122 classes hiérarchisées, et une base de 80 règles heuristiques pour annoter en XML les différents segments discursifs. Le traitement des textes peut être décomposé en trois étapes :

- une phase de segmentation qui transforme le texte en une suite de propositions ;
- une phase d'annotation qui identifie les référentiels associés à chaque proposition ;
- une phase de construction de la structure textuelle à partir de ces annotations.

Ces trois étapes seront illustrées à partir de l'exemple [12].

[12] 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.<sup>25</sup>

### 5.1.1. Segmentation du texte

La segmentation du texte est effectuée au niveau propositionnel. Nous avons supposé comme acquis jusqu'ici que l'unité de base de la structure discursive était d'ordre propositionnel. En effet, les mécanismes de prise en charge font que le niveau phrastique est inapproprié. Des segments distincts peuvent être repérés au niveau infraphrastique comme dans l'exemple [13] où deux segments coexistent dans une phrase puisque la seconde proposition assumée par « la chaîne privée de la télévision israélienne » est en rupture avec la première et la troisième assumées par l'énonciateur global. Les segments peuvent également être repérés au niveau supraphrastique comme dans l'exemple [14] où un segment cohérent inclut plusieurs phrases.

[13] La deuxième chaîne privée de la télévision israélienne avait affirmé qu'Hassan Nasrallah avait été blessé lors d'un raid, ce que le mouvement chiite a démenti.

[14] « Quand au problème du Proche-Orient, il faudra tôt ou tard le régler par la négociation, a dit en conclusion le général d'armée Gareev. Plus tôt nous le comprendrons et mieux ce sera. Pour Israël, pour ses voisins et même pour les forces qui se tiennent derrière eux. »

La segmentation automatique en propositions est réalisée par un outil externe, PROPOS [Wonsever 2004], elle permet de définir les éléments de base, constitutifs des segments cohérents. Ainsi, par exemple, PROPOS permet de segmenter l'exemple [12] en deux propositions distinctes comme le montre la sortie du système présentée en [14] et interprétée en [15].

[14] <PHRASE\_TEXTE><PROP COD=pli1 NIVEAU=1 INDEP VERBE =  
affirment>Ses cadres affirment </PROP><PROP COD=prl1 NIVEAU=2 INSER  
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>

[15] <P1>Ses cadres affirment qu'<P2> : ils ont les moyens de procéder à une  
escalade de grande envergure <P3>: si Tsahal poursuit ses raids sur le sol libanais  
</P3> </P2></P1>.

Une fois segmenté, le texte est considéré comme un ensemble de propositions qu'il faut qualifier et dont il faut définir les interactions.

### 5.1.2. Annotation des propositions

La seconde phase consiste en une analyse de surface du texte qui identifie la nature des différents référentiels présents dans le texte. Ce repérage repose sur la présence de marqueurs

<sup>25</sup> "L'armée israélienne investit le Liban sud.", [http://www.rfi.fr/actufr/articles/079/article\\_45021.asp](http://www.rfi.fr/actufr/articles/079/article_45021.asp), (publié le 12/07/06 et consulté le 15/09/06).



qui déclenchent des règles qui vont, sous certaines conditions, permettre l'annotation des propositions.

Deux bases de connaissances linguistiques conditionnent le traitement : une base de marqueurs et une base de règles. La base de marqueurs contient des motifs linguistiques organisés en classes. Un motif peut faire appel à des annotations morphosyntaxiques ou à d'autres motifs linguistiques. De plus, les motifs peuvent être simples ou combinés puisqu'il est possible de créer des motifs utilisant des connaissances attribuées par une autre tâche du système comme ici les annotations ajoutées par PROPOS. Pour construire l'ensemble des motifs linguistiques, nous avons défini des classes élémentaires que nous avons ensuite réutilisées pour former des classes d'indices plus complexes. Ainsi, par exemple, le motif *&Verbe-Citation*<sup>26</sup> renvoie à toutes les occurrences de verbes de citation conjugués comme dans les exemples [16]<sup>27</sup>.

[16a] Visiblement en bonne santé, il **a déclaré** que le bombardement de Haïfa n'était « que le début».

[16b] A défaut d'un tel cessez-le-feu dans les 24 heures, « toute la région du Proche-Orient sera en danger », **a prévenu** le président pro-syrien.

[16c] « Toutes les victimes sont des employés de la compagnie nationale ferroviaire », **a précisé** un pompier intervenant sur les lieux du drame.

