

**Intégration d'informations
sur la structure communicative
dans un système de génération de textes**

Mémoire présenté pour l'obtention du

**DEA en Linguistique Théorique, Descriptive et Automatique
Mention Linguistique Informatique**

par

Laurence DELORT

Sous la direction de **Laurence DANLOS**
et de **Sylvain KAHANE**

Université Denis Diderot, Paris 7

**UFR de Linguistique
École Doctorale Sciences du Langage**

Année 2001-2002

Remerciements

Je tiens à exprimer ma plus profonde reconnaissance à Laurence Danlos, ma directrice de recherches, tant pour son encadrement dynamique et sa disponibilité que pour ses remarques qui ont souvent éclairé mon travail.

Je tiens également à adresser mes plus vifs remerciements à Sylvain Kahane qui a été un directeur de recherches disponible, attentionné et encourageant. Je pense également à ses nombreuses réflexions qui ont guidé ma recherche.

Je voudrais remercier Igor Mel'čuk et Lidija Iordanskaja pour avoir répondu à mes questions de novice en structure communicative.

Mes remerciements vont également à Laurent Roussarie, pour ses éclairages théoriques, et à Sophie Prévost, pour ses références bibliographiques.

Et je remercie l'équipe Lattice-Talana pour m'avoir donné l'occasion de présenter mon travail lors du colloque Théorie Sens-Texte 2002. Je pense à tous ceux que j'y ai rencontrés et avec qui j'espère avoir de nombreuses occasions de travailler.

Une pensée amicale pour Aurélie Névéol, ma camarade de DEA, avec qui j'ai eu le plaisir de travailler cette année.

Je tiens enfin à remercier chaleureusement Benoît qui malgré la distance a su me donner toute son affection et ses encouragements.

Table des matières

Introduction	5
1. Notions préliminaires et cadres théoriques	7
1.1. Structure et relations de discours	7
1.1.1. La <i>Discourse Representation Theory</i>	7
1.1.2. Les objets abstraits du discours	9
1.1.3. Les relations de discours	10
1.2. Structure communicative de la phrase	17
1.2.1. La partition thème~rhème	17
1.2.2. Le topique	18
1.2.3. Le focus rhématique	19
1.2.4. Les spécifieurs communicatifs	19
1.2.5. L'opposition donné~nouveau	20
1.2.6. La focalisation.....	21
Conclusion première partie	23
2. Intégration d'informations communicatives dans un système de génération de textes ..	24
2.1. Structure du discours en génération de textes	24
2.1.1. Tâches d'un système de génération	24
2.1.2. La structuration de document.....	26
2.1.3. Un pas vers l'organisation communicative du discours : <i>une autre raison de</i> 30	
2.1.3.1. Étude linguistique de <i>une raison de</i>	31
2.1.3.2. Étude linguistique de <i>une autre raison de</i>	33
2.2. Structure communicative en génération de textes	36
2.2.1. De la forme logique à une SDRS	36
2.2.2. D'une SDRS à la structure communicative des phrases	39
2.2.2.1. La relation Narration.....	40
2.2.2.2. La relation Parallèle.....	40

2.2.2.3.	La relation Commentaire	41
2.2.2.4.	La relation Élaboration	42
2.2.2.5.	La relation Particularisation	46
2.2.2.6.	La relation Explication	48
2.2.2.7.	La relation Résultat	49
Conclusion deuxième partie		51
3.	Génération d'un texte avec des informations communicatives	52
3.1.	Présentation du texte à générer.....	52
3.2.	Processus de génération du texte	54
3.2.1.	Génération du segment discursif 1.....	54
3.2.1.1.	De la forme logique aux SDRSs.....	54
3.2.1.2.	Des SDRSs aux structures communicatives	63
3.2.2.	Génération du segment discursif 2	66
3.2.2.1.	De la forme logique aux SDRSs.....	66
3.2.2.2.	Des SDRSs aux structures communicatives	70
Conclusion troisième partie.....		72
Conclusion générale		73
Annexe 1		74
Annexe 2		75
Annexe 3		76
Annexe 4		77
Annexe 5		78
Annexe 6.....		79
Références bibliographiques		82

Introduction

Le présent travail est parti d'une observation. Si nous regardons, par exemple, le discours suivant :

P1 : Paul a acheté un roman.

P2 : C'est le dernier David Lodge.

nous pouvons remarquer deux choses : d'une part, il y a une relation entre les événements dénotés dans les deux phrases. En l'occurrence, l'objet 'un roman' est introduit dans P1 et on l'explique dans P2. D'autre part, si nous analysons la structure thème~rhème des phrases, nous avons pour P1, la partition thème 'Paul'~rhème 'a acheté un roman' (avec un focus sur 'un roman') et pour P2, la partition thème 'cela' (anaphore de 'un roman')~rhème 'est le dernier David Lodge'. Ainsi, nous pouvons voir que le focus rhématique de P1 'un roman' est repris en thème dans P2 sous la forme du pronom démonstratif *cela* (*c*).

À partir de cette observation, nous pouvons nous demander si la cohérence d'un discours est tributaire de la corrélation entre une relation de discours et la structure communicative des phrases. Autrement dit, une relation de discours peut-elle imposer une organisation communicative particulière aux phrases ? Pour une relation de discours, existe-il des progressions thématiques privilégiées ?

Si l'interaction relation de discours/structure communicative s'avère systématique, ces informations pourraient être pertinentes dans un système de Traitement Automatique des Langues et notamment dans un système de génération de textes. En effet, le but d'un générateur de textes est de produire des discours les plus cohérents possible et intégrer de telles informations communicatives permettrait d'assurer la cohérence du discours. Il s'agit pour nous de démontrer comment des informations communicatives peuvent garantir la cohérence du discours à générer et de répondre à la question plus générale : en quoi des informations communicatives peuvent-elles améliorer un système de génération de textes ?

Dans l'architecture d'un générateur, le module de planification de texte prend en charge les décisions pour la structure du discours. C'est à ce niveau que nous allons intervenir : notre travail va tout d'abord consister à proposer le plan de document approprié en fonction du propos du discours (le topique). Par la suite, il va s'agir de voir en quoi la structure du discours, et en particulier les relations de discours, peut imposer des structures communicatives phrastiques particulières et ainsi de proposer la progression thématique correspondante au plan de document. Plus généralement, l'objectif de notre travail est d'enrichir le système de planification de texte existant avec des informations communicatives afin de faciliter la construction d'un discours cohérent.

Afin d'intégrer des informations communicatives dans un système de génération de textes, la première étape sera de définir les outils théoriques, dont les relations de discours définies par la *Segmented Discourse Representation Theory* et des notions relatives à la structure communicative, qui vont nous fournir les bases nécessaires à l'étude de la structure communicative du discours. La deuxième étape consistera à mettre au jour les contraintes communicatives qui régissent la cohérence d'un discours. Ces contraintes seront testées lors de la dernière étape.

Ce mémoire s'articule selon trois grandes parties. Dans la première partie, nous établirons les cadres théoriques dans lesquels notre travail se situe. En particulier, nous présenterons la *Segmented Discourse Representation Theory*, théorie de la représentation du discours qui intègre des relations afin de rendre compte de la structure du discours. Ensuite, nous exposerons les notions relatives à la structure communicative de la phrase qui pourront nous guider pour atteindre un de nos objectifs : privilégier une structure communicative phrastique en fonction de la relation de discours en jeu entre deux phrases.

Dans la deuxième partie de ce mémoire, nous détaillerons le processus de génération de textes et expliciterons l'importance du module de planification de texte. Puis, nous décrirons comment des informations communicatives (comme le topique ou la partition thème~rhème) peuvent permettre la génération de textes (plus) cohérents. Nous présenterons la démarche que nous proposons pour la génération d'un texte : de la forme logique à un plan de document, et d'un plan de document à la structure communicative des phrases.

Enfin, dans la troisième partie, nous illustrerons la démarche décrite dans la deuxième partie avec deux segments discursifs. Nous tenterons de démontrer concrètement en quoi des informations communicatives permettent de produire un texte plus cohérent.

1. Notions préliminaires et cadres théoriques

Les phénomènes discursifs de la langue ont fait l'objet de nombreuses recherches en linguistique théorique et descriptive. Notre travail ne vise pas à faire l'état de ces recherches mais à proposer un nouvel axe d'étude. En effet, nous nous plaçons dans une optique particulière qui est le traitement automatique de la langue, et, plus particulièrement, la génération de textes qui requiert une modélisation du discours et des phrases qui le constituent.

C'est pour cette raison que nous nous fondons dans un premier temps sur la *Segmented Discourse Representation Theory* (abrégée en SDRT). Il s'agit d'une théorie qui propose de représenter formellement la sémantique du discours en articulant les segments phrastiques avec des relations rhétoriques, dites relations de discours.

Nous verrons que dans une perspective de génération, la formalisation de la structure discursive requiert des informations supplémentaires sur la structure communicative des phrases qui composent le discours. Nous consacrerons donc la deuxième partie à la présentation de quelques notions sur la structure communicative.

1.1. Structure et relations de discours

Pour permettre une représentation formelle de la sémantique des énoncés et de la structure du discours, Nicholas Asher a proposé la SDRT dès 1993 (Asher, 1993). Cette théorie se fonde sur la *Discourse Representation Theory*, théorie de représentation du discours (partie 1.1.1.) et sur la notion d'objet abstrait du discours (partie 1.1.2.). Nous verrons également que la SDRT propose des relations de discours qui permettent de relier les représentations des énoncés afin de mettre au jour la structure du discours (partie 1.1.3.).

1.1.1. La *Discourse Representation Theory*

La *Discourse Representation Theory* (abrégée en DRT) est une théorie de la représentation du discours présentée par Hans Kamp et par Uwe Reyle en 1993. Le principe de cette théorie est la formalisation du discours afin de rendre compte de phénomènes tels que la référence temporelle ou encore la résolution d'anaphore.

Le discours est représenté dans une « boîte » appelée DRS pour *discourse representation structure*. La construction d'une DRS se fait en fonction des représentations des phrases précédentes. Par exemple, si nous prenons le discours (1) suivant :

- (1) a. Jean a offert une rose à Marie.
- b. Elle était très contente.

La représentation de l'énoncé (1a) va être la DRS¹ :

x, y, z, e1
x = Jean y = Marie rose (z)
e1-offrir (x, z, y)

¹ Les DRSs et SDRSs proposées dans cette partie sont très simplifiées afin de faciliter les définitions.

Une DRS est composée de deux ensembles :

- un univers (dans la partie supérieure de la DRS) qui comprend l'ensemble des référents (entités ou événements) présents dans le discours ;
- des conditions sur les éléments de l'univers (dans la partie inférieure de la DRS).

Une condition peut correspondre à une égalité entre deux référents, à un prédicat, ou encore à une négation de DRS. La notation $x = Jean$ signifie que le référent 'Jean' est connu, défini, alors que la notation $rose(z)$ signifie que le référent est indéfini².

