

Quelle(s) relation(s) de discours pour les structures énumératives ?

Myriam Bras, Laurent Prévot & Marianne Vergez-Couret

Université de Toulouse, CLLE-ERSS UMR 5263 CNRS & Université Toulouse Le Mirail
{bras, prevot, vergez}@univ-tlse2.fr

1 Introduction

Un certain nombre de phénomènes linguistiques ne peuvent être résolus qu'en prenant en considération le niveau discursif. On peut citer par exemple le problème de l'anaphore pronominale interphrastique, étudié dans un cadre discursif formel dès (Hobbs, 1979), les relations temporelles interphrastiques (Lascarides et Asher, 1993) et les questions de parallélisme autorisant certaines formes d'ellipses (Asher et al, 2001), (Kehler, 2002). Cependant si l'importance du niveau discursif est largement reconnue, la nature exacte de sa contribution est beaucoup plus discutée. Pour analyser le discours, et donc étudier sa structure et la manière dont ses parties s'intègrent pour former un tout cohérent, les théories du discours ont été amenées à définir des segments de discours et des relations de cohérence entre ces segments. Elles ont en outre mis l'accent sur la nécessité d'identifier des marques pour repérer les relations du discours. Cette tâche identificatoire permet d'avancer dans deux directions complémentaires. Elle permet d'une part de valider les structures de discours à l'aide de formes de surfaces leur correspondant. Elle permet d'autre part de découvrir de nouvelles structures et d'affiner les connaissances des structures existantes sur la base d'études systématiques des marqueurs discursifs. Enfin, le repérage systématique de ces relations ouvre des perspectives applicatives (résumé automatique, extraction d'information,...) et requiert un inventaire aussi riche que possible des marqueurs disponibles pour le repérage automatique. Les indices de structure du discours évoqués plus haut incluent les marqueurs lexicaux, certaines structures syntaxiques, des marqueurs aspectuo-temporels, des informations de sémantique lexicale ou même des informations de nature plus encyclopédique. Cependant, la sémantique du discours et en particulier la SDRT (Théorie des Représentations Discursives Structurées) (Asher, 1993) dans laquelle nous situons nos analyses, utilise avant tout la sémantique des énoncés individuels pour atteindre compositionnellement le sens du discours et néglige les aspects textuels, qu'ils soient liés à l'organisation du texte ou à sa mise en forme matérielle. Un des objectifs de cet article est de prendre en compte cette source d'information particulièrement pertinente pour contribuer à l'établissement de la cohérence d'un texte.

Dans l'étude de la structure du discours, l'identification de relations introduisant une dimension hiérarchique s'avère cruciale, tant pour l'interprétation – car la structure hiérarchique participe à la définition de contraintes pour les développements discursifs et l'accessibilité anaphorique – que pour des applications comme le résumé automatique pour lesquelles les éléments dominés hiérarchiquement devraient pouvoir être laissés de côté à l'inverse des éléments dominants qui seraient essentiels au propos. L'étude de ces relations introduisant une hiérarchie entre les éléments est étroitement liée à la relation de discours d'Elaboration. Cette relation est difficile à caractériser (si l'on en juge par le large spectre de propositions pour sa définition dans différentes théories), et également difficile à repérer. Knott (1996) avance par exemple que c'est une relation qui n'est pas repérable à l'aide de marqueurs lexicaux en anglais.

Les structures énumératives, que nous nous proposons d'étudier dans cet article, sont des structures hiérarchiques au plan textuel et sont fréquemment associées à des marques de surface typo-dispositionnelles. Notre étude vise à explorer la contribution d'une source d'information quelque peu négligée (même si nous n'étudierons pas directement les marques typo-dispositionnelles) à l'établissement de la cohérence discursive, et plus particulièrement la contribution de cette source d'information au repérage de la relation d'Elaboration, une tâche notoirement difficile. Plus précisément, nous voulons analyser les structures énumératives telles que décrites par Luc (2000), Péry-Woodley (2000), *inter alia*, ce qui nous amènera à regarder une classe de marqueurs particuliers appelés marqueurs d'intégration

linéaire (MIL) mis au jour par Turco et Coltier (1988), Adam et Revaz (1989). Après avoir introduit les théories du discours (section 2), nous présenterons plus en détails le cadre théorique de référence que nous utilisons (section 3). Ensuite, dans la section 4, nous présenterons l'analyse d'un MIL, *d'abord*, qui nous a servi de point de départ, avant de nous focaliser sur les structures énumératives (section 5) et de proposer une analyse de ces dernières à l'aide de notre cadre théorique dans la section 6.

2 Les théories du discours

L'étude de la macro-structure linguistique du discours et des relations qui lient ses composants remonte aux travaux fondateurs de Halliday et Hasan (1976), prolongés, entre autres, par des travaux comme ceux de Mann et Thompson (1988) ou de Hovy et Maier (1992). Dans ces cadres théoriques, un discours est un modèle structurel établi sur la base de la compréhension des analystes. Les sources d'informations pour établir cette structure sont en conséquence très riches (marqueurs de discours, cohésion lexicale, ...) mais peu explicitées. Par exemple, des considérations sémantiques sophistiquées peuvent être mises en œuvre mais sans que soit systématiquement proposé un modèle du fonctionnement discursif. La RST (*Rhetorical Structure Theory*) de Mann et Thompson (1988) a initié un mouvement vers la spécification de ces relations et en a établi une liste ouverte. En effet, cette théorie a été initialement conçue pour aider des analystes humains à représenter la structure des textes et ne propose pas de définition formelle pour les relations. En revanche, Marcu (1997) a proposé un système complet d'analyse du discours (*Discourse Parsing*) dans ce cadre théorique, mais en se restreignant aux indices de surface. D'autres approches computationnelles ont également utilisé ces indices de surface pour proposer des systèmes comme les grammaires de discours de (Polanyi et Scha, 1984).

L'utilisation des relations rhétoriques de l'analyse discursive dans des cadres plus formels a été initiée par Hobbs (1979). Par la suite de nombreux travaux (Hobbs, 1985, Grosz et Sidner, 1986, Webber, 1991, Polanyi, 1988, Asher, 1993) ont suivi cette idée mais ont grandement divergé quant à (i) la nature exacte des segments discursifs, (ii) celle des relations, (iii) la structure hiérarchique que ces relations engendrent et (iv) leur reconnaissance.

Concernant d'abord les segments discursifs, ils se répartissent entre segments minimaux, correspondant à une phrase ou à un énoncé simple, et qui ont par conséquent un contenu sémantique clair, et segments complexes correspondant à des groupes de segments minimaux, et qui forment en quelque sorte l'emballage organisateur des segments minimaux. Les segments complexes sont souvent représentés par des cadres regroupant un ensemble de constituants comme l'illustreront les figures des sections suivantes.

Concernant ensuite les relations de discours, les auteurs s'accordent sur la difficulté d'établir une liste finie de relations de discours, mais divergent sur la manière de contourner cette difficulté. Grosz et Sidner (1986) préfèrent mettre de côté l'aspect sémantique et ne considèrent que l'aspect structurel des relations. Ils ne définissent que deux relations qui ont des propriétés structurelles différentes dont la distinction est fondée sur des primitives intentionnelles. Plus précisément, un but sous-jacent (*Discourse Segment Purpose, DSP*) est associé à chaque segment et les relations sont définies à l'aide de ces buts (une relation requiert que le DSP du premier segment soit résolu avant de passer au second, l'autre que le DSP du second constituant ne soit qu'une étape dans la résolution du premier). C'est la raison pour laquelle on parle de théorie intentionnelle, car elle place les intentions des locuteurs au cœur du modèle.

Les autres approches estiment (i) soit qu'à l'aide de définitions précises on peut sélectionner un sous-ensemble de relations principales qui suffisent pour interpréter les discours (Hobbs 1985, Asher 1993) ; (ii) soit qu'une liste ouverte ne pose pas de problème particulier (Mann et Thompson, 1988). Dans les approches opérationnelles qui nous intéressent, les relations sont caractérisées d'une part par leurs règles de déclenchement et d'autre part par leurs effets sur le contenu du discours. En SDRT, en particulier, ces éléments constituent la définition des relations.

Les structures hiérarchiques définies par les théories peuvent varier quelque peu mais certains principes se retrouvent dans la plupart des propositions. Par exemple, à un point donné d'un discours, la contribution d'un énoncé peut être de deux types : continuer ce qui est en cours dans l'énoncé précédent

ou le préciser. De ces différents types de contributions naît une 'structure hiérarchique' formée par la distinction des relations 'coordonnantes' et 'subordonnantes'. Cette hiérarchie peut être représentée dans des schémas arborescents où les relations subordonnantes sont des traits verticaux et les relations coordonnantes des traits horizontaux comme illustré dans les figures des sections suivantes. Webber n'utilise que l'emboîtement des segments, sans distinguer ces types de relations. Enfin Mann et Thompson n'utilisent eux, qu'une distinction entre des segments *nucléaires* et *satellites* qui sont en quelque sorte accessoires. La définition des relations de la RST inclut en conséquence la nature des segments qu'elles relient.

Dans le cas de la SDRT, de la LDM (*Linguistic Discourse Model*) de Polanyi, et de la théorie de Hobbs, ces groupes de segments coordonnés doivent être subordonnés à un segment commun, le 'topique discursif'. Ce topique est un constituant simple qui résume le contenu du groupe de segments coordonnés. Pour la SDRT, si ce topique n'a pas été introduit explicitement dans le discours, il doit être construit à partir du contenu des constituants qu'il domine. Dans ce cas, le contenu de ce topique implicite doit être mis à jour quand le constituant complexe qu'il domine est étendu par un nouveau constituant (Asher, 2004).