La base de règles contient les heuristiques exprimées dans le langage propriétaire de *Semantext*, la plate-forme de collecte et d'organisation des connaissances linguistiques dans laquelle Chronotexte a été implémenté. Les règles sont exprimées par le triplet *<Déclencheur, Condition, Action>*. Un moteur de règles coordonne l'exécution des règles quand une occurrence pertinente de marqueur est reconnue : le système détermine alors l'ensemble des règles candidates à l'exécution puis les active une à une selon un ordre prioritaire.

L'annotation des propositions se fait en deux temps : tout d'abord seule la partie des propositions contenant des marqueurs saillants est étiquetée. Ensuite, si aucun indice n'a été repéré et que, par conséquent, aucune condition n'a été satisfaite pour attribuer une valeur à une proposition, une seconde tâche annote les propositions restantes en prenant en compte une seconde base de règles. Des règles de récursivité définissent alors les conditions de propagation des annotations d'une proposition à l'autre. Ainsi une proposition non balisée qui suit une proposition ayant pour attribut un référentiel énonciatif local sera annotée de la même étiquette si elle ne contient pas d'indice de fermeture. Les règles de propagation permettent de vérifier que chaque proposition a été annotée.

Le système fonctionne en appliquant en cascade différents modules : les connaissances générées par le module précédent sont capitalisées et utilisées par la tâche suivante. Étape par étape, toutes les propositions d'un texte sont ainsi balisées suivant le référentiel auquel elles appartiennent. L'exemple [12] donné plus haut suit ainsi le traitement suivant :

- Le système repère tout d'abord les référentiels modaux. P3 vérifie la règle suivante (\$proposition\$ starts-with /&si/) AND (\$proposition\$ contains /&present/) AND ((\$proposition justbefore \$proposition\$ contains /&present/) OR(\$proposition justbefore \$proposition\$ contains /&present/)) qui peut être glosée comme « si une proposition est introduite par un marqueur qui relève de la classe SI et si la proposition contient un verbe conjugué au présent et si la proposition précédente ou suivante contient un présent ». Par conséquent, le système attribue P3 à un référentiel possible éventuel (noté RP<sub>E</sub>).
- Le système repère dans un second temps les référentiels énonciatifs locaux. Ainsi P2 vérifie la règle suivante (\$proposition justbefore \$proposition\$)contains /&ouverture-citation/) AND (\$proposition\$ startswith / &que/) qui peut être glosée comme « si la proposition précédente contient un marqueur appartenant à la classe *&ouverturecitation*

<sup>26</sup> Le symbole & permet de déclarer des classes d'indices.

<sup>27</sup> « Israël frappe lourdement le Sud-Liban », *Le Figaro*, <http://www.lefigaro.fr>, édition en ligne du 15/10/07.

et si la proposition courante est introduite par un marqueur de la classe *&que* ». Par conséquent, P2 est identifié comme un référentiel énonciatif local (noté RE<sub>L</sub>).

- Puisque 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. Le système vérifie donc la règle (NOT (\$proposition\$ contains (/Referentiel-possible/ OR /Referentiel-mental/ OR /Referentiel-enonciatif2/ OR /Referentiel-Nactualise/)) : une fois que tous les autres types de propositions ont été reconnues, le système considère que les propositions restantes relèvent du référentiel énonciatif global (noté RE<sub>G</sub>)

Par conséquent, en sortie, l'extrait est balisé comme en [17].

```
[17]<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>
```

### 5.1.3. Construction de la structure discursive.

À partir des connaissances générées dans l'étape précédente, il est possible d'exploiter la structure sous-jacente ainsi dégagée. Cette structure est formalisée selon la DTD définie par [Couto 2006] pour la plate-forme NaviTexte [Couto et Minel 2006] et [18] montre le résultat associé à notre extrait.

```
[18]<?xml version="1.0" encoding="iso-8859-1"?>
<!DOCTYPE Texte SYSTEM "DocumentNaviTexte.dtd">
<Texte><Tete></Tete><Corps>
<UT Type="Referentiel" Nro="1">
<Attribut Nom="Nature">Enonciatif_global</Attribut>
<UT Type=" Proposition " Nro="1">
<Chaine>Ses cadres affirment qu'</Chaine></UT>
<UT Type=" Référentiel " Nro="2">
<Attribut Nom="Nature">Enonciatif_local</Attribut>
<UT Type=" Proposition " Nro="2">
<Chaine>ils ont les moyens de procéder à une escalade de grande
envergure</Chaine></UT>
<UT Type=" Référentiel " Nro="3">
<Attribut Nom="Nature">Enonciatif_possible_eventuel</Attribut>
<UT Type=" Proposition " Nro="3">
<Chaine>si Tsahal poursuit ses raids sur le sol
libanais.</Chaine></UT></UT></UT></UT>
</Corps></Texte>
```