Les événements sont réifiés (suivant l'approche de D. Davidson), c'est-à-dire qu'ils correspondent à des référents de discours, ils apparaissent donc dans l'univers et peuvent être un argument dans un prédicat ou faire partie d'une égalité (dans les cas de co-référence événementielle où $e_i = e_j$).

Pour représenter l'énoncé (1b), nous avons la DRS (temporaire) suivante :

w, e2
w = ? e2-contente (w)

Dans cette DRS, la variable w correspond au pronom personnel *elle* dont nous ne connaissons pas l'antécédent. Si nous interprétons cette DRS au vue de la DRS précédente, il est possible de résoudre l'anaphore $w = ?$. Voici donc la représentation en DRT du discours (1) :

x, y, z, e1, w, e2
x = Jean y = Marie rose (z) e1-offrir (x, z, y) w = y e2-contente (w)

Ici, l'anaphore est résolue par le biais de l'équation $w = y$.

Une DRS peut contenir des informations temporelles sur les événements. En effet, comme les événements sont considérés comme des référents de discours, il est possible d'établir des relations temporelles entre eux, comme dans la nouvelle DRS du discours (1) :

x, y, z, e1, w, e2
x = Jean y = Marie rose (z) e1-offrir (x, z, y) e1 < now w = y e2-contente (w) e1 < e2

² D'après le séminaire de DEA sur la DRT donné par Francis Corblin (année 2001-2002).

La DRS comprend au final des informations de co-référence (avec, par exemple, l'équation anaphorique $w = y$) et des informations comme $e1 < e2$ qui nous indiquent la structure temporelle du discours.

La représentation du discours en une DRS rend compte de nombreux phénomènes discursifs (comme l'anaphore, par exemple) mais elle ne permet pas la référence à un segment phrastique entier. C'est ce sur quoi N. Asher s'est penché et que nous allons présenter maintenant.

1.1.2. Les objets abstraits du discours

N. Asher s'est rendu compte que dans un discours comme :

- (2) a. Jean a offert une rose à Marie.
b. Cela lui a fait plaisir.

le pronom démonstratif *cela* fait référence à l'événement décrit dans (2a). La DRT permet la résolution d'anaphore dont la référence est un groupe nominal mais elle ne tient pas compte de la référence à ce que N. Asher nomme les objets abstraits du discours i.e. des segments discursifs décrivant des événements. Il distingue les éventualités, les faits et les propositions.

- Les éventualités sont des expressions qui dénotent la réalité (elles ne sont ni vraies, ni fausses). Il peut s'agir de phrases simples ou de noms déverbaux. C'est ce que nous avons représenté par e dans les différentes DRSs reproduites plus haut.

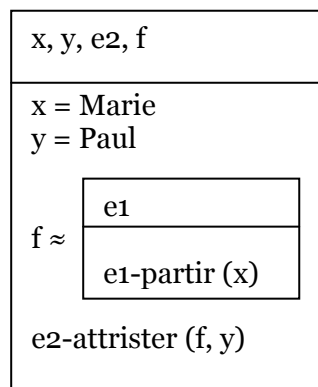
- Les faits sont des expressions intermédiaires entre les éventualités et les propositions. Il est possible de les enchâsser dans une expression comme *le fait que*. Par exemple, le discours :

- (3) a. Marie est partie.
b. Paul est triste.

peut être paraphrasé par :

- (4) Le fait que Marie soit partie attriste Paul.

et la représentation en DRT du discours (3) va être :



Un fait est caractérisé par une sous-DRS dans la DRS principale : il a des conditions propres, ce qui signifie que Paul peut être attristé par un autre fait (avec d'autres conditions).

- Les propositions ont une valeur de vérité par opposition aux éventualités qui ne sont ni vraies ni fausses.

Par exemple, si ‘Marie a réussi son examen’ est une proposition, on pourra dire :

(5) Il est vrai que Marie a réussi son examen.

mais on ne pourra pas dire :

(6) # La réussite de Marie à son examen est vraie.

où ‘la réussite de Marie à son examen’ correspond à une éventualité.

Un autre test permettant de distinguer une proposition d’une éventualité consiste à faire une prédication avec des verbes cognitifs :

(7) a. Paul sait que Marie a réussi son examen.

b. Paul demande si Marie a réussi son examen.

c. # Paul demande la réussite de Marie à son examen.

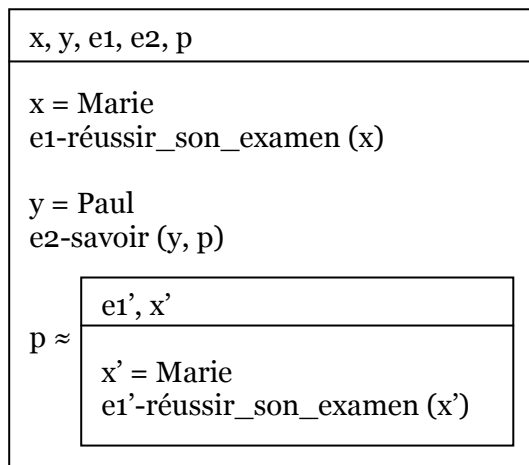
L’énoncé (7c) n’est pas acceptable car ‘la réussite de Marie à son examen’ est un groupe nominal qui n’exprime pas la proposition ‘Marie a réussi son examen’. Les propositions correspondent très souvent à des phrases simples et non à des groupes nominaux.

On peut représenter le discours (8) :

(8) a. Marie a réussi son examen.

b. Paul le sait.

par la DRS suivante :



1.1.3. Les relations de discours

Nous avons vu qu’il était possible de faire référence à des segments discursifs dans un texte. Afin de représenter ce phénomène, N. Asher propose de construire des SDRSs (pour *segmented discourse representation structures*) : il s’agit en fait de mettre des relations discursives entre des DRSs.

Formellement, une SDRS est composée d’un univers U et de conditions Con :

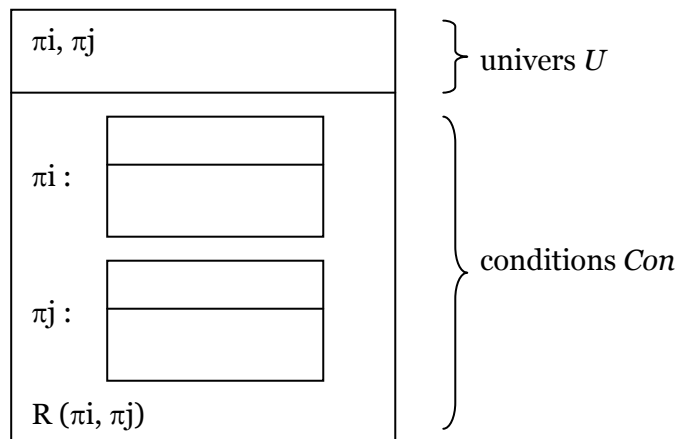
SDRS = $\langle U, Con \rangle$

L’univers U est l’ensemble des étiquettes pour des sous-DRSs (ou sous-SDRSs). Une sous-(S)DRS est étiquetée π_i et elle représente toutes les informations relatives à une proposition.

Les conditions Con représentent les différentes paires $\pi_i : K$, où K est une sous-(S)DRS.

$R(\pi_i, \pi_j)$ représente la relation de discours entre π_i et π_j .

Schématiquement, une SDRS ressemble à :



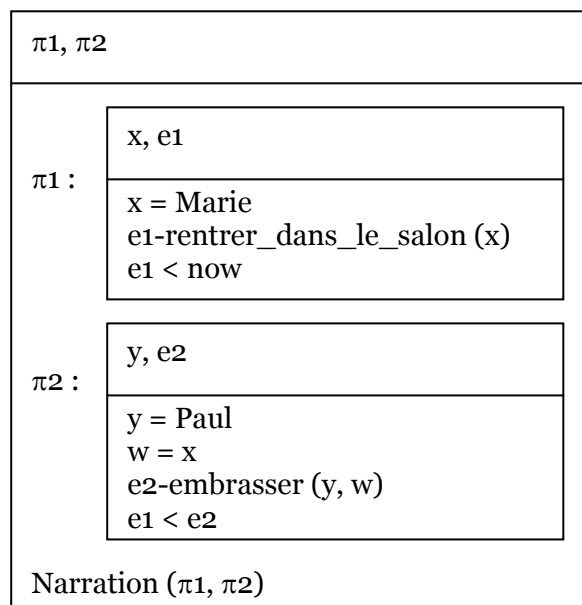
Si π_1 étiquète la DRS qui représente l'événement e_1 et si π_2 étiquète la DRS qui représente l'événement e_2 , alors les relations de discours en jeu entre les événements e_1 et e_2 (décrits dans les deux phrases constituant le discours) peuvent être les suivantes :

a. La relation Narration

La relation Narration est caractérisée par la succession temporelle des événements décrits dans le discours. L'événement e_2 est successif à e_1 ($e_1 < e_2$). Par exemple, dans le discours (9), l'événement décrit dans (9b) se produit après l'événement décrit dans (9a) :

- (9) a. Marie entra dans le salon.
- b. Paul alla l'embrasser.

Une manière de représenter ce discours est la SDRS suivante :



La relation Narration est justifiée car les événements sont successifs et cela se traduit dans la SDRS par $e_1 < e_2$ dans la DRS π_2 . En génération, l'ordre des DRSs est très important car il donne lieu à l'ordre des propositions dans le discours à produire.

Par ailleurs, N. Asher introduit un topique de narration qui permet de traduire la cohérence thématique du discours : il s'agit de retrouver dans la SDRS impliquant une relation Narration un élément commun aux deux événements représentés. Nous reviendrons sur la notion de topique dans la partie 1.2.2..

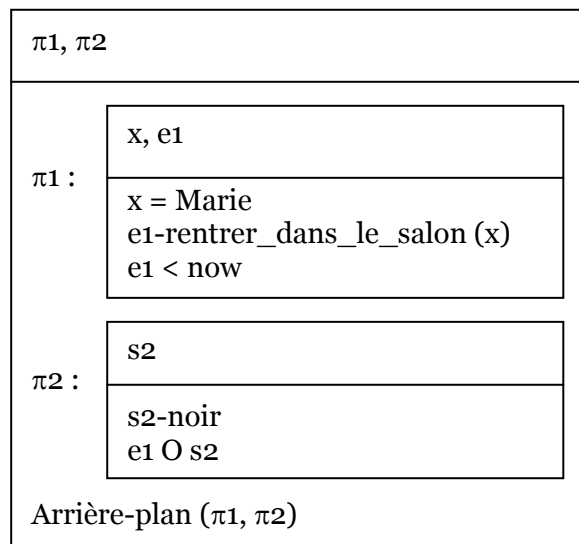
b. La relation Arrière-plan

La relation Arrière-plan se caractérise par le fait que l'un des deux événements décrits dans le discours représente le contexte, les circonstances de l'autre événement. Dans le discours (10) :

- (10) a. Marie entra dans le salon.
b. Il faisait noir.

l'état décrit dans (10b) est simultanément à l'événement e_1 décrit dans (10a). On parle alors d'*overlapping*, de chevauchement entre e_1 et s_2 décrit dans (10b) : $e_1 \text{ O } s_2$.