3 Cadre théorique de l'étude : la SDRT

La SDRT (Asher, 1993, Lascarides et Asher, 1993, Asher et Lascarides, 2003) est une théorie formelle du discours qui s'est développée dans le prolongement de la DRT –Discourse Representation Theory– (Kamp et Reyle, 1993) pour prendre en compte la notion de structure du discours via les relations de discours (Narration, Elaboration, Résultat, Explication, Contraste, etc...). Comme nous l'avons déjà évoqué plus haut, la SDRT cherche à rendre compte de la façon dont différents type d'informations – information sémantique lexicale et grammaticale, connaissance du monde, principes pragmatiques – interagissent pour établir des relations de discours.

Les représentations du discours, les SDRSs (pour Segmented Discourse Representation Structures), sont des structures récursives constituées de constituants, i.e. de DRSs (les Discourse Representation Structures de la DRT) étiquetées représentant le contenu propositionnel des segments et décrivant des états ou des événements, et de sous-DRSs étiquetées reliées entre elles par des relations de discours. On utilise des lettres grecques (α, β, \dots) pour représenter les étiquettes, qui sont des marqueurs d'actes de langage ; des notations de la forme K_α pour représenter les constituants étiquetés avec α ; des notations de la forme e_α, e_β, \dots pour représenter l'événement ou l'état décrit dans le constituant. La construction d'une SDRS, à partir d'une SDRS représentant le contexte discursif et du nouveau constituant à attacher au contexte, s'appuie sur une logique non monotone chargée d'inférer les liens de cohérence entre le nouveau constituant et le contexte discursif, et sur une fonction de mise à jour chargée de l'intégration de ce constituant dans la structure du discours proprement dite.

La SDRT distingue comme on l'a dit les relations coordonnantes comme Narration ou Continuation, des relations subordonnantes comme Explication ou Elaboration. Nous décrivons ci-dessous les relations de Narration, Elaboration et Explication.

3.1 Narration

La relation de Narration en SDRT est fondée sur la maxime de manière de Grice (1975) concernant l'organisation des contributions linguistiques : deux segments reliés par Narration décrivent, dans l'ordre d'occurrence, deux événements d'une même histoire comme en (1) et (2) :

- (1) Pierre trébucha. Il tomba.
- (2) Pierre tomba. Puis Paul le poussa.

On peut inférer une relation de Narration entre deux constituants en s'appuyant sur la présence d'un connecteur comme *puis* (Bras et al., 2001), comme en (2), on a dans ce cas une inférence monotone (\rightarrow).

On peut aussi inférer Narration en s'appuyant sur le prédicat *occasion* appliqué aux types des éventualités décrites par les segments : ce prédicat exploite des informations basées sur la connaissance lexicale et la connaissance du monde, il est vérifié dans le cas où les deux segments contiennent des indices indiquant que les éventualités qu'ils décrivent peuvent appartenir à une même histoire, ce qui est le cas par exemple pour des types d'événements comme x *trébucher* / x *tomber*, nous amenant donc à l'inférence de Narration pour (1). On a dans ce cas une inférence non monotone ($>$). Si aucune autre relation de discours ne peut être inférée, Narration peut être inférée de façon non monotone à partir de ce qu'on sait des relations de discours liant les constituants suivants (Asher et Lascarides, 2003). Ainsi, on infère Narration entre les deux premiers segments de (3) grâce à la présence de Narration entre les deux derniers segments :

(3) Pierre tomba. Paul le poussa. Puis il tomba à son tour.

Ces trois cas de figure sont formalisés par les axiomes suivants (où λ représente le constituant minimal qui contiendra la formule reliant α à β quand la SDRS du discours complet sera construite, et R représente une relation de discours quelconque de la SDRT) :

Inférer_Narration $(? (\alpha, \beta, \lambda) \wedge \text{Occasion}(\alpha, \beta)) > \text{Narration}(\alpha, \beta, \lambda)$

Inférer_Narration_Puis $(? (\alpha, \beta, \lambda) \wedge \text{puis}(\beta)) \rightarrow \text{Narration}(\alpha, \beta, \lambda)$

Relations_Environnantes $(? (\alpha, \beta, \lambda) \wedge R(\beta, \gamma, \lambda')) > R(\alpha, \beta, \lambda)$

Narration pose des contraintes temporelles sur le contenu du discours. Les éventualités de deux segments reliés par Narration sont temporellement ordonnées :

Contraintes spatio-temporelles $\text{Narration}(\alpha, \beta) \rightarrow e_\alpha < e_\beta$

Narration impose aussi des contraintes structurelles, en particulier en termes de continuité thématique : deux événements reliés par Narration doivent avoir le même « topique de discours ». Par exemple pour (1) le topique serait « la chute de Pierre ». Les topiques sont des constituants simples qui peuvent être explicitement donnés par le discours, ou non, comme en (1). Dans ce cas, on ajoute un constituant, dit implicite, au moment de la mise à jour de la SDRS par la contrainte suivante (Asher et Lascarides, 2003, p.219) où \Downarrow représente la relation de Topique qui est subordonnante et l'opération $\alpha \sqsupset \beta$ dénote une proposition généralisant α et β :

Contrainte de topique $\text{Narration}(\alpha, \beta, \lambda) \rightarrow \exists \delta ((\delta = \alpha \sqsupset \beta) \wedge \exists \gamma (\Downarrow(\delta, \lambda, \gamma)))$

3.2 Elaboration

Elaboration est une relation subordonnante, elle introduit un constituant complexe, c'est-à-dire une SDRS qui contient d'autres SDRSs, comme illustré en Figure 1 qui représente la SDRS de (4) :

(4) (a) Mixel a escaladé le Vignemale hier matin. (b) Il est parti du refuge des Oulettes au lever du jour. (c) Puis il a passé la Hourquette d'Ossoue vers 9h, (d) et il est arrivé au sommet vers midi.

Le discours (4) décrit un événement en (a), et propose ensuite une élaboration de cette description en (b)-(d). En SDRT, on dit que l'étiquette du premier constituant π_a , représentant la phrase (a), est élaborée par les étiquettes des constituants π_b , π_c , π_d représentant respectivement les phrases (b), (c) and (d). π_a est l'étiquette du constituant complexe enchâssant les constituants élaborant π_a , ce dernier est un topique de discours explicite.

Elaboration peut être inférée de façon non monotone grâce à la présence d'une relation de subsumption entre les types des éventualités des segments à relier, ce qui est exprimé par le prédicat *Subtype_D* indiquant que le type de la seconde éventualité est un sous-type de celui de la première dans la sémantique lexicale des éventualités ou grâce à des connaissances du monde restreintes au contexte

discursif. L'axiome suivant restreint la portée de *Subtype_D* à σ , lui-même défini (à l'aide de *Top*) comme étant le sommet de la structure discursive dans laquelle apparaît α .

Inférer_Elaboration $(?(\alpha, \beta, \lambda) \wedge \text{Top}(\sigma, \alpha) \wedge \text{Subtype}_D(\sigma, \beta, \alpha)) > \text{Elaboration}(\alpha, \beta, \lambda)$

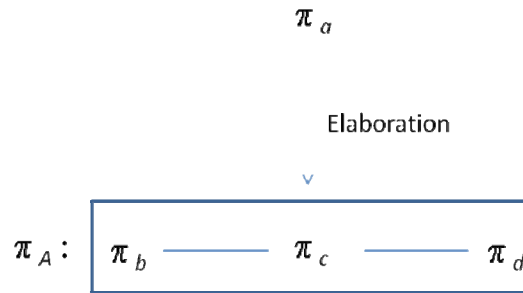


Figure 1: SDRS de (4)

Ainsi, pour (4), des connaissances du monde relatives au pyrénéisme et à la sémantique lexicale des verbes *escalader* et *partir* nous permettent d'inférer *Subtype_D*, puis, grâce à l'axiome ci-dessus d'inférer *Elaboration*(π_a, π_b). π_c est ensuite attaché à π_b par *Narration*, grâce à l'axiome *Inférer Narration_Puis*. On a donc *Narration*(π_b, π_c). On peut alors appliquer un principe général de la SDRT, stipulant que si on attache un constituant à un autre constituant qui fait partie d'une *Elaboration*, ce qu'on attache doit aussi faire partie de l'*Elaboration* (Asher et Lascarides, 2003, p.207). On a donc *Elaboration*(π_a, π_c). Grâce au connecteur *et* et aux informations temporelles contenues dans π_c et π_a , on peut inférer *Narration*(π_c, π_d), et comme ci-dessous *Elaboration*(π_a, π_d). On peut ensuite construire le constituant complexe π_A qui élabore π_a .

D'autres exemples de discours comme le classique (5) ont le même type de structure que (4), seule diffère la relation coordonnante qui relie les segments élaborants : il s'agit ici d'une relation de *Continuation*, qui a les mêmes contraintes sur le topique que *Narration*, mais pas de conséquences temporelles, elle est choisie pour relier un constituant qui a pour fonction de continuer à élaborer le même topique.

(5) (a) L'été de cette année-là vit plusieurs changements dans la vie de nos héros. (b) François épousa Adèle, (c) Jean-Louis partit pour le Brésil et (d) Paul s'acheta une maison à la campagne.