La partie « Corps » contient les différentes propositions du texte dans leur ordre syntagmatique. À chaque proposition sont associées les balises qui définissent les propriétés référentielles : chaque segment est défini grâce à une *Unité Textuelle* (UT) à laquelle est associé un attribut *Nature* qui type le référentiel. Il est également possible d'ajouter des attributs sachant que chaque type de référentiel va convoquer des attributs différents afin d'encoder les propriétés qui lui sont associées. Ainsi, par exemple, le référentiel énonciatif possède un attribut *Enonciateur* qui peut être vide ou qui peut contenir le nom de l'énonciateur si celui-ci est renseigné.

La partie « Tête » construit la structure par des opérations de séquençage : chaque opération de *Séquence* rétablit la cohésion entre des segments discontinus dans le texte. Une fois que le texte balisé est encodé dans cette DTD, il est possible d'exploiter les résultats par des feuilles XSLT, par des outils de visualisation comme Navitexte ou par la construction automatique de graphes comme nous l'avons proposé dans cet article.

## 5.2 Exploitation des résultats

La représentation à l'aide d'un graphe présente plusieurs avantages: (i) un accès immédiat à la complexité structurelle d'un texte; (ii) l'économie d'un méta-langage ; (iii) la possibilité de matérialiser la structure de textes volumineux dans un espace restreint.

L'accès immédiat à la complexité structurelle d'un texte est l'élément le plus pertinent pour notre étude : il permet de caractériser un texte suivant la complexité de son graphe. Exploiter la profondeur du graphe ouvre la perspective d'une typologie des articles basés sur des critères énonciatifs et modaux. Il est ainsi possible de fonder intrinsèquement une distinction entre d'une part les articles produits par les autorités israéliennes et palestiniennes et d'autre part les articles de la presse française. Les premiers se caractérisent par une structure simple avec peu de référentiels distincts dont la profondeur n'excède jamais 2 alors que les seconds présentent des emboîtements complexes de référentiels.

L'économie du méta-langage est rendue possible par la représentation iconique : il suffit de savoir qu'à chaque noeud correspond un type de prise en charge énonciative (un énonciateur ou une modalité) et de distinguer le premier noeud pour savoir si un texte est fortement citationnel ou fortement modalisé.

La possibilité de matérialiser la structure dans un espace restreint permet un affichage graphique qui modifie le rapport au texte puisqu'il est possible de visualiser directement la complexité du texte pour faire émerger d'autres caractéristiques de la structure discursive.

### 5.3 Limites de l'analyse

Utiliser *Chronotexte* pour annoter la structure discursive de ce corpus nous a permis d'évaluer la pertinence de notre approche méthodologique comme la qualité du système implémenté. Par conséquent, les difficultés rencontrées sont de deux ordres : les premières relèvent des choix théoriques, les secondes d'une couverture insuffisante des phénomènes par *Chronotexte*.

#### 5.3.1 Limites du cadre méthodologique

En choisissant comme unité sémantique de base des propositions construites autour de verbes conjugués, une partie des informations pertinentes échappe parfois à l'analyse. Tout d'abord, il est impossible de rendre compte des phrases nominales comme dans l'exemple [19] où « une information confirmée peu après par une source aéronautique libanaise » ne peut être traduite dans le graphe.

[19]La télévision du Hezbollah rapporte que l'aviation israélienne a bombardé jeudi matin l'aéroport international de Beyrouth. Une information confirmée peu après par une source aéronautique libanaise.<sup>28</sup>

Ensuite, l'absence de marqueurs explicites au sein du discours indirect entraîne souvent une ambiguïté sur la personne qui prend en charge certains termes. Ainsi, dans l'exemple [20], le qualificatif de « sévère » est attribué à l'énonciateur global alors qu'il est plus probable qu'il participe à la citation des propos de l'énonciateur local, Israël.

[20]Israël a aussitôt annoncé une riposte sévère.<sup>29</sup>

Enfin, comme l'illustre l'exemple [21a], considérer la proposition comme unité de base fait que des constructions syntaxiques internes à la proposition ne sont pas prises en compte et entraîne une perte d'information. La modification de l'exemple [21a] en [21b] montre que pour un contenu sémantique presque équivalent, seule une analyse de [21b] identifiera les principes de citations.