La SDRS sous-jacente au discours (10) peut être la suivante :



Nous pouvons remarquer que la relation Arrière-plan ne reproduit pas l'ordre textuel des segments (π_1 n'est pas l'arrière-plan de π_2) comme cela est le cas pour la relation Narration (π_1 est avant π_2). Dans un contexte de génération, cela peut être gênant et c'est pour cette raison que Laurent Roussarie propose dans (Roussarie, 2000) deux relations Arrière-plan selon que l'arrière-plan soit avant ou après l'événement d'avant-plan. L. Roussarie introduit également un pseudo-topique pour conserver une cohérence discursive mais nous n'irons pas dans les détails ici.

c. La relation Parallèle

La relation Parallèle permet de décrire deux événements simultanés : on peut comparer cela à une relation Narration sans succession ou à une relation Arrière-plan avec deux événements dynamiques (et non avec un événement statique). La relation Parallèle peut être illustrée avec le discours (11) :

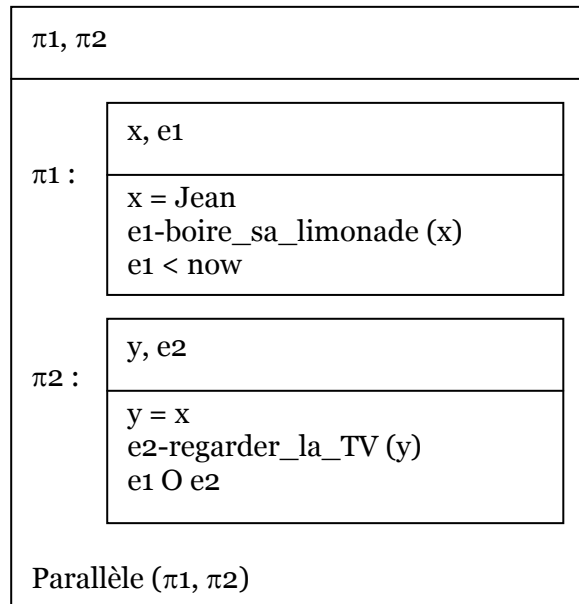
- (11) a. Jean buvait sa limonade.
b. Il regardait aussi la télévision.

Le discours (11) peut être paraphrasé par :

(12) Jean buvait sa limonade en regardant la télévision.

Les événements décrits dans les deux phrases sont simultanés. En règle générale, il y a une marque de présupposition comme l’adverbe *aussi* ou *également* dans la deuxième phrase (ici, dans (11b)) pour signaler l’occurrence d’un premier événement (ici, dans (11a)).

Pour représenter le discours (11), nous pouvons construire la SDRS suivante :



Nous ne marquons pas ici que l’événement e_1 est présupposé quand e_2 est posé dans la DRS π_2 ³.

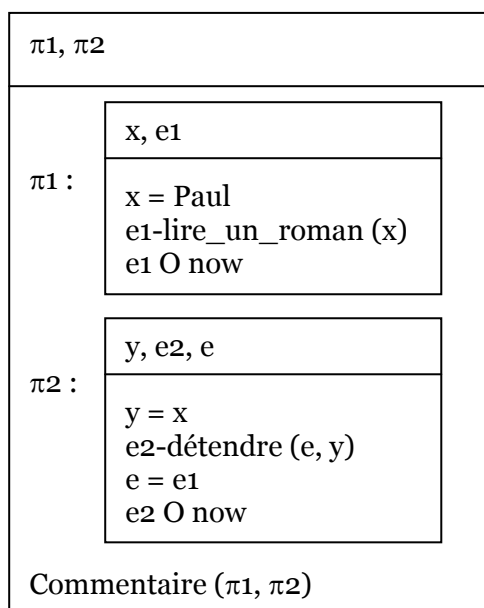
d. La relation Commentaire

La relation Commentaire est instanciée entre deux DRSs lorsque la deuxième phrase est une glose, un commentaire sur la première, comme dans le discours :

(13) a. Paul lit un roman.
b. Cela le détend.

Il s’agit ici d’un discours que la DRT n’aurait pas pu traiter car le pronom démonstratif *cela* dans (13b) reprend l’événement décrit dans (13a). La relation Commentaire permet la représentation formelle de ce discours comme ce qui suit :

³ Pour plus détails sur la présupposition et notamment sur l’adverbe *aussi*, nous invitons le lecteur à voir (Raynal, 2002).



Dans la DRS π_2 , il existe un événement ' e qui détend l'entité ' y ' : cet événement correspond à e_1 (de π_1). Il y a co-référence événementielle entre les deux phrases : e_1 est repris sous la forme du pronom démonstratif *cela* et il aurait pu également être repris sous la forme du substantif déverbal *la lecture*.

Nous verrons que la relation Commentaire permet d'exprimer des relations causales complexes dans la partie 2.1.2..

e. La relation Élaboration

La relation Élaboration permet de donner des informations supplémentaires sur le premier événement décrit dans le discours. Plus précisément, l'événement e_2 élabore, explicite, est une partie de l'événement e_1 . Soit le discours :

- (14) a. Paul est rentré à Paris.
b. Il est passé par Toulouse.

L'événement e_2 décrit dans (14b) est un sous-événement de e_1 décrit dans (14a). Il apporte des informations sur e_1 .

Il existe selon nous un autre type de relation Élaboration. Elle concerne l'apport d'informations sur une entité et non sur un événement comme nous venons de le voir. Dans le discours (15) suivant :

- (15) a. Paul a acheté un roman.
b. C'est le dernier David Lodge.

ce n'est pas l'événement e_1 qui est élaboré mais seulement l'entité 'un roman'. Dans (15b), on donne des informations sur 'un roman'. Nous appellerons ce type de relation Élaboration objet (vs. Élaboration événement, comme dans le discours (14)).

f. La relation Particularisation

La relation Particularisation est un cas particulier de relation Élaboration car elle permet l'explicitation ET LA REPRISE du premier événement dans la deuxième partie du discours. Il s'agit d'une co-référence événementielle comme dans le discours :

- (16) a. Paul lit un roman.
b. Il lit le dernier David Lodge.

Le même événement est décrit dans les deux phrases ('Paul lit un roman') et il y a un ajout d'informations dans (16b). L'apport d'informations correspond à la spécification d'un élément de (16a). Il peut y avoir également adjonction d'informations :

- (17) a. Paul lit un roman.
b. Il lit un roman dans le salon.

L'ajout d'informations correspond ici au circonstant *dans le salon* dans (17b). Et nous pouvons aussi avoir un discours particularisant par spécification et par adjonction :

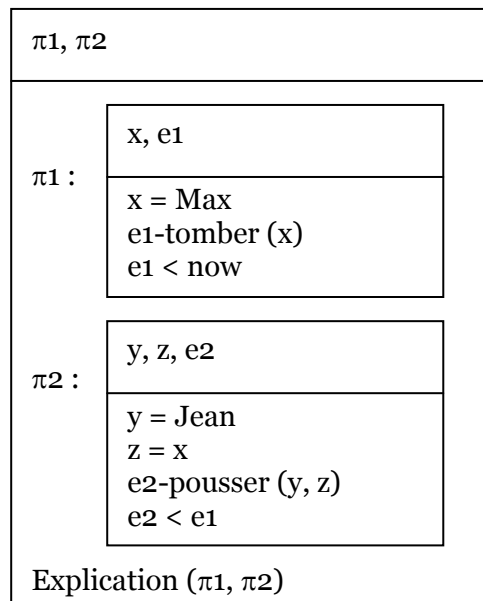
- (18) a. Paul lit un roman.
b. Il lit le dernier David Lodge dans le salon.

g. La relation Explication

La relation Explication est définie si le deuxième événement cause (ou explique) le premier, comme dans le discours :

- (19) a. Max est tombé.
b. Jean l'a poussé.

Dans le discours (19), e2 (décrit dans (19b)) cause e1 (décrit dans (19a)). Voici une représentation du discours (19) :

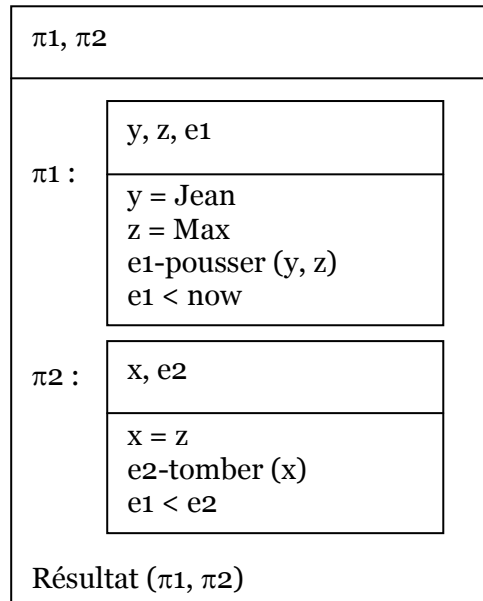


h. La relation Résultat

Si le premier événement cause le deuxième, alors il y a une relation Résultat. Le discours :

- (20) a. Jean a poussé Max.
b. Il est tombé.

illustre la relation Résultat car e1 (dans (20a)) cause e2 (dans (20b)). La SDRS sous-jacente au discours (20) peut être :



En génération, les relations Explication et Résultat jouent un rôle très important dans la planification de document car ce sont elles dont dépend l'ordre des constituants phrastiques dans le discours. Nous reviendrons sur ce point dans la partie 2.1.3..

Conclusion

L'adaptation de la SDRT à la génération de textes (depuis (Roussarie, 2000)) permet d'organiser les propositions (au sens sémantique) de manière à produire un texte le plus cohérent possible. Mais nous avons vu dans cette première partie que les informations contenues dans une (S)DRS peuvent ne pas suffire. Certaines informations, comme le topique, semblent nécessaires à la représentation et (surtout) à la génération d'un discours cohérent. C'est pour cela que nous allons présenter maintenant quelques notions relatives à la structure communicative, comme le topique ou la partition thème~rhème, qu'il serait pertinent d'intégrer dans un formalisme de représentation de discours.