Pour construire cette structure de discours, on procède un peu différemment que pour (3) : ici on s'appuie sur des informations relevant de la sémantique lexicale telles que « les mariages, les départs, les achats de maison sont des changements (dans la vie) » pour inférer *Subtype_D* et, en suivant, *Elaboration* (π_a, π_b) et *Elaboration* (π_a, π_c). On applique ensuite un autre principe général qui nous permet d'attacher π_c à π_b avec *Continuation* (Prévoit, 2004, p. 60) :

$$(?(\beta, \delta, \lambda) \wedge \text{Elaboration}(\alpha, \beta, \lambda) \wedge \text{Elaboration}(\alpha, \delta, \lambda)) \rightarrow \text{Continuation}(\beta, \delta, \lambda)$$

Elaboration pose des contraintes spatio-temporelles sur le contenu du discours. Les éventualités de deux segments reliés par *Elaboration* sont telles que l'éventualité élaborante est incluse ou égale spatio-temporellement dans l'éventualité élaborée (\subseteq_{st} dénote l'inclusion spatio-temporelle) :

Contraintes spatio-temporelles $\text{Elaboration}(\alpha, \beta) \rightarrow e_\beta \subseteq_{st} e_\alpha$

3.3 Explication

Explication est la relation à l'œuvre dans un discours comme (6) où (b) explique pourquoi la proposition décrite par (a) est vraie :

(6) (a) Pierre est tombé. (b) Paul l'a poussé.

C'est une relation subordonnante dans laquelle le constituant expliquant est dominé par le constituant expliqué. Explication peut être inférée de façon non monotone grâce à la présence d'une relation de causalité potentielle entre les types des éventualités des segments à relier. Ceci est exprimé par le prédicat $Cause_D$ indiquant que le type de la première éventualité est une cause possible, dans l'environnement du discours courant (σ), de la survenue de la deuxième éventualité.

Inférer_Explication $(?(\alpha, \beta, \lambda) \wedge Top(\sigma, \alpha) \wedge Cause_D(\sigma, \beta, \alpha)) > Explication(\alpha, \beta, \lambda)$

Les contraintes temporelles posées par Explication sur le contenu du discours sont les suivantes l'éventualité expliquante (la cause) précède temporellement l'éventualité expliquée dans le cas où il s'agit d'un événement, elle ne peut lui être postérieure dans tous les cas :

Contraintes Spatio-Temporelles $Explication(\alpha, \beta) \rightarrow \neg(e_\alpha < e_\beta)$

$Explication(\alpha, \beta) \rightarrow (event(e_\beta) \rightarrow e_\beta < e_\alpha)$

4 Analyse de *d'abord*

Dans les études sur les adverbes du français, *d'abord* a la particularité d'être considéré à la fois comme un adverbe conjonctif – c'est-à-dire comme établissant un lien entre le segment de discours qu'il introduit et un autre segment de discours situé dans le contexte gauche (Guimier, 1996, Molinier et Lévrier, 2000) – et comme un adverbe organisateur du discours, marqueur d'ouverture d'une série qui va être décrite dans le contexte droit (Turco et Coltier, 1988, Adam et Revaz, 1989, Nøjgaard, 1992, Guimier, 1996). Cette propriété de « double attachement » discursif a motivé l'analyse de Bras (2007) en termes de relations subordonnantes et coordonnantes dans le cadre de la SDRT. Nous la présentons brièvement ci-dessous.

4.1 Les relations de *d'abord* avec ses contextes gauche et droit

Concernant les relations de *d'abord* avec son contexte gauche, Bras (2007) montre que le constituant contenant l'adverbe est attaché, par une relation de discours subordonnante, à un constituant considéré comme topique de discours. Les relations subordonnantes peuvent être des relations d'Elaboration, d'Explication ou de Résultat comme nous le verrons dans les exemples ci-dessous.

Pour ce qui est des relations avec son contexte droit, *d'abord* introduit un premier élément et annonce qu'il sera suivi d'un autre élément au moins, dont il prépare en quelque sorte l'interprétation et qui sera attaché par une relation coordonnante comme Narration ou Continuation.

Les exemples suivants, empruntés à (Bras 2007), illustrent les différentes combinaisons de ces relations subordonnantes et coordonnantes.

(7) (a) Une campagne de dénigrement, trop systématique pour n'être pas organisée, se déchaîna contre la personne même de Sélim. [...] (b) **D'abord** on répandit le bruit absurde que Sélim était alcoolique, alors qu'il était la sobriété même. (c) **Puis** on prétendit partout que Sélim était stérile, c'est-à-dire maudit. (Michel de Grèce, La nuit du sérail, 1982, p.373)

En (7), le segment (b), contenant *d'abord*, élabore (a), qui a ici le rôle de topique de discours explicite. Le segment (c), relié à (b) par une relation de Narration grâce à la présence du connecteur *puis*, poursuit cette élaboration. La structure de ce discours est schématisée en Figure 2 (en ignorant la relation de Contraste existant au sein de (b)).

L'usage de *d'abord* pour introduire, non plus la première étape d'une élaboration, mais le premier élément d'une explication est également très fréquent, que la relation subordonnante d'Explication soit explicitement marquée comme en (8), ou non, comme en (9) :

(8) La discipline essentielle est alors le latin, **d'abord parce que** le latin est la langue de l'église, **parce qu'aussi** c'est à Rome plus qu'à Athènes qu'on cherche ses maîtres à penser. (Encyclopédie pratique de l'éducation en France, 1960, p. 125)

(9) (a) Tu devrais épouser Laura, Jean-Pierre. (b) **D'abord**, elle est adorable. (c) **Ensuite**, c'est une des plus riches héritières du Brésil. (Romain Gary, Au-delà de cette limite votre ticket n'est plus valable, 1975, p. 253)

Dans les deux cas, la structure de discours obtenue, schématisée en Figure 3, a la même allure générale que la précédente.

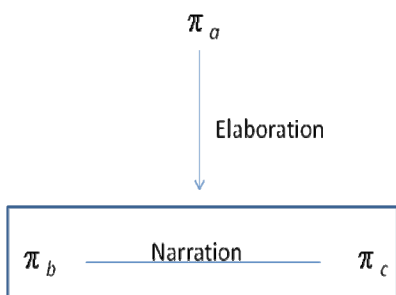


Figure 2 : structure de discours pour (7)

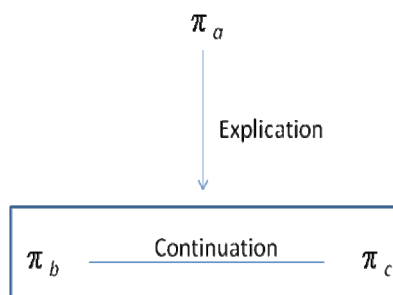


Figure 3 : structure de discours pour (8) et (9)

Bras (2007) présente également des cas de structures similaires avec une relation de Résultat subordonnante comme en (10) dont la structure est représentée en figure 4 :

(10) Une nuit, (a) il lui a demandé de faire la même chose pour moi [il lui a demandé de raconter mon enfance], [...] (b) **D'abord**, Mariella a fait une scène de jalousie, (c) **puis** elle a répondu qu'il fallait qu'elle me demande l'autorisation de dévoiler mon passé. (Evangé Hanska, Les Amants foudroyés, 1984, p. 62)

Finalement, Bras conclut que *d'abord* requiert des structures du type de celle schématisée en Figure 5.

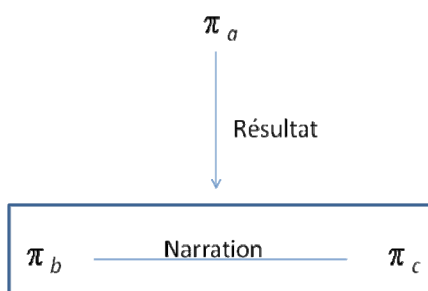


Figure 4 : structure de discours pour (10)

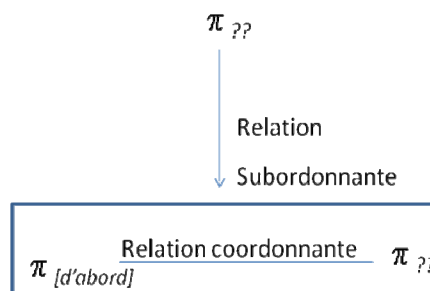


Figure 5 : structure de discours requise par *d'abord*

Il s'agit maintenant d'examiner la façon dont on construit ces structures pour comprendre le rôle que *d'abord* joue dans cette construction qui cherche à capturer le processus interprétatif des destinataires de ces discours.

Si on analyse (7) comme nous le proposons dans la section 3, nous allons pouvoir conclure que le prédicat Subtype_D est vérifié, grâce à un ensemble d'informations alliant connaissances du monde et sémantique lexicale et exprimant qu'un événement du type de celui qui est décrit en (b) est potentiellement un sous-événement du type de celui qui est décrit en (a). On peut donc inférer

Elaboration sans tenir compte de la présence de *d'abord*. Pourtant, la présence de *d'abord* semble faciliter cette inférence, ce que Guimier (1996, p.126) relève également en soulignant que « l'adverbe, s'il n'apporte pas de signification par lui-même, facilite [...] le travail interprétatif du destinataire ». En revanche, l'interprétation de *d'abord* en tête de (b) met le destinataire dans l'attente du segment (c), il se prépare ainsi à interpréter une autre étape de la campagne de dénigrement.

En (8), la présence de *parce que*, marqueur explicite de la relation d'Explication, rend inutile la présence de *d'abord* pour inférer la relation subordonnante, mais elle prépare là aussi l'interprétation de la seconde partie de l'explication. En (9), la relation d'Explication est inférable à partir de connaissances générales (avec le prédicat Cause_D présenté en 3.3) sur les types des événements en jeu. Pourtant, comme en (7), la présence de *d'abord* semble faciliter l'inférence de la relation subordonnante. Pour explorer plus avant cette intuition, nous reprenons des exemples de Péroz (2001), que nous avons légèrement transformés :

(11) (a) On a mangé sur la terrasse. (b) Mon père est arrivé.