<sup>28</sup> « Israël en guerre contre le Hezbollah », <http://www.liberation.fr/actualite/monde/193193.FR.php>.

<sup>29</sup> « Israël frappe lourdement le Sud-Liban », *Le Figaro*, <http://www.lefigaro.fr>, édition en ligne du 15/10/07.

[21a]Le bilan des blessés est confirmé par le Maguen David Adom, l'équivalent israélien de la Croix-Rouge.<sup>30</sup>

[b] Maguen David Adom, l'équivalent israélien de la Croix-Rouge, a confirmé que les blessés étaient au nombre de 11.

### 5.3.2 Limites du système de traitement automatique

Il est également difficile de repérer automatiquement les reformulations diverses d'une même information. Dans l'exemple [22], d'une part le système ne permet pas d'articuler la première et la seconde proposition - le fait que c'est un énonciateur local, le général Amir Eshel qui assume la probable longue durée de l'opération - et d'autre part il n'est pas possible d'établir une correspondance discursive entre les propos cités directement par le général -. « Cela va durer longtemps » - et les mêmes propos rapportés par l'auteur de l'article - « L'opération au Liban serait de longue durée ».

[22]L'opération au Liban serait de longue durée. C'est ce que déclare le chef d'état-major adjoint de l'armée de l'air israélienne, le général Amir Eshel. Il prévient l'opinion internationale et intérieure que : «Il ne faut pas voir cela comme une question de jours. Cela va durer longtemps».<sup>31</sup>

Enfin, le système de règles reste encore à compléter pour pouvoir traiter certains problèmes complexes que ce corpus a fait émerger. À titre d'exemple, nous pouvons citer le cas de la citation partielle comme dans l'exemple [23] : la citation alterne des marques de discours direct comme les guillemets et des marques de citation indirecte. Le phénomène de citation partielle est très présent dans le corpus étudié et devra être modélisé afin que *Chronotexte* soit apte à le gérer.

[23]Et le Hezbollah de prévenir : "En cas de riposte israélienne", le mouvement n'épargnera "aucun objectif dans la ville de Haïfa et ses environs".

Un autre problème soulevé par le corpus est le cas de l'emboîtement de citations comme dans l'exemple [24]. Le système n'identifie que deux référentiels énonciatifs : le référentiel énonciatif global de l'énonciateur et le référentiel énonciatif local de la radio militaire israélienne. Or, en réalité, trois référentiels devraient être distingués puisqu'à partir du référentiel énonciatif local s'ouvre un second référentiel énonciatif local, celui des responsables de Tsahal. C'est d'ailleurs sur ce référentiel qu'est située la proposition selon laquelle un blocus général est imposé sur le Liban.

[24]La radio militaire israélienne, citant des responsables de Tsahal, annonce qu'un blocus général (aérien, maritime et terrestre) est imposé sur le Liban.

Du point de vue de l'automatisation, le repérage des citations contextualisées apparaît très délicat. Si dans des exemples comme [25], les indices lexicaux (« agence de presse ») et typographiques (présence de guillemets) permettent un repérage de citation, les indices sont souvent difficiles à identifier et occasionnent parfois des erreurs d'analyse. Ainsi, dans l'exemple [26], la polysémie de « conception » rend plausible une citation par une autorité iranienne alors que le texte n'en contient pas.

[25] Selon l'agence de presse gouvernementale égyptienne Mena, les deux hommes ont discuté des «moyens de calmer l'escalade et d'empêcher la situation dans les territoires palestiniens et au Liban de se détériorer»<sup>32</sup>.

[26]Le Hezbollah utilise différents types de missiles, parmi lesquels des roquettes Katioushas et Fajr 2, ces derniers étant fabriqués en Syrie selon une conception iranienne.

La multiplication de ces phénomènes fait qu'actuellement la tâche de repérage des citations contextualisées est trop bruitée pour être pertinente. Aussi, le repérage automatique des

---

<sup>30</sup> « Israël en guerre contre le Hezbollah », <http://www.liberation.fr/actualite/monde/193193.FR.php>.

<sup>31</sup> *Ibid.*

<sup>32</sup> "L'armée israélienne investit le Liban sud.", Arsenault, C., <http://www.rfi.fr>

citations contextualisées sans marque explicite de citation comme les guillemets a été exclu de *Chronotexte* en attendant de pouvoir intégrer des règles de repérage plus fines.