1.2. Structure communicative de la phrase

Toute phrase possède une structure communicative, informationnelle, et cette structure permet au locuteur de traduire le sens de la phrase en un message tel qu'il a l'intention de le faire passer. La structure communicative met en jeu différentes valeurs (le thème et le rhème notamment) dont l'organisation peut varier pour produire des énoncés distincts d'un point de vue informationnel. Nous allons nous attacher ici à définir quelques unes de ces valeurs communicatives : nous décrirons en particulier la partition thème~rhème (partie 1.2.1.), les notions de topique (partie 1.2.2.), de focus rhématique (partie 1.2.3.) et de spécifieur communicatif (partie 1.2.4.). Nous verrons également une définition de l'opposition donné~nouveau (partie 1.2.5.) et de la focalisation (partie 1.2.6.).

1.2.1. La partition thème~rhème

Au niveau informationnel, la caractéristique de tout message est que l'on dit quelque chose (le rhème) à propos de quelque chose (le thème) :

Le rhème (ou focus, *comment*) est

- ce qui est communiqué
- un élément obligatoire
- le focus informationnel de la phrase
- une information nouvelle quant au propos du message, s'il y en a.

Le thème est

- l'entité au sujet de laquelle la proposition (plus précisément le rhème) est
- le propos du message, ce dont on parle
- un élément facultatif : on peut communiquer quelque chose sans propos particulier.

L'organisation de ces deux valeurs va permettre au locuteur de présenter le message comme il le souhaite⁴ :

- Les phrases thème-rhème ou *topic-comment (predicate-focus structure)* permettent de faire une prédication (i.e. le rhème -en majuscules-) sur l'élément dont on parle (i.e. le thème -en minuscules-) :

- Q⁵ : que fait Benoît ?
- (21) Benoît JOUE DE LA GUITARE.
- ou
- (22) Benoît, il JOUE DE LA GUITARE.

Dans la phrase (22), le thème 'Benoît' est mis en valeur par le biais d'une construction appelée dislocation gauche. Nous décrirons ce phénomène dans la partie 1.2.6..

- Les phrases d'identification (*argument-focus structures*) permettent de mettre au jour un nouvel argument :

- Q : qui écoute de la musique ?
- (23) NATHALIE écoute de la musique.
- ou
- (24) C'est NATHALIE qui écoute de la musique.

⁴ Nous utilisons la terminologie employée par Igor Mel'čuk et par Knud Lambrecht.

⁵ 'Q' représente la question sous-jacente à l'énoncé pour donner une idée du contexte dans lequel peut apparaître la phrase. Il ne s'agit en rien d'une paire question-réponse attestée.

- Q : qu'étudie Jean ?
 (25) Jean étudie LA PHYSIQUE.
 ou
 (26) C'est LA PHYSIQUE que Jean étudie.

Dans les phrases (24) et (26), le nouvel argument (le rhème) est mis en valeur grâce à une construction clivée en « c'est...que ».

Il peut y avoir des phrases sans thème : elles sont appelées phrases rhématiques (ou thétiqes). Elles sont de trois types :

- Les phrases événementielles, *event-reporting (sentence-focus structures)*, permettent d'introduire un nouvel événement impliquant un nouveau référent de discours :

- Q : que se passe-t-il ?
 (27) SANDRINE ARRIVE DEMAIN.

- Les phrases performatives représentent l'action qui y est dénotée :

- Q : que se passe-t-il ?
 (28) IL PLEUT.
 (29) JE VOUS DÉCLARE MARI ET FEMME.

- Les phrases de présentation (*topic-setting sentences*) permettent d'introduire une nouvelle entité, un nouveau référent de discours :

- Q : que se passe-t-il ?
 (30) Je voudrais te parler de MA MERE.

1.2.2. Le topique

Selon certains auteurs, dont Knud Lambrecht, certains éléments intra-phrastiques (comme un pronom anaphorique) et inter-phrastiques (comme un connecteur) ne peuvent pas être interprétés dans une phrase isolée mais seulement en discours. Plus précisément, la structure communicative de la phrase ne concerne pas l'organisation du discours mais l'organisation de la phrase dans un discours.

Certes, il existe une partition thème~rhème dans toute phrase : le locuteur dit quelque chose (i.e. le focus ou rhème) au sujet de quelque chose (i.e. le thème). Mais quand on étudie la structure informationnelle de la phrase en contexte, en relation avec d'autres phrases, il convient de distinguer le thème du topique :

Le thème (phrastique) est l'objet de la prédication dans une phrase :

- (31) Mon père LIT LE JOURNAL.

Dans la phrase (31), on fait une prédication (le rhème) sur un thème ('mon père').

Le topique (discursif) est l'objet du discours et correspond à un « hyper-thème » :

- (32) a. Mon père est parti pour Toulouse en train.
 b. Durant le trajet, certains écoutent de la musique.
 c. Mon père, il lit le journal.

Le discours (32) est un texte dont le topique est 'mon père' (le topique peut également être 'le voyage de mon père'). Il est question de 'mon père' dans tout le discours. Dans la phrase (32c), le thème phrastique est mis en valeur pour prendre une fonction de « cadre » pour référer au topique discursif. Dans ce cas, l'élément mis en valeur est appelé topique ou topique *scene-setting* par K. Lambrecht.

Le topique permet d'assurer une cohérence discursive à un texte. En effet, un texte doit traiter du même référent, il ne doit pas passer « du coq à l'âne » pour rester cohérent.

1.2.3. Le focus rhématique

Dans ce qui veut être communiqué par le locuteur, i.e. dans le rhème, il peut y avoir un élément plus saillant d'un point de vue informationnel : c'est ce qui est appelé le focus rhématique. Dans l'énoncé (33) :

- Q⁶ : que fait Claire ?
(33) Claire LIT UN MAGAZINE.

l'élément du rhème qui apporte le plus d'informations est 'un magazine' (par rapport à 'la lecture') : 'un magazine' est le focus rhématique.

Le focus rhématique permet d'introduire un nouveau référent sur lequel il sera possible de faire une prédication par la suite :

- Q : que fait Fabien ?
(33) a. Fabien REGARDE UN MATCH.
b. C'est France-Angleterre.

En discours, le focus rhématique est souvent repris en thème dans la phrase suivante, comme dans (33).

Dans des discours comme (34) :

- Q : que fait Fabien ?
(34) a. Fabien REGARDE UN MATCH.
b. Ça le détend.

ce n'est pas le focus rhématique qui est repris mais le rhème 'regarder un match' qui est repris en thème dans la phrase suivante sous la forme du pronom démonstratif *ça*. Nous reviendrons sur ces types d'enchaînements discursifs dans la partie 2.2.2..

1.2.4. Les spécifieurs communicatifs

Nous avons vu que le rhème constituait le cœur du message à communiquer et que des éléments comme le thème pouvaient s'y ajouter. C'est également le cas des spécifieurs communicatifs : ce sont des éléments de la structure communicative qui dépendent du rhème et qui ne sont pas obligatoires.

Il existe des spécifieurs communicatifs internes et externes :

⁶ Le focus rhématique est souligné.

- Le spécifieur interne fonctionne dans la phrase même :
 - le circonstant indique les circonstances objectives de la phrase :
 - le circonstant factuel indique les circonstances de l'événement décrit dans la phrase (le temps, l'espace, les actants, etc.)
ex : la nuit dernière, à Paris
 - le circonstant discursif indique les circonstances de l'énonciation
ex : pour votre information
 - le modal est une marque de la subjectivité du message :
 - le modal factuel indique l'attitude de locuteur par rapport à ce qu'il dit
ex : malheureusement, pour être honnête
ou par rapport au statut épistémologique du propos tenu
ex : bien sûr, sans aucun doute
 - le modal discursif indique la façon qu'a le locuteur de présenter ce qu'il dit
ex : pour résumer, en conclusion
- Le spécifieur externe fonctionne entre plusieurs phrases. C'est un connecteur qui exprime les relations de discours (ou rhétoriques) :
 - le connecteur factuel marque la relation entre deux événements
ex : Il est venu. *Après*, il est parti.
 - le connecteur discursif exprime la relation de discours entre deux phrases
ex : Il est parti. *Contrairement à ça*, elle est restée.

Nous avons explicité en détails les relations de discours dans la partie 1.1. consacrée à la SDRT, nous n'y revenons donc pas dessus.

1.2.5. L'opposition donné~nouveau

Nous tenons ici à bien distinguer :

- l' (les) élément(s) dont on parle = le thème
- ce qu'on dit de cet (ces) élément(s) = le rhème
- l' (les) élément(s) donné(s)
- l' (les) élément(s) nouveau(x)

car, s'il y a souvent amalgame entre thème et donné, et, entre rhème et nouveau, il existe des discours dans lesquels ces quatre valeurs communicatives correspondent à des éléments phrastiques distincts.

Un élément est dit donné lorsqu'il correspond à un référent connu, à une information ancienne, partagée des interlocuteurs. Il est un focus cognitif car il est activé dans l'esprit des interlocuteurs.

Un élément donné ne correspond pas toujours au thème. Dans le discours :

- Q : qui prépare le dîner ?
- (35) a. Fred est dans la cuisine.
b. C'est LUI qui prépare le dîner.

'Fred' est un élément donné (reprise anaphorique dans la phrase (35b)) et il s'agit du rhème : on apprend que c'est Fred qui prépare le dîner et pas quelqu'un d'autre.

Un élément donné peut

- faire référence à un élément introduit dans le contexte antérieur, dans la ou les phrases précédentes ;

- faire référence à un élément présupposé connu des interlocuteurs. On trouve souvent cela en début de discours.

Au niveau lexical, un élément donné peut être spécifié par un adjectif démonstratif (*ce, cette, ces*), par un article défini (*le, la*), ou être un pronom personnel (*il, elle, je*), ou encore être un pronom démonstratif (*celui-ci, celle-là*).

Un élément est dit nouveau lorsqu'il correspond à un référent nouveau, à une information nouvelle. Un élément nouveau fait très souvent partie du rhème.

1.2.6. La focalisation

Un élément de la phrase peut être focalisé, mis en valeur en fonction de ce que le locuteur souhaite mettre en avant. Il est marqué indépendamment des organisations thème~rhème et donné~nouveau. Il ne faut pas confondre ce sur quoi le locuteur veut attirer l'attention et ce que le locuteur souhaite dire d'un point de vue strictement informationnel (thème~rhème, donné~nouveau). La focalisation est une « couche » supplémentaire (et facultative) sur l'information : le thème aussi bien que le rhème peut être focalisé.

En français, la focalisation d'un élément⁷ est très marquée syntaxiquement :

- les constructions clivées permettent de focaliser le rhème :

Q : qui court tous les matins ?
(36) C'est CAROLINE qui court tous les matins.

Q : qu'a mangé Richard à midi ?
(37) C'est UNE PIZZA que Richard a mangé à midi.

- les constructions pseudo-clivées permettent de focaliser le thème :

Q : qui a acheté une voiture ?
(38) Celui qui a acheté une voiture est THIERRY.

- la dislocation gauche, le *fronting*, permet de focaliser le thème :

Q : que fait Xavier ?
(39) Xavier, il RÉVISE SES EXAMENS.