(12) (a) On a mangé sur la terrasse. (b) D'abord, mon père est arrivé.

L'intuition de Péroz est qu'on interprète (11) comme une succession de procès considérés sur le même plan alors que (12) met en évidence une dépendance de (b) envers (a), avec l'idée que (a) et (b) ne sont pas au même niveau. Notre interprétation de ce couple d'exemples va dans le même sens : les constituants de (11) peuvent être interprétés par défaut comme étant reliés par Narration, relation coordonnante, en (12), en revanche, (b) ne fait pas progresser le temps du récit et *d'abord* nous contraint en quelque sorte à revenir sur (a) pour élaborer la description du procès en commençant par décrire les participants au repas par exemple :

(13) (a) On a mangé sur la terrasse. (b) D'abord, mon père est arrivé. (c) Puis on s'est tous assis autour de la table (d) et on s'est souhaité bon appétit. (e) Alors, j'ai servi à chacun une part de la quiche aux poivrons que j'avais préparée dans l'attente fébrile de ces retrouvailles familiales....

Bras (2007) traduit ce rôle de *d'abord* par l'introduction de la relation subordonnante minimale, notée \Downarrow , sans restreindre cette relation à Elaboration, parce que la relation imposée par *d'abord* doit rester compatible avec les autres relations subordonnantes inférables grâce aux autres indices en présence. Notons que pour l'exemple (13), les constituants suivants, en particulier (d) et (e), décrivent des événements qui sont clairement compris comme des sous-événements typiques d'un repas. On peut donc conclure que le prédicat Subtype_D décrit en 3.2 est vérifié et qu'on a Elaboration(π_a , π_d) et Elaboration(π_a , π_e). On peut alors appliquer le principe général de la SDRT évoqué en 3.2 pour inférer Elaboration(π_a , π_b) et Elaboration(π_a , π_c), en propageant en quelque sorte la relation d'Elaboration aux liens subordonnés minimalement par \Downarrow .

Il reste maintenant à examiner la construction de la SDRS de (10). Les événements décrits sont du type *demander/réagir à la demande/répondre*. Le type de lien entre *demande/réaction* et *demande/réponse* est encodé en SDRT par le prédicat Cause_D et permet d'inférer une relation de Résultat entre π_a et π_b d'une part, et entre π_a et π_c d'autre part. La relation de Résultat est normalement coordonnante en SDRT, mais elle peut, sous certaines conditions, devenir subordonnante (Asher et Vieu, 2005), et c'est le cas ici sous l'action de *d'abord* qui impose \Downarrow , ce qui permet de rendre compte correctement du résultat en deux temps exprimé par (b)-(c) en (10).

Nous pouvons maintenant résumer l'analyse de (Bras, 2007) en récapitulant les contraintes que pose *d'abord* sur la structure du discours. Concernant ses relations avec le contexte gauche, *d'abord* requiert un attachement vers le haut de la structure (vers l'arrière dans la linéarité du discours) : si une relation de contenu (comme Elaboration ou Explication) avec un constituant déjà introduit est inférable en dehors de la présence de *d'abord*, cette relation est toujours subordonnante (cf. (7), (8), (9)), sinon, c'est *d'abord* qui impose une relation subordonnante (cf. (10), (12), (13)). Concernant sa relation avec le contexte droit, *d'abord* requiert un attachement vers la droite (vers l'avant dans la linéarité du discours) : il introduit un premier élément et annonce qu'il sera suivi d'un autre élément au moins, dont il prépare en quelque sorte

l'interprétation et qui sera attaché par une relation coordonnante. Ces contraintes sont exprimées par l'axiome suivant :

Inférer_Subord_Coord_d'abord $(\exists(\alpha, \beta, \lambda) \wedge [d'abord](\beta)) > (\Downarrow(\alpha, \beta, \lambda) \wedge \exists \gamma \exists R \in \text{Coord } R(\beta, \gamma, \lambda))$

qui indique que le constituant β contenant *d'abord* doit être attaché par une relation subordonnante (\Downarrow) à un constituant α situé au dessus dans la structure de discours λ et qu'il attend un constituant γ auquel il sera attaché par une relation coordonnante R . *D'abord* est donc un marqueur discursif qui impose des contraintes sur le type des relations de discours, à rapprocher de *and* décrit par Gómez Txurruka (2003) comme imposant une relation coordonnante.

4.2 *D'abord* dans des structures énumératives

Bras (2007) décrit aussi le rôle de *d'abord* quand il entre dans des structures telles que celle qui apparaît dans l'exemple (14), une paraphrase de l'exemple (9). En suivant la terminologie de (Luc 2000), (14) peut être décrit comme une 'structure énumérative' faisant apparaître une 'amorce', *au moins deux raisons*, suivie d'une 'énumération' formée par l'ensemble des 'items' (b) et (c). Ici l'amorce précise le nombre d'arguments de l'explication et explicite la relation de discours d'Explication qui lie (a1) à (a2).

- (14) (a1) Tu devrais épouser Laura, Jean-Pierre, (a2) pour au moins deux raisons.
 (b) **D'abord**, elle est adorable. (c) **Ensuite**, c'est une des plus riches héritières du Brésil.

Avec les éléments donnés ci-dessus (4.1), nous pouvons construire la structure discursive de (14), qui est assez proche de celle de (9). Dans la suite nous allons nous intéresser à la relation subordonnante qui lie l'amorce à l'énumération, représentée ici par le constituant complexe regroupant les constituants représentant les items (b) et (c).

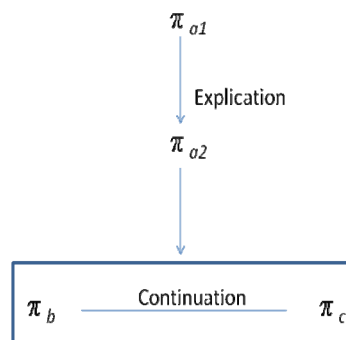


Figure 6 : structure de discours pour (14)

En dehors de Dalmas (1998) dans une étude contrastive avec l'allemand, et de Péroz (2001), qui décrit l'ensemble des emplois de *d'abord* quand il est détaché à gauche (emplois temporel, discursif, argumentatif), c'est cet emploi de *d'abord* comme "marqueur d'ouverture" de la série énumérative qui est le plus décrit dans les travaux de linguistique textuelle sur les organisateurs énumératifs (Adam et Revaz, 1989), ou "marqueurs d'intégration linéaire" (Auchlin, 1981, Turco et Coltier, 1988). Les structures énumératives seront présentées plus en détail dans la section 5, et analysées dans la section 6.

Nous allons simplement présenter ici deux des exemples de structures énumératives impliquant *d'abord* comme marqueur de premier item analysés par Bras (2007), et qui ont servi de point de départ à l'analyse que nous poursuivons dans cet article.

Elle analyse des exemples comme (15) où l'amorce contient un nom abstrait sous-spécifié qui attend des spécifications (Legallois, 2006), ce que l'établissement de la relation subordonnante imposée par *d'abord* doit permettre. Autrement dit, le topique de discours présenté en (a) dans (15) n'a pas le même statut que

les topiques explicites de (4), (5), (7), il attend des spécifications, et il faut pouvoir le mettre à jour au fur et à mesure de l'interprétation, un peu comme pour les topiques implicites qui sont construits quand on a une Narration sans topique explicite (cf. exemple (1) section 3). C'est pour cette raison qu'il n'est pas souhaitable d'imposer une relation d'Elaboration, qui ne permet pas une telle mise à jour.

(15) (a) L'hygiène dentaire [...] se fonde sur **plusieurs éléments**. (b) **D'abord**, il importe d'assurer un bon nettoyage des dents. (c) **Puis** il faut savoir faire appel au dentiste dès que l'on constate quelque chose d'anormal. (Encyclopédie médicale Quillet : nouvelle encyclopédie pratique de médecine et d'hygiène, 1965, p. 179)

Bras analyse aussi des cas d'amorces comportant des noms plus riches sémantiquement tels que des noms d'événement comme en (16) :

(16) (a) Tout en roulant vers l'île Saint-Louis, il prit **deux décisions**. (b) **D'abord**, il allait persuader sa mère de quitter Paris au moins jusqu'au procès et de descendre dans le Midi, peut-être même en Italie où elle ne risquerait rien. (c) **Et puis**, de son côté, il abandonnerait la rue de Longchamp et louerait un studio meublé. (Michel Droit, *Le Retour*, 1964, p. 298)

Elle se demande si, quand on interprète ce discours, c'est l'ensemble de la structure énumérative (l'amorce et les items) qui nous permet d'inférer que les événements décrits respectivement en (b) et (c) sont des décisions, ou si ce sont des connaissances externes qui feraient que nous saurions que des événements comme persuader quelqu'un de faire quelque chose ou abandonner un lieu sont des événements de décision. Elle penche vers un rôle plus fort de la structure énumérative et s'appuie sur cela pour lui accorder un rôle dans l'inférence d'une relation subordonnante entre l'amorce et chaque item. Nous reviendrons sur ce point à la section 6 et retenons pour le moment la proposition de détecter la présence de structures énumératives à partir de la présence d'une amorce et d'un introducteur d'item comme *d'abord*, nous verrons dans la section 6 ce qu'implique cette proposition.