## 6. Conclusion et perspectives

Notre objectif était de montrer que des propriétés de nature typologique peuvent émerger du texte lui-même par le biais d'une représentation formelle de sa structure. Pour l'atteindre, nous avons proposé un cadre méthodologique et un outil qui construisent une représentation de la dynamique discursive sous forme de graphe. Cette représentation rend compte des degrés et mécanismes de prises en charge énonciative et peut dès lors renseigner sur le caractère polémique de certains textes.

L'originalité de notre approche réside d'une part dans la dynamique de construction de la représentation discursive et d'autre part dans l'implantation qui permet de confronter notre modèle à des corpus. Elle propose également de considérer les phénomènes de prise en charge énonciative comme des éléments structurants du discours. Cependant, nous avons montré qu'un certain nombre des difficultés persiste dans la reconnaissance automatique. Par ailleurs, notre système est pour l'instant incapable de résoudre le difficile problème des anaphores pronominales et nominales. Afin que le graphe proposé présente autant de noeuds que de prises en charge repérées, il serait également nécessaire de regrouper les segments homogènes pris en charge par un même énonciateur.

Si la génération automatique du graphe à partir des fichiers balisés reste encore à mener à terme, elle offre de nouvelles perspectives de recherche dans le nouveau rapport au texte qu'elle suppose. En soustrayant le texte à sa traditionnelle linéarité, elle permet d'imaginer des parcours de navigation qui restent à exploiter. L'une des perspectives que nous souhaiterions également poursuivre pour évaluer la pertinence de notre approche est celle qui permettrait de corroborer ces résultats avec d'autres expressions du caractère polémique des mêmes sujets. En effet, un article peut être sujet à caution de par les réactions qu'il suscite auprès des lecteurs. Un premier exemple est celui de la *Wikipedia*, encyclopédie collaborative où chaque article est rédigé par les utilisateurs. La possibilité de suivre les débats entre les différents rédacteurs *via* l'historique des contributions et des discussions, souvent passionnées, des participants offre alors un outil privilégié pour repérer les articles les plus polémiques : plus les débats sont nombreux et moins le sujet est consensuel. Dans le cas où les désaccords sont trop importants, l'article peut même être soumis à modération. Un second exemple est offert par la possibilité que proposent maintenant les blogs ou certains sites de presse de réagir et commenter chaque article en ligne. Ici encore, certains articles suscitent nombre de réactions de lecteurs alors que d'autres n'entraînent que peu de commentaires. L'examen quantitatif et qualitatif des réactions des lecteurs permettrait de distinguer les articles plus polémiques que d'autres et de comparer l'approche linguistique défendue ici aux approches statistiques comme celles adoptées par [Viegas *et al* 2004] et [Nigam and Hurst 2004].

### Références bibliographiques

- Asher, N. 1993. *Reference to Abstract objects in Discourse*. Kluwer Academic Publishers.
- Authier, J. 1981. *Paroles tenues à distances. Matérialités discursives*. Lille, Presses Universitaires de Lille.
- Battistelli, D., Chagnoux and M., Desclés J.-P. 2007. «Référentiels et ordonnancements temporels dans les textes», *Information temporelle, procédures et ordre discursif, Cahiers Chronos* 18, Amsterdam/New York.

Battistelli, D. and Chagnoux, M. 2007. «Représenter la dynamique énonciative et modale de textes », *TALN 2007*, Toulouse : 5-8 juin 2007.

Bilhaut, F. 2006. « Introduceurs intra-prédicatifs d'univers de discours et leur détection automatique ». *Schedae*, prépublication n°6, (fascicule 1) 41-50.

Chagnoux, M. 2006. *Temporalité et aspectualité dans les textes français. Modélisation sémantico-cognitive et traitement informatique*. Paris : Thèse de l'Université Paris-Sorbonne.

Charolles, M. 1997. « L'encadrement du discours : univers, champs, domaines et espaces », *Cahiers de Recherche Linguistique*, 6 : 1-73. Nancy : LANDISCO.

Charolles, M., Le Draoulec, A., Pery-Woodley M.-P. and Sarda, L. 2005, "Temporal and spatial dimensions of discourse organisation", *Journal of French Language Studies* : 115-130.

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

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

Culioli, A. 1980. « Valeurs aspectuelles et opérations énonciatives : l'aoristique », in *Notions d'aspect*, J. Davis et R. Martin (eds), 182-193. Paris: Klincksieck.

Desclés, J.-P. 1980. « Construction formelle de la catégorie grammaticale du temps et de l'aspect », in *Notions d'aspect*, J. Davis et R. Martin (eds), 198-237. Paris: Klincksieck.