Q : qu'y a-t-il à propos de cet appartement ?
(40) Cet appartement, ÉMILIE l' A HABITÉ.

Q : qu'y a-t-il à propos d'Émilie et de cet appartement ?
(41) Émilie, cet appartement, elle l' A HABITÉ.

- la dislocation droite, le *postponing*, permet également de focaliser le thème :

Q : que fait Céline ?
(42) Elle RANGE SA CHAMBRE, Céline.

Q : qu'y a-t-il à propos de la cuisine ?
(43) ÉDITH l' A REPEINTE, la cuisine.

⁷ L'élément focalisé est souligné.

Le thème focalisé est repris par un pronom anaphorique (dans les cas de *fronting*) ou cataphorique (dans les cas de *postponing*) qui dépend du verbe principal dans le rhème.

Conclusion

Nous avons défini ici quelques notions relatives à la structure communicative de la phrase envisagée en discours. La pertinence de ces concepts apparaît dès lors que nous nous plaçons dans une perspective de génération de textes et que nous considérons que la structure communicative de la phrase est déduite en contexte et que ce n'est pas la somme des interprétations des phrases successives qui constitue l'interprétation du discours. Même du point de vue du locuteur, le discours n'est pas une suite aléatoire de phrases : la façon dont va être exprimée une phrase dépend de la façon dont a été exprimée la phrase juste avant. D'où l'importance d'informations communicatives, comme celles décrites ici, en génération de textes.

Conclusion première partie

Après cette première partie, nous pouvons conclure à ceci : aucune phrase ne peut être interprétée ou énoncée indépendamment du contexte discursif dans lequel elle se trouve. En effet, nous avons vu que les énoncés entretiennent entre eux des relations qui permettent de les organiser en discours et que des informations communicatives (comme le topique ou la partition thème~rhème) pouvaient contribuer à la cohérence d'un discours. La cohérence discursive va donc dépendre de deux choses : des relations de discours en jeu et de la structure communicative des phrases. Ces deux paramètres sont guidés par le choix du topique du discours. En effet, le topique impose un ordre particulier des propositions (et impose donc des relations de discours) et cet ordre contraint l'organisation communicative interne des propositions.

La cohérence du discours va être tributaire de la structure communicative à deux niveaux : au niveau de l'organisation inter-propositionnelle i.e. de l'organisation communicative des propositions entre elles, et au niveau de l'organisation intra-propositionnelle i.e. de la structure communicative de chaque proposition. Nous allons voir dans la deuxième partie comment ces deux niveaux peuvent s'articuler afin de permettre la génération de textes cohérents.

2. Intégration d'informations communicatives dans un système de génération de textes

Rappelons que l'objectif du présent travail est d'intégrer des informations communicatives dans un système de génération de textes. Nous avons exposé dans la première partie les différents outils théoriques dont nous allons maintenant nous servir pour atteindre notre objectif : générer un texte le plus cohérent possible grâce à des informations sur la structure communicative.

Nous allons tout d'abord présenter le contexte général dans lequel s'inscrit notre étude, en particulier, la structuration du discours dans un système de génération de textes. Après avoir décrit les différentes tâches d'un système de génération, nous verrons en quoi la SDRT apporte des solutions pour la planification du discours à générer. Enfin, nous analyserons un cas particulier de relation causale qui soulève le problème du choix du topique.

Ensuite, nous passerons au processus de génération proprement dit (d'un point de vue théorique). Nous verrons comment le topique peut guider le choix d'un plan de document et comment un plan de document impose une structure communicative particulière pour chaque proposition à générer.

2.1. Structure du discours en génération de textes

La génération automatique de textes (abrégée en GAT) a pour objectif de produire à partir de données abstraites des énoncés en langage naturel dans une langue bien formée. Un générateur doit pouvoir transformer des données conceptuelles (représentant la sémantique du message à produire) en un discours cohérent.

Nous détaillerons d'abord les différents tâches d'un système de GAT (partie 2.1.1.) puis, nous expliciterons le fonctionnement du module de structuration de document et l'apport de la SDRT dans cette tâche (partie 2.2.2.).

2.1.1. Tâches d'un système de génération

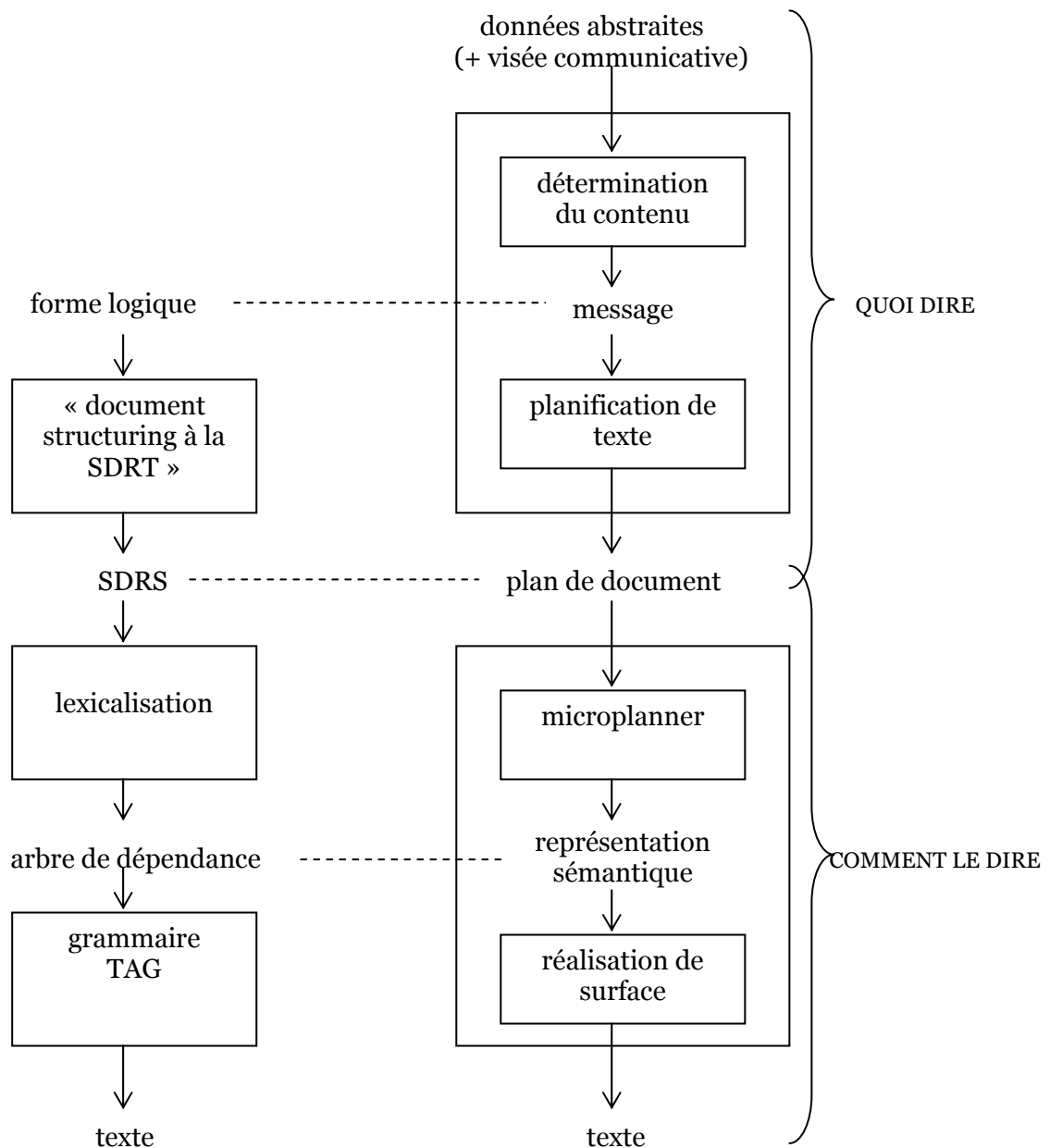
Un générateur de textes est composé de plusieurs niveaux correspondant à des tâches précises. Chaque niveau transforme les données qu'il reçoit en entrée. Les tâches peuvent être regroupées dans deux modules : le « Quoi-dire » et le « Comment-le-dire ».

La première tâche consiste à déterminer le contenu du message que l'on veut générer. Cela peut se présenter sous la forme d'une forme logique (logique des prédicats par exemple). Cette forme logique peut simplement représenter le contenu sémantique du texte à générer mais elle peut aussi être enrichie d'informations communicatives (« visée communicative »).

Ensuite, il s'agit d'organiser la forme logique dans un texte : c'est la planification de document. Nous arrivons là à la frontière entre le « Quoi-dire » et le « Comment-le-dire » car si le module de structuration de document organise les propositions dans un texte cela a des conséquences directes sur la réalisation syntaxique finale. Lors de la construction du plan de document, on est déjà « en langue ».

Le « Comment-le-dire » comprend, entre autres, la tâche de planification syntaxique dont le but est de privilégier une réalisation syntaxique par rapport à une autre pour exprimer une structure sémantique. La tâche suivante consiste à lexicaliser les éléments de la structure sémantique. Enfin, les tâches d'ajustement morphologique et typographique sont les touches finales au processus de génération.

Le schéma qui suit représente l'architecture générale d'un générateur de textes. Le processus de génération représenté dans la partie de gauche correspond au cadre de notre travail.



Le module de structuration du document correspond à une étape essentielle dans le processus de génération. Nous allons détaillé ce passage de la forme logique à un plan de document dans la partie qui suit.

2.1.2. La structuration de document

Le module de structuration de document (*document structuring*) prend en entrée une forme logique (sans informations communicatives) qui représente la sémantique du texte à générer, soit par exemple la forme logique (1) :

$$(1) \exists e1, s2, x, y \left(\begin{array}{l} e1\text{-partir}(x) \wedge \\ s2\text{-triste}(y) \wedge \\ \text{cause}(e1, s2) \wedge \end{array} \begin{array}{l} x = \text{Marie} \wedge \\ y = \text{Paul} \wedge \\ e1 < \text{now} \wedge \\ s2 \text{ O now} \end{array} \right)$$

Cette forme logique inclut un événement ($e1$) et un état ($s2$), des entités (x et y) et une relation conceptuelle entre l'événement $e1$ et l'état $s2$ (*cause* ($e1, s2$)).

L'objectif pour le module de planification de texte est de permettre la génération de toutes les paraphrases possibles à partir de la forme logique (1). À la fin du processus de génération, nous voudrions obtenir, par exemple, les paraphrases suivantes :

- (2) Paul est triste parce que Marie est partie.
- (3) Marie est partie donc Paul est triste.
- (4) a. Marie est partie.
b. Paul est triste à cause de cela.
- (5) a. Marie est partie.
b. Cela attriste Paul.
- (6) a. Marie est partie.
b. C'est la raison pour laquelle Paul est triste.
- (7) a. Paul est triste.
b. La raison de sa tristesse est que Marie est partie.