5 Généralisation aux structures énumératives

Nous souhaitons maintenant aller au delà des seules structures énumératives contenant *d'abord*. Il s'agit de comprendre le rôle d'une structure énumérative en tant que telle et de savoir si sa présence doit systématiquement déclencher l'inférence d'une relation d'Elaboration, ou d'une autre relation subordonnante. Pour cette étude, nous partons des définitions des structures énumératives de (Luc, 2000, 2001, Luc et Virbel, 2001, Péry-Woodley, 2000, Porhriel, 2007).

5.1 Structure visuelle vs. Structure discursive

Un premier aspect qu'il nous paraît important de signaler lorsque l'on aborde la question des structures énumératives est le « continuum textuel » (Porhriel, 2007) qui va d'une structure plus linéaire (discursive) à une structure plus visuelle. Virbel (1989), qui a proposé le Modèle de l'Architecture Textuelle, émet l'hypothèse d'une équivalence fonctionnelle entre des manifestations visuelles (typographiques, dispositionnelles et ponctuationnelles) et des manifestations discursives (marques lexico-syntaxiques). Nous nous intéresserons ici aux énumérations linéaires discursives, également appelées structures énumératives « horizontales » (Porhriel 2007). Elles sont signalées par des marqueurs lexico-syntaxiques.

5.2 Description de l'amorce et des items

L'amorce est une phrase introductrice précédant l'énumération. Elle est caractérisée par sa complétude ou son incomplétude syntaxique :

Amorce Complète Notre plan se déroulera en trois parties : 1. (...) ; 2. (...) ; 3. (...).

Amorce Incomplète Les trois étapes de notre plan seront : 1. (...) ; 2. (...) ; 3. (...).

Nous nous limiterons ici au cas où l'amorce est complète. Généralement les amorces complètes se caractérisent par la présence d'un Groupe Nominal pluriel. Ils sont appelés 'classificateurs' chez Jacquemin et Bush (2000). Leur rôle est de définir la nature des items de l'énumération. Ils sont souvent introduits par un numéral (*deux, trois*) ou un article indéfini pluriel (*quelques, plusieurs*). Dans certains cas, l'amorce comprend un élément lexical cataphorique tel que *ci-dessous, suivant* comme nous le verrons dans l'exemple (20).

Une question importante pour la définition de l'énumération concerne le nombre minimum d'items requis. Sur cette question, les avis divergent. Certains travaux considèrent qu'il y a énumération à partir de deux items (Turco et Coltier, 1988). Pour d'autres, il est nécessaire d'avoir au moins trois items. Schnedecker (1998) par exemple inscrit les corrélats anaphoriques comme *l'un/l'autre, d'une part/d'autre part, le premier/le second, soit...soit, tantôt...tantôt* en dehors du champ des structures énumératives. Nous ne souhaitons pas trancher sur cette question. Cependant, même si nous sommes d'avis d'exclure les structures à corrélats anaphoriques des structures énumératives, nous souhaitons prendre en compte dans le cadre de nos travaux des structures énumératives à deux items, telles que (16).

Enfin, les marqueurs lexicaux des items, les Marqueurs d'Intégration Linéaire, sont quelques fois spécialisés, tel *d'abord* qui est un marqueur de premier item ou *enfin* qui est un marqueur de dernier item (Cadiot et al., 1985).

Dans cet article, nous allons considérer les structures énumératives d'au moins deux items. Nous observerons plus particulièrement les "énumérations prototypiques", dans lesquelles il y a présence conjointe de plusieurs types de marques. Plus précisément, les structures que nous étudierons dans un premier temps auront une amorce complète présentant un GN quantifié pluriel introduit par un numéral ou un indéfini et chacun des items sera introduit par un marqueur d'intégration linéaire lexical (*d'abord, puis, ensuite*). Dans un premier temps, nous ne considérerons pas les corrélats anaphoriques (*l'un/l'autre ; le premier/le second*) qui mettent en place un mécanisme de référence plus complexe.

6 Structures énumératives et structure du discours

6.1 Les Structures énumératives dans les théories du discours

Les structures énumératives sont parfois évoquées de manière implicite dans certaines approches du discours. Knott et Dale (1995) en particulier identifient dans leur liste de marqueurs de relations des marqueurs de début de 'séquence', de position dans une séquence, d'étape suivante. Ils identifient des séquences présentationnelles et des séquences temporelles. Ces données, très proches des énumérations considérées ici, sont cependant peu illustrées et peu discutées dans leurs travaux qui cherchent à recenser le plus de relations de discours possible plutôt qu'à détailler les particularités de chacune.

Le modèle de Polanyi (1988) ne traite pas spécifiquement, à notre connaissance, de structures énumératives, mais les structures N-aires récemment introduites (Polanyi et al, 2004) nous semblent intéressantes pour les analyser. Les structures N-aires sont une sorte de généralisation des relations binaires qui incluaient principalement des paires antécédent/conséquent ou des paires dialogiques comme question/réponse. L'extension à des structures N-aires est justifiée d'une part par l'existence de structures pragmatiques de plus haut niveau, d'autre part par la découverte de marqueurs signalant de telles structures, comme par exemple les adverbes en position initiale. Cependant, bien qu'ils identifient ces structures, Polanyi et ses collègues ne proposent ni une analyse fine des marques de surfaces les signalant, ni une analyse linguistique de leurs conséquences pour l'interprétation du discours. Sur ce dernier point, une différence est cependant avancée entre les constructions coordonnantes et les constructions N-aires. Ces deux structures supposent un constituant commun aux éléments coordonnés ou aux éléments groupés sous la relation N-aire, et auquel ils sont attachés. Dans le cas des relations coordonnantes, ce constituant contient les informations communes à tous les éléments coordonnants. A l'inverse, pour les constructions N-aires, ce constituant contient les informations de chaque élément de la construction ainsi que la relation

elle-même. La relation d'Énumération que nous proposons à la section 6 permet de construire des structures qui peuvent être rapprochées des structures N-aires de (Polanyi et al, 2004).

Bien que les travaux mentionnés ci-dessus pointent vers des structures discursives auxquelles les structures énumératives semblent correspondre, ils ne traitent pas spécifiquement du niveau textuel et en particulier des structures énumératives auxquelles nous nous intéressons ici. En revanche, le travail de Luc (2000) traite spécifiquement de ces structures dans un cadre discursif.

Luc (2000) propose de combiner la RST et le Modèle d'Architecture Textuelle (MAT). Ces deux modèles considérés séparément ne rendent pas compte de tous les aspects de la structure énumérative et, selon Luc, gagnent à être combinés. Le MAT fournit des informations pour la délimitation des segments complexes, à travers la mise en forme matérielle, tandis que la RST fournit les relations entre les objets textuels, mis au jour par le MAT, et les relations entre les segments à l'intérieur des objets textuels, lorsqu'un objet textuel est un segment complexe. Afin de prendre en compte les deux modèles, Luc a proposé d'utiliser un formalisme de représentation semblable pour la RST et pour le MAT et dresse entre les deux modèles des liens multiples afin de pouvoir les superposer et les comparer. Ce travail est un des rares à prendre en compte le niveau textuel (via le MAT) et celui des relations de discours en étudiant les correspondances et la complémentarité entre les deux niveaux. Nous voulons poursuivre la prise en compte conjuguée de ces deux dimensions, en examinant la façon dont elles interagissent avec le contenu sémantique des segments. Nous allons pour cela choisir un autre modèle des relations de discours, la SDRT, et intégrer en son sein des informations d'ordre textuel et architectural.

6.2 Les Structures Enumératives en SDRT

Nous avons déjà mentionné en section 2 et 3 que la SDRT allait au delà de l'analyse qui consiste à poser des relations de discours, et qu'elle proposait de rendre compte des effets de ces relations sur le contenu sémantique. Pour cela, elle définit formellement les relations de discours et les règles qui permettent de déclencher leur établissement d'une part, et d'inférer leurs effets sémantiques d'autre part.

6.2.1 Une première formalisation

Bras (2007) donne les premiers éléments de formalisation des structures énumératives en SDRT. L'axiome « Inférer_Subord_Coord_d'abord » présenté en section 4 et repris ci-dessous exprime formellement le rôle de *d'abord* :

Inférer_Subord_Coord_d'abord ($?(α, β, λ) \wedge [d'abord](β)$) >

$$(\Downarrow(α,β,λ) \wedge \exists \gamma \exists R \in \text{Coord } R(β,γ,λ))$$

En outre, elle propose que la reconnaissance d'une structure énumérative, i.e d'une amorce et de *d'abord* en tant que marqueur de premier item, permette d'inférer par défaut une nouvelle relation de discours, Énumération :

Inférer_Enumération ($?(α, β, λ) \wedge \text{amorce}(α) \wedge [d'abord](β)$) > Enumération(α,β,λ)

Dans cet axiome 'amorce(α)' est un prédicat qui est vrai si des indices de surface permettent d'établir que le segment α contient une amorce, suivant (Jacquemin et Bush, 2000). 'Enumération' est une relation subordonnante qui a la particularité de rendre compte de relations affleurant à la surface du texte, comme les relations Parallèle et Contraste, qu'on qualifie de "textuelles", par opposition aux relations sémantiques comme Elaboration, Explication, Résultat, ou Narration, concernant le contenu propositionnel des segments.