Desclés, J.-P. 1995. *Les référentiels temporels pour le temps linguistique*. Modèles linguistiques 16: 9-36.

Desclés, J.-P., and Guentchéva, Z. 2000. « Enonciateur, locuteur, médiateur. » Erikson Ph et Monod-Becquelin A (éds). *Les rituels du dialogue* : 79-112. Nanterre : Société d'ethnologie. Paris X.

Fauconnier, G. 1985. *Mental spaces: Aspects of meaning construction in natural language*. Cambridge : MIT Press.

Espuny, J. 1998. « De l'énonciation singulière à l'énonciation plurielle du locuteur », *Estudios de lengua y literatura francesas*, 12 : 53-70 Cadix.

Gosselin, L. 2005. *Temporalité et modalité*. Bruxelles, De Boeck Duculot

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

Guiguet, 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.

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*. Tübingen: Max Niemeyer, *Linguistische Arbeiten* 469: 131-151.
- Leeman-Bouix, D. 1994 *Grammaire du verbe français*. Fernand Nathan.
- Le Querler, N. 1996. *Typologie des modalités*, Caen, Presses Universitaires de Caen.
- Mangueneau, D. 1986. *Éléments de linguistique pour le texte littéraire*. Bordas
- Martin, R. 2005. « Définir la modalité », *Revue de linguistique romane* 273: 7-18. Strasbourg.
- Mann, W.C., Thompson, S.A. 1988. "Rhetorical Structure Theory: toward a functional theory of text organization", *Text* 8(3), 243-281.
- Nigam, K. and Hurst, M. 2004. "Towards a robust metric of opinion. In *Proc. AAAI Symp. Exploring Attitude and Affect in Text*.
- Péry-Woodley, M.-P. 2001. « Présentation du numéro : Cohérence et relations de discours à l'écrit », *Verbum*, XXIII, 1.
- Porhiel, S. 2001, "Linguistic expressions as a tool to extract thematic information", *Corpus Linguistic*, Lancaster : 477-482.
- Rabatel, A. 1998. *La Construction textuelle du point de vue*. Lausanne, Paris : Delachaux et Niestlé.
- Rabatel, A. 2006. « L'effacement de la figure de l'auteur dans la construction événementielle d'un « journal » de campagne électorale et la question de la responsabilité, en l'absence de récit primaire *Semen*, 22.
- Recanati, F. 2000. *Oratio Obliqua, Oratio Recta : An Essay on Metarepresentation*. Cambridge : MIT Press.
- Rosier, L. 1993. « De la stylistique sociologique suivie d'une application pratique : discours direct, presse et objectivité », *Revue belge de philosophie et d'histoire* 71. 625-644.
- Rosier, L. 1999, *Le discours rapporté. Histoire, théories, pratiques*, Bruxelles, Paris : Duculot.
- Roulet, E. 1999, *La description de l'organisation du discours*, Paris, Didier.
- Sandor, A., Kaplan, A. and Rondeau, G. 2006. "Discourse and citation analysis with concept matching". *Schedae*, prépublication n°19, (fascicule 1) 147-152.
- Sauri, R., Verhagen, M. and Pustejovsky, J. 2005. Annotating and recognizing Event Modality in Text. Actes *FLAIRS'06*, Melbourne Beach, Florida.

Schrepfer-Andre, G. 2006. *La portée phrastique et textuelle des expressions introductrices de cadres énonciatifs: Les syntagmes prépositionnels en selon X*. Paris : Thèse de l'Université Paris III Sorbonne Nouvelle.

Terran, E. 2002. *Le cadrage temporel en français*. Thèse de doctorat, Université Paris III - Sorbonne Nouvelle.

Tuomarla, U. 2000. *La citation mode d'emploi. Sur le fonctionnement discursif du discours rapporté direct*. Helsinki : Academia scientiarum Fennica.

Viegas, F. B., Wattenberg, M. and Dave, K. 2004. "Studying cooperation and conflict between authors with history flow visualizations". In *Proc. SIGCHI Conf. Human factors in computing systems* : 575–582.

Vinzerich, A. and Desclés J.-P. 2006. « Référentiels des possibles : Représentation des situations potentielles et irréelles. ». Actes *Chronos 7*. Anvers, 18-20 septembre 2006.

Wonsever, D. 2004. *Repérage automatique des propositions par exploration contextuelle*. Paris : Thèse de l'Université Paris-Sorbonne.