Rappelons que tous ces discours ont le même contenu sémantique (i.e. ils sont issus de la même forme logique) mais qu'ils diffèrent de par leurs structures communicatives.

Afin de produire les discours (2) à (7) ci-dessus, il faut faire des plans de document, en l'occurrence, il faut construire des SDRSs. Selon la relation de discours en jeu, la relation causale entre l'événement $e1$ et l'état $s2$ va être exprimée en langue de différentes manières.

Une relation de discours sous-jacente va donner lieu à un connecteur phrastique (comme pour les discours (2) et (3)), un élément non réifié i.e. *cause* ($e1, s2$) donnera lieu à un modifieur (comme à *cause de* dans le discours (4)) et un élément réifié i.e. *f-cause* ($e1, s2$) donnera lieu à un nom ou un verbe, comme dans les discours (5), (6) et (7).

Nous analyserons en détails un cas particulier de relation causale réifiée qui donne lieu au nom *raison*, comme dans les discours (6) et (7). Mais pour le moment, voyons comment les discours (2) à (5) peuvent être générés.

La première étape consiste à construire une DRS à partir de la forme logique (1). Cela consiste à rassembler tous les éléments de la forme logique dans une seule DRS comme la suivante :

(8)	$e1, s2, x, y$ $x = \text{Marie}$ $y = \text{Paul}$ $e1\text{-partir}(x)$ $s2\text{-triste}(y)$ $\text{cause}(e1, s2)$ $e1 < \text{now}$ $s2 \text{ O now}$
-----	--

Pour chaque discours à générer, on va avoir différentes SDRSs sous-jacentes. Ces SDRSs sont issues de la DRS (8), équivalente à la forme logique (1).

La SDRS qui va permettre la génération du discours (2) peut être la suivante :

(9)	$\pi1, \pi2$ <table border="1"> <tr> <td>$\pi1 :$</td> <td> $y, s2$ $y = \text{Paul}$ $s2\text{-triste}(y)$ $s2 \text{ O now}$ </td> </tr> <tr> <td>$\pi2 :$</td> <td> $x, e1$ $x = \text{Marie}$ $e1\text{-partir}(x)$ $e1 < \text{now}$ </td> </tr> </table> Explication ($\pi1, \pi2$)	$\pi1 :$	$y, s2$ $y = \text{Paul}$ $s2\text{-triste}(y)$ $s2 \text{ O now}$	$\pi2 :$	$x, e1$ $x = \text{Marie}$ $e1\text{-partir}(x)$ $e1 < \text{now}$
$\pi1 :$	$y, s2$ $y = \text{Paul}$ $s2\text{-triste}(y)$ $s2 \text{ O now}$				
$\pi2 :$	$x, e1$ $x = \text{Marie}$ $e1\text{-partir}(x)$ $e1 < \text{now}$				

Pour construire cette SDRS, nous avons regroupé tout ce qui concerne l'événement $e1$ dans une DRS et tout ce qui concerne l'état $s2$ dans une autre DRS.

La DRS étiquetée $\pi1$ représente l'état $s2$, soit la sémantique d'une phrase comme :

Paul est triste.

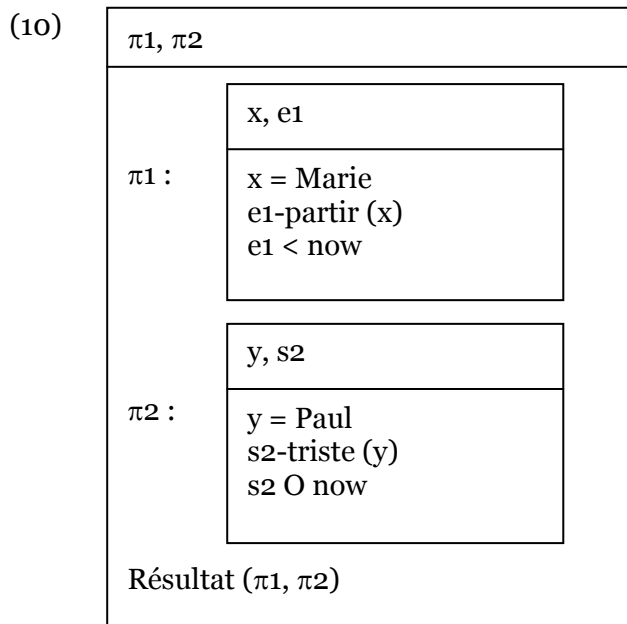
La DRS étiquetée $\pi2$ représente l'événement $e1$, soit la sémantique d'une phrase comme :

Marie est partie.

Entre les deux DRSs, la relation Explication traduit *cause* ($s2, e1$). Cette relation causale peut s'exprimer dans des connecteurs phrastiques tels que *parce que, du fait que, car*, etc. comme dans le discours (2) ou ne pas s'exprimer, comme dans le discours (2') :

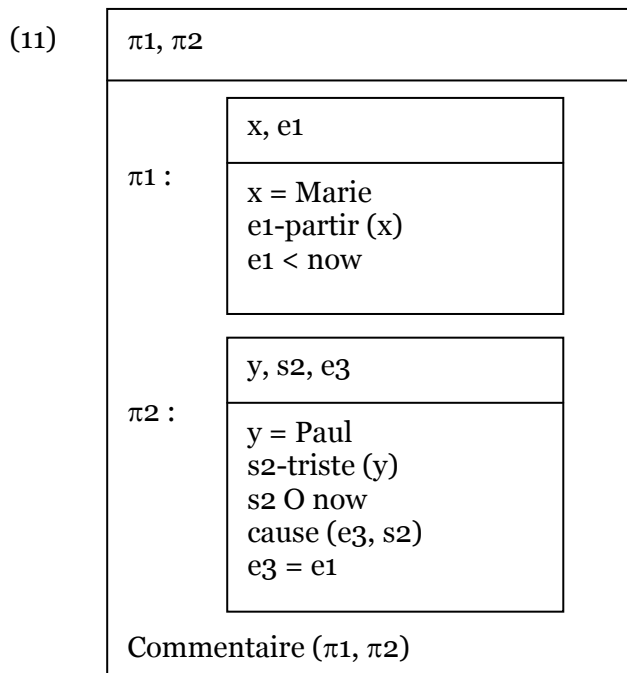
(2') Paul est triste. Marie est partie.

La SDRS sous-jacente au discours (3) peut être la SDRS (10) :



Ici, la relation causale est traduite par la relation Résultat. Cette relation causale peut être lexicalisée dans les connecteurs *donc*, *par conséquent*, etc. comme dans le discours (3) ou ne pas l'être.

La SDRS sous-jacente au discours (4) peut être la suivante :



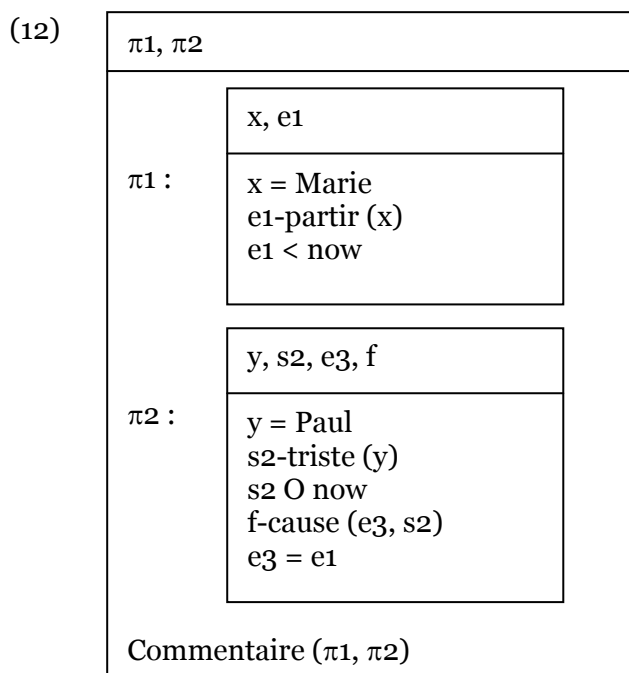
On retrouve tout ce qui concerne l'événement e1 dans la DRS π_1 et le reste des conditions est dans la DRS π_2 .

Le prédicat *cause* (e1, s2) traduisant la relation causale dans la DRS (8) ne peut pas apparaître tel quel dans la DRS π_2 : le référent de discours e1 est déjà présent dans la DRS π_1 . Un référent de discours (ici, e1) ne peut pas apparaître dans deux DRSs (i.e. on ne peut pas avoir e1 dans π_1 et *cause* (e1, s2) dans π_2). Il faut donc créer une nouvelle variable e3 dans π_2

pour référer à e_1 . La co-référence va se traduire par le biais d'une équation anaphorique, soit $e_3 = e_1$ dans π_2 . Grâce à cela, nous pouvons avoir le prédicat *cause* (e_3, s_2) dans π_2 . Le prédicat *cause* (e_3, s_2) -qui n'est pas réifié- permet de lexicaliser la relation causale entre e_3 et s_2 par la locution prépositionnelle à *cause de*, comme dans le discours (4).

La relation Commentaire est une relation qui n'ajoute aucune information sur les relations entre les événements. Comme toutes les conditions de la DRS (8) sont épuisées, la relation Commentaire convient parfaitement. Par ailleurs, la contrainte de co-référence imposée par cette relation est respectée avec $e_3 = e_1$, comme cela est souligné dans (Danlos, Gaiffe et Roussarie, 2001).

Avec la relation Commentaire, nous pouvons également adopter une approche davidsonienne : les événements sont considérés comme des référents du discours, des entités. Il peuvent donc apparaître dans l'univers du discours et être des arguments dans certains prédicats. Cette alternative permet de construire la SDRS sous-jacente aux discours (5) et (6) :



Le prédicat *f-cause* (e_1, e_2), dans π_2 , signifie que l'événement, ou le fait, f est considéré comme un référent de discours et pourra se lexicaliser par un nom ou un verbe. La réification va donc permettre de produire des textes différents par le biais d'un lexique plus large. Elle permet de traduire le lien de cause à effet entre e_3 et s_2 par le verbe *causer* ou *provoquer* :

Marie est partie. Cela a provoqué de la tristesse chez Paul.
ou par le nom *cause* :

Marie est partie. C'est la cause de l'état de tristesse de Paul.
Lors de la phase des choix lexicaux dans le processus de génération, nous pourrions obtenir les discours (5) et (6), par exemple.

Par rapport à la SDRS (12) proposée ci-dessus, le discours (6) fonctionne parce que la cause est exprimée dans la première phrase et qu'elle est reprise dans la deuxième phrase sous la forme de l'expression *c'est la raison pour laquelle*. En revanche, si nous voulons exprimer l'effet puis la cause, il va falloir adopter une autre approche car pour produire des discours comme (7), que nous rappelons ici :

- (7) a. Paul est triste.
- b. La raison de la tristesse de Paul est que Marie est partie.

il faut prendre en compte le fait que l'antécédent de *la raison* dans (7b) n'est pas encore exprimé.