Nous proposons maintenant d'ajouter un axiome rendant compte du rôle des marqueurs d'item de suite ou de clôture pouvant suivre *d'abord* dans une structure énumérative comme l'illustrent les exemples (7), (9), (10), (13), (14), (15), (16) :

Inférer_Subord_Coord_MAI ($?(β,γ,λ) \wedge \exists α \text{Enumération}(α,β,λ) \wedge [\text{suite|puis|et puis|enfin}](γ)$) >

Enumération(α, γ, λ) \wedge item(γ) \wedge Continuation (β, γ, λ)

L'inférence de Enumération, relation subordonnante, permettra d'établir entre les constituants représentant les items la relation coordonnante minimale, Continuation. Cet axiome pourrait être complété par des axiomes spécifiques aux marqueurs de suite comme *puis* ou *ensuite* pouvant introduire une relation coordonnante plus riche comme Narration dans les cas où l'amorce est temporelle ou événementielle, mais nous laisserons ce problème de côté ici.

6.2.2 Généralisation

Nous souhaitons maintenant généraliser l'analyse d'une structure énumérative en *d'abord* à d'autres types de structures énumératives. Pour cela nous allons regarder des structures énumératives dont les items sont introduits par d'autres marqueurs tels que *premièrement/deuxièmement*, *en premier lieu/en deuxième lieu* mais qui ont nécessairement une amorce explicite. Observons l'exemple (17) :

(17) (a) *Deux conséquences importantes en découlent :*

(b) - **Premièrement**, aucune politique de stimulation de l'offre énergétique non pétrolière (relance du nucléaire ou du charbon, développement des énergies renouvelables, ou autres) n'est susceptible de réduire significativement la croissance de la demande de pétrole, qui tend à être concentrée sur ses usages spécifiques. En particulier, la croissance des besoins énergétiques liés au transport se traduira nécessairement par une augmentation de la demande de pétrole ;

(c) - **Deuxièmement**, les politiques de maîtrise de la demande énergétique ne peuvent avoir qu'un impact limité sur la demande de produits pétroliers sauf à modifier la consommation du secteur des transports. Dans ses usages non spécifiques, le pétrole a déjà été évincé par des technologies plus performantes. (corpus GEOPO (HoDac, 2007))

En s'appuyant sur les travaux de Guimier (1996), de Molinier (2000) et de Nøjgaard (1992), Schnedecker (2001) note que l'emploi des adverbes ordinaux sériels comme *premièrement*, *deuxièmement* requiert un contexte gauche et un contexte droit. Elle décrit leur contribution comme le déclenchement d'un double mécanisme d'anticipation et de rétroaction, ce qui rejoint la description que nous avons faite plus haut de la contribution de *d'abord*, en tant que marqueur de premier item (MPI). En (17), *premièrement* est un MPI qui est subordonné à l'amorce et qui attend un constituant à sa droite (un énoncé formé seulement de l'amorce et du premier item serait mal formé), *deuxièmement* est un marqueur d'autre item (MAI). Nous proposons donc les axiomes suivants pour les marqueurs d'items et l'énumération, qui sont une généralisation des axiomes proposés à la section 6.2.1 :

Inférer_Subord_Coord_MPI $(?(\alpha, \beta, \lambda) \wedge \text{MPI}(\beta)) > (\Downarrow(\alpha, \beta, \lambda) \wedge \exists \gamma \exists R \in \text{Coord } R(\beta, \gamma, \lambda)) \wedge \text{item}(\gamma)$

Inférer Enumération_bis $(?(\alpha, \beta, \lambda) \wedge \text{amorce}(\alpha) \wedge \text{MPI}(\beta)) > \text{Enumération}(\alpha, \beta, \lambda)$

Inférer_Subord_Coord_MAI_bis $(?(\beta, \gamma, \lambda) \wedge \exists \alpha \text{Enumération}(\alpha, \beta, \lambda) \wedge \text{MAI}(\gamma)) >$

$(\text{Enumération}(\alpha, \gamma, \lambda) \wedge \text{item}(\gamma) \wedge \text{Continuation}(\beta, \gamma, \lambda))$

Ces axiomes permettent de construire la structure de discours représentée en Figure 7. Notre traitement des MIL met en œuvre le double principe de rétroaction/anticipation de Schnedecker, mais nous opérons une distinction entre le MPI et les MAI : nous traitons le MPI comme s'attachant au contexte gauche par une relation subordonnante et au contexte droit par une relation coordonnante tandis que le MAI s'attache par deux fois au contexte gauche, d'abord à l'amorce par une relation subordonnante, puis au segment décrivant l'item précédent avec une relation coordonnante. Dans des travaux futurs, nous envisageons également de traiter les marqueurs de dernier item, comme *enfin* ou les adverbes ordinaux précédés de *et*, par exemple *et troisièmement*.

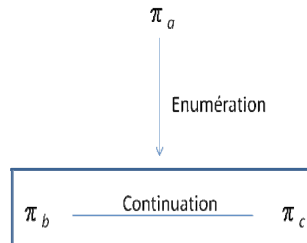


Figure 7 : structure de discours pour (17)

Cette structure repose sur une relation textuelle d'Enumération mais ne fait pas apparaître de relation de contenu entre l'amorce et les items. Or, il semble que tout locuteur a une connaissance générale qui dit qu'une conséquence est nécessairement une éventualité (i.e. un événement ou un état). Ce type de connaissance suffit en SDRT pour inférer la relation de contenu Elaboration, en plus de la relation Enumération. C'est d'ailleurs ce qui est à l'œuvre quand on interprète l'exemple (18) adapté de (17) :

(18) *Une conséquence importante en découle. Aucune politique [...] n'est susceptible de réduire significativement la croissance de la demande de pétrole [...]*

Le même constat pourrait être fait pour l'exemple (16) examiné en section 4.2, sur la base de la connaissance générale que toute action intentionnelle est susceptible de faire l'objet d'une décision préalable, i.e. que le type "décision" domine le type "décision de persuader quelqu'un de faire quelque chose..." dans une hiérarchie de types d'éventualités. Ainsi nous pourrions inférer une relation d'Elaboration dans un exemple comme (19) et donc le faire aussi pour (16) :

(19) *Il prit une décision. Il allait persuader sa mère de*

Mais ce type de connaissance qui se situe au niveau des types des éventualités, généralement encodé en SDRT par le prédicat subtype_D, s'il est en théorie disponible, est, en pratique, difficile à établir. En revanche, l'information de surface apportée par la structure énumérative est mise à disposition par le discours et facilite l'interprétation. Nous plaçons ici pour l'exploiter, même si l'inférence de la relation d'Elaboration rend en principe inutile la présence de la relation Enumération qui n'apporte pas plus d'information sur le plan sémantique. En outre, comme nous l'avons signalé en section 4.2, l'inférence de la relation d'Elaboration en SDRT bloquera la mise à jour du topique, i.e. du constituant correspondant à la première phrase. Or ce n'est peut-être pas souhaitable pour un topique aussi sous-spécifié, et on pourrait préférer ne pas exploiter les relations de subsumption avec les types les plus hauts dans la hiérarchie des types proposée par la SDRT pour inférer subtype_D, pour laisser, dans de tels cas, la place à la seule relation de topique.

6.2.3 Cas d'items structurés

Examinons maintenant un exemple présentant des items plus complexes sur le plan de la structure du discours :

(20) *Ces exemples appellent à la prudence. (a) Les facteurs essentiels sont ici les suivants. (b) En premier lieu, (b1) la guerre a en général pour les acteurs locaux un intérêt politique de premier ordre, souvent bien plus important que celui des Etats-Unis. (b2) Leur tolérance à la souffrance est donc plus grande. (c) En deuxième lieu, en dépit de leur taille réduite, ces acteurs supplantent d'ordinaire les Etats-Unis dans une ressource précieuse : le nombre d'hommes en âge de combattre. Même s'il n'est plus l'élément déterminant de la guerre terrestre, il reste un facteur critique, notamment en ville, dans la jungle ou en montagne. (d) Troisièmement, les " locaux " disposent en général d'un avantage [...]. (e) Quatrièmement, nombre des chefs militaires de ces Etats ou entités ont été formés dans le monde développé - [...]. (f) Cinquièmement,*

l'arsenal nécessaire au combat rapproché, [...] est beaucoup moins coûteux que les armements nécessaires à la guerre dans les "espaces communs".

Cet exemple montre que les items d'une énumération peuvent être des éléments discursifs relativement complexes et que les marqueurs d'items peuvent avoir portée au delà de la phrase syntaxique dans laquelle ils se trouvent. C'est le cas pour le premier item de cet exemple : (b) a portée sur (b1) et (b2). Schnedecker (2001) et Jackiewicz (2002), inter alia, ont montré que les marqueurs d'intégration linéaire en position détachée à l'initiale ont un rôle cadratif, selon le modèle de Charolles (1997), c'est-à-dire qu'ils introduisent un cadre pouvant aller au delà de la phrase. Ce rôle d'introducteur de cadre a été formalisé en SDRT dans (Vieu et al., 2005) pour le cas des adverbiaux de localisation spatio-temporelle détachés à l'initiale (i.e. en position d'IP-adjoint), à partir des travaux de Le Draoulec et Péry-Woodley (2005). Ces adverbiaux détachés y sont traités comme des introducteurs de nouveaux topiques sous-spécifiés, considérés comme des constituants du discours à part entière, qui dominent discursivement les segments dans la portée du cadre. Nous proposons de traiter les marqueurs d'items détachés à l'initiale de la même façon que (Vieu et al., 2005), ce qui nous permet de grouper π_{b1} et π_{b2} , correspondant aux segments (b1) et (b2), sous la dominance de π_b , un constituant équivalent à "ce qui est à prendre en compte premièrement", et, ce faisant, d'introduire un niveau de structuration supplémentaire.