Dans le discours (7), l'état 'Paul est triste' est causé par un événement qui est exprimé linguistiquement dans la deuxième partie du discours (dans (7b)) : 'Marie est partie'.

Une représentation en SDRS du discours (7b) peut être :

(13)

π_1, π_2	
$\pi_1 :$	y, s2, e3, f
	y = Paul s2-triste (y) s2 O now f-cause (e3, s2) e3 = ?
$\pi_2 :$	x, e1
	x = Marie e1-partir (x) e1 < now e1 = e3
Commentaire-Égalité (π_1, π_2)	

Dans la représentation (13), nous avons la DRS π_1 qui représente un événement e_3 , exprimé dans une cataphore ($e_3 = ?$), qui cause l'état s2 'Paul est triste'.

Le prédicat *f-cause* (e_3, s_2) permet de réifier la cause (i.e. l'événement qui cause s2) dans un référent de discours (un verbe ou un nom).

Par ailleurs, nous avons la DRS π_2 qui représente l'événement e_1 'Marie est partie'. On apprend ici que cet événement e_1 correspond à l'événement e_3 inexprimé dans π_1 ($e_1 = e_3$) : l'événement 'Marie est partie' est la cause de l'état 'Paul est triste', d'où la relation Commentaire-Égalité entre π_1 et π_2 .

2.1.3. Un pas vers l'organisation communicative du discours : *une autre raison de*

Nous avons vu que les discours (6) et (7) faisaient intervenir une réification de la cause notamment exprimée dans le nom *raison*. Les constructions dans lesquelles apparaît ce nom ont particulièrement attiré notre attention pour deux raisons :

- l'événement dénotant la cause est toujours nouveau ;
- l'événement exprimant l'effet peut être nouveau, donné ou présupposé.

Selon la valeur communicative de l'événement exprimant l'effet (nouveau, donné ou présupposé), nous allons avoir plusieurs constructions syntaxiques possibles. C'est ce que nous proposons de regarder en détails maintenant.

2.1.3.1. Étude linguistique de *une raison de*

En fonction de l'organisation communicative du discours, notamment le choix du topique (événement, ou entité, sur lequel le discours porte), il va falloir privilégier un plan de document par rapport à un autre lors du processus de génération. C'est ce plan de document qui va être à l'origine de la réalisation syntaxique finale. Nous reviendrons sur le choix des SDRSs en fonction du topique dans la partie 2.2.1..

Pour le moment, nous allons regarder quelles sont les réalisations syntaxiques envisageables pour l'emploi de l'expression *une raison de* SN_{effet} / *pour laquelle* P_{effet} ⁸ lorsque :

- l'effet est une information nouvelle ;
- l'effet est une information donnée dans le contexte linguistique antérieur ;
- l'effet est un élément présupposé
 - o par rapport à la situation d'énonciation ;
 - o par rapport aux connaissances communes des interlocuteurs.

● Cas n°1 : l'effet est nouveau

La cause et l'effet sont des informations nouvelles : on apprend les deux événements dans deux propositions différentes.

Quand l'événement dénotant l'effet est nouveau (introduit pour la première fois dans le discours), il est exprimé dans une proposition dans laquelle le sujet et le verbe sont clairement définis (car introduits pour la première fois dans le discours).

Les discours privilégiés pour exprimer la première fois la cause et l'effet sont constitués de deux phrases distinctes avec une reprise anaphorique de la cause dans le pronom démonstratif *c'* (*cela, ça*).

Par exemple :

- (14) a. Marie est partie.
b. C'est la raison pour laquelle Paul est triste.
- (15) a. Marie est partie.
b. C'est pourquoi Paul est triste.
- (16) Marie est partie, raison pour laquelle Paul est triste.

Dans le discours (16), la phrase exprimant l'effet est apposée.

● Cas n°2 : l'effet est donné

L'effet est un élément donné car il est déjà introduit linguistiquement dans le contexte linguistique antérieur et repris dans un nom (représenté par SN_{effet}).

Il peut être repris dans l'expression *la raison de* SN_{effet} qui a pour fonction objet du verbe. Très souvent, il s'agit de verbes d'activité cognitive comme *confier, demander, expliquer, etc.* qui régissent *une raison de*.

Il peut également être repris dans l'expression *la raison de* SN_{effet} *est* P_{cause} , avec un verbe copule, comme dans le discours (7) qui nous intéressait plus haut.

Par exemple :

- (17) a. Paul est triste.
b. Il m'a confié la raison de sa tristesse : Marie est partie.
- (18) a. Paul est triste.
b. Je vais lui demander la raison de sa tristesse.

⁸ Nous utilisons la notation 'SN' pour désigner un syntagme nominal, et 'P' pour désigner une phrase. Les indices 'effet' et 'cause' signifient que le syntagme ou la phrase dénote l'effet ou la cause.

- (19) a. Paul est triste.
b. La raison de sa tristesse est que Marie est partie.

Dans les discours (17) et (19), la cause est exprimée linguistiquement alors que dans (18), elle ne l'est pas.

• Cas n°3 : l'effet est présupposé

a) car il fait référence à la situation d'énonciation

L'effet n'a pas son antécédent dans le contexte linguistique mais plutôt dans le contexte extra-linguistique : usage déictique des pronoms, des adjectifs démonstratifs, des adjectifs possessifs. Par ailleurs, l'événement dénotant l'effet est ponctuel.

Par exemple :

- (20) J'ai découvert la raison de la tristesse de Paul / de sa tristesse : Marie est partie.
(21) Je lui ai demandé la raison de cette tristesse.

Dans l'énoncé (21), l'événement désignant la cause est annoncé (dans *la raison*) mais il n'est pas exprimé linguistiquement.

b) car il fait référence à des connaissances communes

L'effet n'a pas son antécédent dans le contexte linguistique. On le trouve dans le contexte extra-linguistique car l'effet fait référence à des connaissances partagées des interlocuteurs. La caractéristique de ce cas est que l'effet dénote un événement permanent, un trait psychologique la plupart du temps.

Par exemple :

- (22) a. Marie n'est jamais revenue.
b. C'est la raison du mauvais caractère de Paul / de son mauvais caractère.
(23) Je n'ai jamais su la raison de son mauvais caractère.

Des exemples issus de corpus illustrant les cas n°1, 2, 3a et 3b sont respectivement en Annexe 1, 2, 3 et 4.

✱ Quand l'effet est exprimé avant la cause, il peut être introduit par des verbes d'activité cognitive comme *demander*, *deviner*, *connaître*, *comprendre*, *être curieux de*, etc. ou par l'expression *la raison de SN_{effet} est P_{cause}*. En d'autres termes, lorsqu'il y a un lien d'effet à cause (exprimé ou non) dans un discours, les constructions privilégiées sont les suivantes :

V activité cognitive + *la raison de SN_{effet} / la raison pour laquelle P_{effet} : P_{cause}*

- a. Paul est triste.
b. J'ai trouvé la raison de la tristesse de Paul : Marie est partie.
ou b'. J'ai trouvé la raison pour laquelle Paul est triste : Marie est partie.
ou b''. J'en ai trouvé la raison : Marie est partie.

V activité cognitive + *la raison de SN_{effet} / la raison pour laquelle P_{effet}*

- a. Paul est triste.
b. J'ai demandé la raison de la tristesse de Paul.
ou b'. J'ai demandé la raison pour laquelle Paul est triste.
ou b''. J'en ai demandé la raison.

La raison de SN_{effet}/ la raison pour laquelle P_{effet} est P_{cause}

- a. Paul est triste.
 - b. La raison de sa tristesse est que Marie est partie.
- ou
- b'. La raison pour laquelle Paul est triste est que Marie est partie.

✱ Quand la cause est exprimée avant l'effet, c'est-à-dire lorsqu'il y a un lien de cause à effet dans un discours, la construction syntaxique privilégiée est la suivante :

P_{cause}. C'est la raison de SN_{effet}/ la raison pour laquelle P_{effet}

- a. Paul est triste.
 - b. Marie est partie.
 - c. C'est la raison de la tristesse de Paul.
- ou
- c'. C'est la raison pour laquelle Paul est triste.

La phrase (a) fournit le contexte linguistique ou extra-linguistique indispensable à la suite du discours.

- ou
- a. Marie est partie.
 - b. C'est la raison pour laquelle Paul est triste.

Ici, le contexte antérieur n'est pas nécessaire. On peut apprendre dans l'énoncé (b) que 'Paul est triste'.

2.1.3.2. Étude linguistique de *une autre raison de*

Nous allons maintenant nous pencher sur une variante de l'expression *une raison de SN_{effet} / pour laquelle P_{effet}*. Nous allons regarder de quelle manière plusieurs causes peuvent être exprimées en langue, notamment par le biais de l'adjectif épithète *autre* dans des discours comme (24) :

- (24)
- a. Paul est triste parce qu'il pleut.
 - b. Une autre raison de sa tristesse est que Marie est partie.

Regardons la première définition de *autre* donnée dans le Trésor de la Langue Française Informatisé⁹ :

Adj. et pron. indéf. Permet de distinguer, de différencier, par rapport à une première partie donnée ou connue (*un, les uns*) servant de point de référence, une ou plusieurs personnes, un ou plusieurs éléments à l'intérieur d'une seconde partie.

Pour le cas qui nous occupe, l'utilisation de l'adjectif *autre* sert à distinguer des éléments donnés (une raison exprimée dans la première phrase) d'éléments nouveaux (une nouvelle raison dans la deuxième phrase).

Afin de décrire les conditions d'utilisation de l'adjectif *autre*, nous avons repéré en corpus des discours exprimant (au moins) deux événements causant le même effet. Ces discours sont en Annexe 5.

⁹ Le TLFi est consultable gratuitement à partir de l'adresse <http://zeus.inalf.fr>

Dans le discours (24), un même état s 'Paul est triste' est causé par deux événements e1 'il pleut' (dans la phrase (24a)) et e2 'Marie est partie' (dans la phrase (24b)). En termes conceptuels, cela revient à dire *cause* (e1, s) et *cause* (e2, s).

Pour exprimer une deuxième cause, les discours privilégiés sont les mêmes que ceux lorsqu'une seule cause est exprimée. Soient, par exemple :

- (25) a. Paul est triste parce qu'il pleut.
b. J'ai trouvé une autre raison de la tristesse de Paul : Marie est partie.
b'. Une autre raison de la tristesse de Paul est que Marie est partie.
- (26) a. Paul est triste parce qu'il pleut.
b. Marie est partie. C'est une autre raison de la tristesse de Paul.