Bien que justifié au niveau linguistique par le rôle discursif de ces marqueurs d'items détachés à l'initiale, ce niveau de structuration intermédiaire pourrait sembler superflu du point de vue de la structuration du discours. En réalité, il n'en est rien car, comme le montre l'exemple (20), π_{b1} peut être le premier argument d'une relation coordonnante introduite, non par l'item suivant mais par une proposition incluse dans l'item en cours de développement : π_{b2} . Cet attachement serait problématique puisqu'il violerait le principe de la frontière droite, bien établi dans les théories de la structure discursive (Hobbs, 1982, Mann, et Thompson, 1988, Webber, 1991, Asher, 1993). La frontière droite est l'ensemble des constituants situés sur le chemin le plus à droite allant du dernier constituant introduit au sommet de la structure. Les constituants situés sur cette frontière sont dits "ouverts" et les autres "fermés". Le principe de la frontière droite affirme que les constituants pouvant être étendus par la suite du discours sont les constituants ouverts. Un attachement coordonnant ferme donc un point d'attachement (ainsi que tous les constituants qu'il domine) bloquant ainsi toute possibilité de continuation future à ce niveau.

Pour l'exemple (20), sans le niveau hiérarchique supplémentaire que nous introduisons entre (b) et (b1), le constituant (π_{b-b1}) qui correspondrait à (b-b1) serait fermé par (π_{b2}). Or, le prochain item (π_c) de l'énumération serait censé précisément s'attacher sur ce constituant (π_{b-b1}) contenant le marqueur d'item précédent.

Essayons maintenant de voir quel doit être le contenu de π_b . Dans (Vieu et al., 2005), le constituant correspondant à l'adverbial détaché à gauche contient la forme sémantique nue de l'adverbial qui est ensuite saturée existentiellement. Dans la suite de la construction de la SDRS, ce constituant sera relié à d'autres constituants représentant les segments du contexte droit, qu'il va dominer pour rendre compte de la propriété de portée du cadre spatio-temporel. Dans notre cas, l'information apportée par l'adverbial n'est pas de même nature : il s'agit seulement d'une information concernant la structuration de l'énumération. Cette différence entre les cadratifs spatio-temporels et les MILs cadratifs est d'ailleurs relevée par Schnedecker (2001) qui constate que seuls les seconds peuvent être supprimés car ils n'ont pas de contenu sémantique. Le traitement que nous proposons est le suivant :

$\exists \pi, \pi: [\exists P \exists e (P(e) \wedge \text{Position-Enum}(e))]$

Cela signifie qu'au marqueur d'item correspond un constituant fortement sous-spécifié qui précise seulement qu'une éventualité va être introduite et donne sa position dans l'énumération.

Cette solution technique est justifiée sur le plan linguistique par le rôle discursif de certains éléments détachés à l'initiale (Vieu et al. 2005, Asher et al. 2007). Elle présente aussi l'avantage de procurer une structure particulièrement homogène, comme illustré en Figure 8, qui pourra faciliter, entre autres, la détection d'énumérations mal formées.

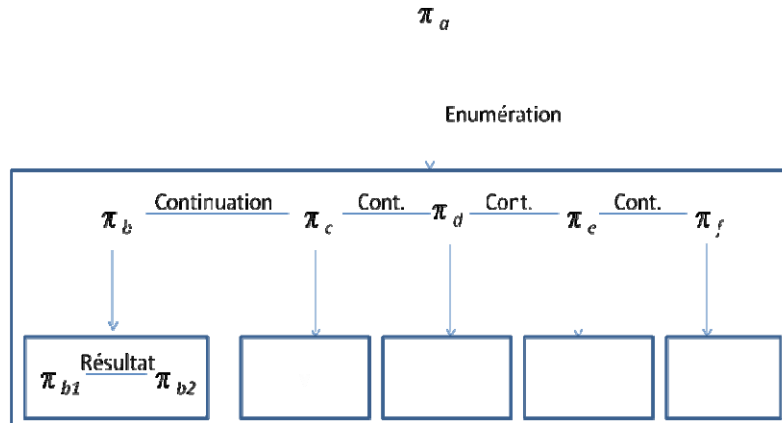


Figure 8 : structure du discours (20)

6.3 Discussion sur le statut des topiques

Grâce à l'examen des nouvelles données que sont les structures énumératives, nous pouvons réexaminer le statut des topiques discursifs en SDRT. Ceux-ci peuvent s'organiser sur une échelle implicite/explicite. A une extrémité de l'échelle se trouvent des constituants explicites correspondant à des segments de discours et qui sont reliés à leurs éléments subordonnés par la relation d'Elaboration (cas de (4) et (7) entre autres). A l'autre extrémité, nous avons des constituants implicites, inférés grâce à plusieurs segments explicites coordonnés pour lesquels un topique commun a pu être construit et auxquels ils sont liés par la relation de topique \Downarrow (cas de (1)). Entre les deux, des topiques hybrides, pour les marqueurs d'items cadratifs (π_b) et pour les amorces de structures énumératives qui forment des constituants topiques fortement sous-spécifiés (π_a). Nous les qualifions de semi-implicites parce qu'ils sont à la fois associés à des formes linguistiques, donc explicites, mais ils doivent être mis à jour par les éléments constitutifs des items ce qui les rend similaires à un topique implicite. Les topiques introduits par les adverbiaux spatio-temporels cadratifs peuvent également être versés dans cette catégorie des topiques semi-implicites.

La division effectuée par le passé en SDRT entre les relations d'Elaboration et de Topique est réductrice par rapport à la richesse des objets décrits ici. Cependant, nous ne souhaitons pas introduire une nouvelle relation qui serait intermédiaire entre les deux, et considérons simplement, pour l'instant, que les topiques hybrides sont intégrés au discours à l'aide de la relation de Topique. La distinction entre Elaboration et Topique met en jeu des considérations sur la nature explicite/implicite des segments, et surtout la non mise à jour du topique explicite vs la mise à jour du topique implicite ou semi-implicite. Nous privilégions donc dans nos analyses la relation de Topique (avec Enumération).

Comme nous l'avons déjà discuté plus haut à propos des exemples (16)-(19), l'amorce de l'énumération introduit fréquemment ce type de topique hybride. Le constituant construit pour l'amorce va recevoir un contenu sémantique fortement sous-spécifié et pourra être mis-à-jour par le contenu des items de l'énumération, qui devront dans la mesure du possible vérifier qu'ils peuvent bien correspondre au « classificateur énumératif » afin d'expliquer par exemple l'étrangeté de (21),

(21) *Hier j'ai pris deux décisions : premièrement qu'il pleuvrait ce matin, deuxièmement que la lune serait pleine à la fin de la semaine.*

Schnecker (2001) décrit cette caractéristique du nom classificateur de l'amorce en lui assignant le rôle d'imposer une contrainte d'identité sémantique sur les items. La question qui reste difficile à résoudre est celle de la plasticité de la contrainte lors de l'interprétation : dans les cas limites, comme ci-dessus, certains destinataires trouveront un contexte permettant de satisfaire la contrainte, d'autres jugeront le discours incohérent.

Pour rendre compte de manière plus fine des structures énumératives en SDRT, il faudrait pouvoir rendre compte du degré de spécification de l'amorce qui influe sur la possibilité d'inférer Elaboration (cas d'une amorce suffisamment spécifiée), ou simplement Enumération (cas d'amorce sous-spécifiée devant être mise à jour), ou encore de conclure à un discours incohérent (cas d'amorce sous-spécifiée incompatible avec les items).

L'introduction de la relation d'Enumération permet d'ores et déjà de séparer la contribution textuelle de la contribution proprement sémantique à l'édifice discursif, et ainsi de rendre compte du rôle de facilitateur d'interprétation des structures énumératives.

De plus, l'hypothèse quelquefois admise d'un isomorphisme entre les structures suggérées par les niveaux textuel et sémantique, plaidant contre la nécessité de relations textuelles, demande encore à être vérifiée Luc (2001), par exemple, suggère que les structures ne sont pas isomorphes. Pour faire cette vérification, il faudra bien être capable de séparer ces deux niveaux et de comparer leur contribution de manière systématique.

Enfin, la relation Enumération permettra peut-être d'aider à distinguer plusieurs cas de relations d'Elaboration ayant des sémantismes différents. Comme nous l'avons déjà introduit, la relation d'Elaboration est difficile à définir. Certaines théories comme la RST proposent plusieurs définitions pour la relation d'Elaboration. En ce qui concerne la SDRT, beaucoup plus précautionneuse en ce qui concerne la prolifération des relations, elle ne propose qu'une définition de la relation d'Elaboration. Mais plusieurs travaux (Prévot, 2004, Fabricius-Hansen et Behrens, 2001) ont dégagé la nécessité d'introduire une relation d'élaboration d'objet pour traiter des segments discursifs qui n'introduisent pas des événements mais des entités objets.

7 Conclusion

Nous avons proposé un traitement des structures énumératives dans le cadre de la sémantique du discours. Les approches discursives qui se sont traditionnellement intéressées à ce sujet se sont focalisées sur l'organisation textuelle, privilégiant par exemple les indices typo-dispositionnels. Inversement, la sémantique du discours et en particulier la théorie dans laquelle nous avons situé nos analyses, la SDRT, utilise avant tout la sémantique des énoncés individuels pour atteindre compositionnellement le sens du discours et néglige quelque peu les aspects textuels. Nous avons traité les énumérations prototypiques en nous appuyant d'une part sur l'étude de certains marqueurs de discours comme *d'abord* (Bras, 2007) et d'autre part sur l'analyse de certains adverbiaux en position initiale détachée (Vieu et al., 2005). Les similarités que nous mettons en évidence entre *d'abord* et les marqueurs d'items énumératifs nous ont conduits à proposer un traitement homogène fondé sur un niveau de structuration intermédiaire correspondant aux items. Nous avons également défendu l'ajout d'une relation d'Enumération à la SDRT pour rendre compte de l'utilité du niveau textuel dans l'établissement efficace de la cohérence discursive. Au delà du traitement des structures énumératives dans un cadre sémantique, notre travail a exploré les interactions entre les niveaux *textuel* et *sémantique* en examinant des structures où les deux niveaux sont explicites et semblent se combiner pour faciliter l'interprétation du texte.

Cependant nous n'avons examiné que des exemples prototypiques pour lesquels les structures textuelles et sémantiques sont isomorphes. Il faut maintenant s'attaquer à des cas moins canoniques comme ceux étudiés par Luc (2000) dans lesquels la structure textuelle semble orthogonale à la structure sémantique. Les structures énumératives à deux temps, décrites par Porhiel (2007) constituent également un riche terrain d'étude pour mettre à l'épreuve notre proposition (Vergez-Couret et al, à paraître). Enfin, nous aimerions approfondir la question des corrélats anaphoriques à structure bipartite étudiés par Schnedecker (1998) dans le cadre que nous avons installé.

Références bibliographiques

- Adam, J.M., Revaz, F. (1989). Aspects de la structuration du texte descriptif: les marqueurs d'énumération et de reformulation, *Langue Française* 81, pp. 59-98.
- Asher, N. (1993). *Reference to abstract objects in Discourse*. Dordrecht: Kluwer Academic Press.
- Asher, N. (2004). Discourse Topic, *Theoretical Linguistics* 30, pp. 161-201.
- Asher, N., Hardt D., Busquets, J. (2001), Discourse Parallelism, Scope, and Ellipsis, *Journal of Semantics* 18(1), pp. 1-25.
- Asher, N., Lascarides, A. (2003). *Logics of Conversation*. Cambridge: Cambridge University Press.
- Asher, N., Prévot, L., Vieu, L. (2007). Setting the Background in Discourse, *Discours* 1.
- Asher, N., Vieu, L. (2005). Subordinating and Coordinating Discourse Relations, *Lingua* 115(4), pp. 591-610.
- Auchlin, A. (1981). Réflexions sur les marqueurs de structuration de la conversation, *Etudes de Linguistique appliquée* 44, pp. 97.
- Bras, M. (2007) French adverb *d'abord* and Discourse Structure, in Aurnague, M., Larrazabal, J. Korta, K. (eds) *Language, Representation and Reasoning. Memorial Volume to Isabel Gómez Txurruka*, pp. 77-102.
- Cadiot, A., et al. (1985). *Enfin*, marqueur métalinguistique, *Journal of pragmatics* 9, pp. 199-239.
- Charolles, M. (1997). L'encadrement du discours : univers, champs, domaines et espaces, *Cahier de Recherche Linguistique* 6, pp. 1-73, Landisco, Nancy.
- Dalmas, M. (1998). D'abord, et après ? Le marqueur d'intégration français et ses cousins germaniques in C. Schnedecker (ed.) *Les corrélats anaphoriques*, *Recherches linguistiques XXII*, Université de Metz, Klincksieck, Paris, pp. 75-95.
- Danlos L. (2006). Partition d'une entité par le truchement d'opérateurs aspectuo-temporels, in G. Kleiber, C. Schnedecker, & A. Theissen (éds), *La relation PARTIE--TOUT*, Peeters, pp 751-770.
- Fabricius-Hansen, C., Behrens, B. (2001). « Elaboration and related discourse relations viewed from an interlingual perspective », SPRIK report of the project *Languages in Contrast*, University of Oslo (www.hf.uio.no/german/sprik).
- Grosz, B. & Sidner, C. L. (1986), Attention, intentions, and the structure of discourse, *Computational Linguistics* 12(3), pp. 175-204.
- Guimier, C. (1996). *Les adverbes du français*. Paris/Gap : Ophrys.
- Grice, H.-P. (1975). *Logic and conversation*, volume 3, pp. 41-58. P. Cole (ed.), New York: Academic Press Edition.
- Halliday, M. & Hasan, R. (1976). *Cohesion in English*., London: Longman.
- Ho-Dac, M. (2007). *La position initiale dans l'organisation du discours : une exploration en corpus*, Thèse de doctorat, Université Toulouse Le Mirail, Décembre 2007.
- Hobbs, J. R. (1979). Coherence and Coreference, *Cognitive Science* 3, pp. 67-90.
- Hobbs, J. R. (1985). On the Coherence and Structure of Discourse (CSLI-85-37), Technical report, Center for the the study of language and information, Stanford University.
- Hobbs, J. R. (1982). Towards an understanding of coherence in discourse in *Strategies for natural language processing*, Erlbaum, Hillsdale, NJ.

- Hovy, E. & Maier, E. (1992). Parsimonious or Profligate: How Many and Which Discourse Structure Relations? (RR-93-373), Technical report, USC Information Sciences Institute.
- Jackiewicz, A. (2002). Repérage et délimitation des cadres organisationnels pour la segmentation automatique des textes in *Actes de CIFT'02*, Hammamet, Tunisie, pp. 95-107.
- Jacquemin, C., Bush, C (2000). Fouille du Web pour la collecte d'Entités Nommées in E. Wehrli (ed), *Proceedings, TALN'2000*, Lausanne, Suisse.
- Kehler A. (2002), *Coherence, Reference, and the Theory of Grammar*, CSLI Publications.
- Knott, A. (1996). *A Data-Driven Methodology for Motivating a Set of Coherence Relations*. Thèse de doctorat, Department of Artificial Intelligence, University of Edinburgh.
- Knott, A. & Dale, R. (1995). Using Linguistic Phenomena to Motivate a Set of Coherence Relations. *Discourse Processes* 18(1), pp. 35-62.
- Le Draoulec, A. & Péry-Woodley, M.-P. (2005). Encadrement temporel et relations de discours, *Langue Française* 148, pp.45-60.
- Legallois, D. (2006). Quand le texte signale sa structure : la fonction textuelle des noms sous-spécifiés, in *Corela, Organisation des textes et cohérence des discours*.
- Luc, C. (2000). *Représentation et composition des structures visuelles et rhétoriques du texte. Application à la génération de textes formatés*. Thèse de doctorat, Université Paul Sabatier, novembre 2000.
- Luc, C. (2001). Une typologie des structures énumératives basée sur les structures rhétoriques et architecturales du texte in *Actes de TALN 2001*, pp. 263-272.
- Luc, C., Virbel, J. (2001). Le modèle d'architecture textuelle : fondements et expérimentation, *Verbum* XXIII(1), pp. 103-123.
- Molinier, C. & Lévrier, F. (2000). *Grammaire des adverbes. Description des formes en -ment*. Genève/Paris : Droz.
- Mann, W.C. & Thompson, S., (1988). Rhetorical Structure Theory: towards a Functional Theory of Text Organization, *Text* 8-3, pp. 243-281.
- Marcu, D. (1997). *The Rhetorical Parsing, Summarization, and Generation of Natural Language Texts*, PhD thesis, Department of Computer Science, University of Toronto.
- Nøjgaard, M. (1992). *Les adverbes français : essai de description fonctionnelle*. Copenhagen : Munksgaard.
- Péroz, P., (2001). La construction des valeurs sémantiques de d'abord en position de détachement gauche, *Buletin Științific "Fascicula Limbi Moderne"*, seria A, vol.XV, Universitatea de Nord Baia Mare, (Roumanie), pp.59-72.
- Péry-Woodley, MP (2000). *Une pragmatique à fleur de texte*, mémoire d'Habilitation à diriger des Recherches, Université Toulouse Le Mirail.
- Polanyi, L. & Scha, P. (1984). A syntactic approach to discourse semantics, in *COLING84*.
- Polanyi, L. (1988). A formal model of the structure of discourse, *Journal of Pragmatics* 12.
- Polanyi, L., Culy, C., van der Berg, M., Thione, G. L. & Ahn, D. (2004). A rule Based Approach to Discourse Parsing, *Proceedings of SIGDIAL 2004*.
- Porhiel, S. (2007). Les structures énumératives à deux temps, *Revue Romane* 42-1, pp. 103-135.
- Schnedecker, C. (1998). Les corrélats anaphoriques : une entrée en matière , *Recherches Linguistiques* 22, pp. 3-36.

Schnedecker, C. (2001). Adverbes ordinaux et introducteurs de cadre : aspects linguistiques et cognitifs, *Lingvisticae Investigationes XXIV:2*, pp. 257-287.

Turco, G., Coltier, D. (1988). Des agents doubles de l'organisation textuelle : les marqueurs d'intégration linéaire, *Pratiques 57*, pp. 40-60.

Gómez-Txurruka, I. (2003). The natural language conjunction *and*, *Linguistics and Philosophy 26*, pp. 255-285.

Vergez-Couret M., Prévot L. & Bras, M. (soumis). Interleaved discourse, the case of two-step enumerative structures.

Vieu, L., Bras, M., Asher, N. & Aurnague, M. (2005). Locating Adverbials in Discourse, *Journal of French Linguistics Studies 15*, pp. 173-193.

Virbel, J. (1989). The contribution of Linguistic Knowledge to the Interpretation of Text Structures, in J. André, V. Quint, R.K. Furuta (eds.) *Structured Documents*, pp. 161-181, Cambridge : CUP.

Webber, B. (1991). Structure and Ostentation in the interpretation of discourse deixis, *Language and Cognitive processes 6*, pp. 107-135.

Webber, B. L. (1988). Discourse Deixis and Discourse Processing (MS-CIS-86-74), Technical report, Department of Computer and Information Science. University of Pennsylvania.