Dans le discours(27) suivant :

- (27) a. Paul est triste parce qu'il pleut.
b. L'autre raison de sa tristesse est que Marie est partie.

il est présupposé qu'on attendait deux raisons : dans la phrase (27a), on explicite la première raison ; dans la phrase (27b), on explicite la deuxième raison. L'article défini *la* présuppose qu'il y a un élément connu : la deuxième raison, et elle est exprimée dans la phrase (27b).

On peut paraphraser le discours (27) par le discours(28) :

- (28) a. Paul est triste pour deux raisons.
b. Il est triste parce qu'il pleut.
c. L'autre raison de sa tristesse est que Marie est partie.

La phrase

- c'. Une autre raison de sa tristesse est que Marie est partie.

semblerait déviante dans le discours (28) si on la mettait à la place de la phrase (28c).

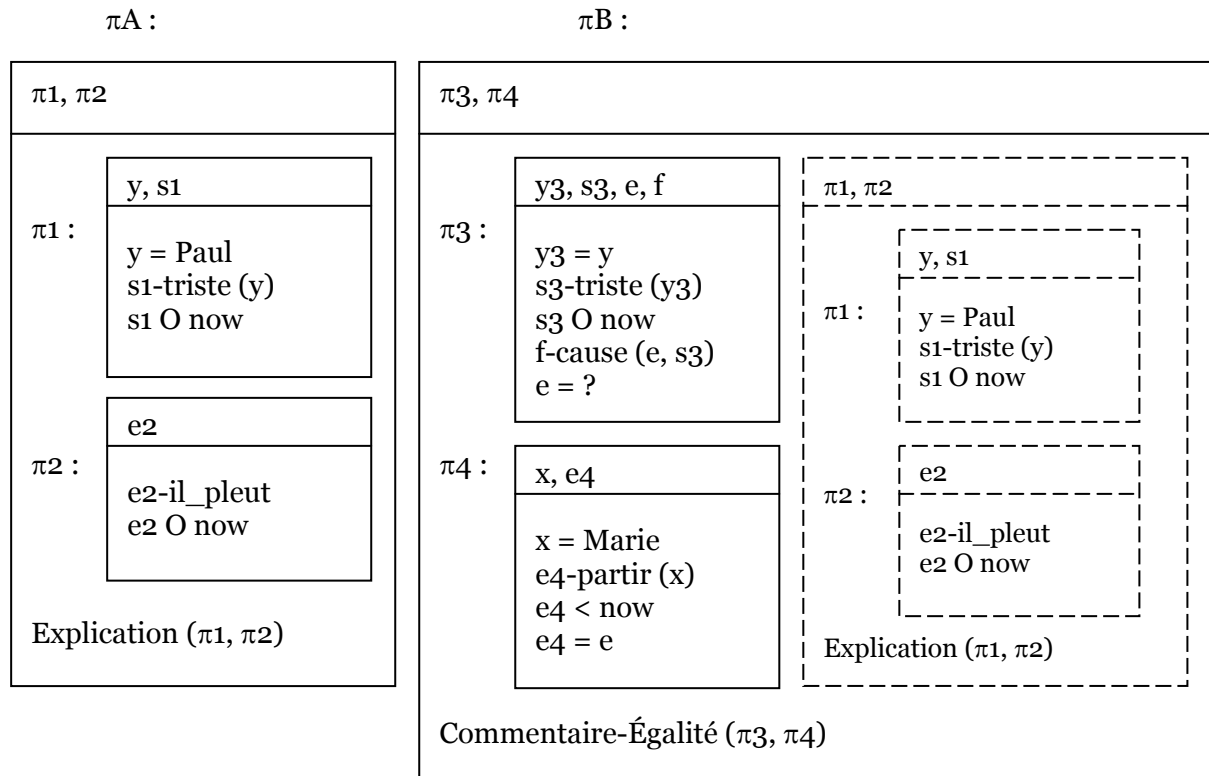
Nous avons regardé précisément en corpus les emplois de *l'autre raison* mais le peu d'exemples (7 dans la base Frantext, avec pour seule restriction l'année : à partir de 1930) ne nous permet pas de généraliser l'emploi de cette expression linguistique.

Nous allons maintenant proposer une représentation du discours (24) rappelé ici :

- (24) a. Paul est triste parce qu'il pleut
b. Une autre raison de sa tristesse est que Marie est partie.

dans lequel la relation causale réifiée de la phrase (24b) est exprimée dans le nom *raison*.

Pour représenter le discours (24), nous proposons les SDRSs π_A et π_B suivantes :



Parallèle (π_A, π_B)

La SDRS π_A représente la sémantique de la phrase (24a) et la SDRS π_B représente la sémantique de la phrase (24b).

Afin de générer l'adjectif *autre* dans (24b), il faut représenter la SDRS qui rend compte que 'Paul a déjà été triste pour une première raison'. Il faut présupposer la SDRS π_A quand on pose la SDRS π_B . La SDRS π_A est alors rappelée en pointillés dans π_B selon la proposition de van der Sandt pour la représentation des présupposés en DRT.

Conclusion

La structuration du document est une étape essentielle en génération de textes. Nous avons vu qu'elle permettait de produire différents discours à partir d'une même forme logique. Nous avons également vu que des informations communicatives peuvent entrer en jeu lors de la construction des SDRSs. Avec l'exemple de *une autre raison de*, nous avons entamé une réflexion sur l'ordre des segments phrastiques. En effet, nous avons proposé différentes réalisations syntaxiques possibles quand l'effet est exprimé avant la cause ou, inversement, quand la cause est exprimée avant l'effet. En fonction du contexte discursif, une réalisation syntaxique sera privilégiée par rapport à une autre : c'est là que le choix de la bonne SDRS doit se faire.

2.2. Structure communicative en génération de textes

Rappelons-le, l'objectif d'un générateur de textes est de produire un texte cohérent. Mais dans quelle mesure un texte est-il cohérent ? À cela, nous répondons qu'il faut qu'un texte parle de la même chose du début à la fin et que les phrases s'enchaînent le plus naturellement possible. Cette vision, *a priori* simpliste, de la cohérence du discours implique en fait la gestion de deux choix, l'un consécutif à l'autre : il faut dans un premier temps définir à partir de la forme logique de quoi le texte va parler (c'est-à-dire rajouter des informations communicatives à la forme logique) et en fonction de cela, construire un plan de document convenable (partie 2.2.1.); dans un deuxième temps, en fonction du plan de document, il faut décider de l'organisation communicative des phrases (partie 2.2.2.).

2.2.1. De la forme logique à une SDRS

Un des critères principaux (si ce n'est le plus important) pour qu'un discours soit cohérent est l'existence d'un topique discursif. Ce topique peut être un objet (un personnage, par exemple) ou un événement. C'est en fonction de ce topique que va devoir se faire le choix des plans de document. En effet, si, par exemple, le texte traite de Paul et que l'on parle tout à coup de Marie, cela va sembler étrange. Mais pour mieux cerner le problème soulevé ici, nous allons étudier cela avec des illustrations.

Afin de simplifier la lecture, nous reprenons la forme logique (1) et les discours (2) à (7) pour illustrer nos propos. Nous avons changé les temps modaux par respect de la concordance des temps ainsi que la sémantique des phrases, à moindre mesure, pour rendre les discours plus compréhensibles.

Parmi les possibilités de choix du topique, nous choisissons ici de traiter deux cas :

- 'Marie' est le topique
- 'Paul' est le topique

● Premier exemple : 'Marie' est le topique discursif

Le passage de la forme logique à une ou plusieurs SDRSs se fait en fonction du topique 'Marie'. Nous partons donc du principe que le discours traite de 'Marie' depuis le début comme dans (29) :

- (29) a. Marie a passé la soirée avec des amis.
b. Elle a revu Paul.
c. Ils ont discuté.
d. Marie a dû partir tôt.
e. Paul était triste.

Le discours (29) semble tout à fait cohérent : il y a une continuité thématique, il traite du même topique 'Marie'. Pour la comparaison, voyons le discours (30) :

- (30) a. Marie a passé la soirée avec des amis.
b. Elle a revu Paul.
c. Ils ont discuté.
d. Paul était triste.
e. Marie a dû partir tôt.

Le discours (30), quant à lui, semble étrange parce que l'on passe d'une entité à une autre : 'Marie' (dans (30a) et (30b)), puis 'Paul' (dans (30d)), puis on revient sur 'Marie' dans (30e). Malgré la présence de la phrase (30c) qui traite des deux personnages, l'enchaînement avec la phrase (30d) dont le thème est 'Paul' est maladroit. Cela s'explique par le fait que 'Marie' est

le sujet central du discours, i.e. le topique. Même si une certaine progression thématique est respectée dans (30), le changement de topique dans (30d) fait que le discours (30) est moins cohérent que le discours (29).

Pour accepter l'enchaînement des phrases (30d)–(30e) il faudrait le discours suivant :

- (31) a. Paul a passé la soirée avec des amis.
b. Il a revu Marie.
c. Ils ont discuté.
d. Paul était triste.
e. Marie a dû partir tôt.

Ici, le topique est 'Paul'. Il n'y a pas de changement de topique au milieu du texte, donc le discours respecte une certaine cohérence. Ceci rejoint parfaitement les travaux décrits dans (Grosz, Joshi et Weinstein, 1995).

Si le choix d'un topique impose un ordre des phrases dans le discours, cela implique que le choix d'un topique impose un plan de document particulier.

Nous allons voir pour la forme logique (1) quelles sont les SDRSs sous-jacentes possibles et impossibles¹⁰ en fonction du topique discursif 'Marie'.

Les SDRSs telles que 'Marie' est le topique sont les suivantes :

- la SDRS (10) peut donner lieu à la réalisation :
(32) Marie a dû partir tôt.
Paul était triste.
comme dans le discours (3)
- la SDRS (11) peut donner lieu à :
(33) Marie a dû partir tôt.
À cause de cela, Paul était triste.
comme dans le discours (4)
- la SDRS (12) peut donner lieu à :
(34) Marie a dû partir tôt.
Cela a rendu Paul triste.
ou à
(35) Marie a dû partir tôt.
C'est la raison pour laquelle Paul était triste.
comme dans les discours (5) et (6).

Dire que 'Marie' est le topique discursif revient à signaler qu'il faut mettre tout ce qui concerne 'Marie' (dans la forme logique) dans la première DRS, soit ici π_1 . À partir de là, tout ce qui reste de la forme logique va dans la DRS π_2 , sauf la condition *cause* (e_1, s_2) si l'on veut qu'elle soit traduite par une relation de discours.

Les SDRSs qui brisent la continuité thématique du discours (29) sont les suivantes :

- la SDRS (9) peut donner lieu à la réalisation :
(36) Paul était triste.
Marie a dû partir tôt.
comme dans les discours (2) et (30).

¹⁰ Nous ne reproduisons pas les SDRSs car nous faisons référence à des représentations existantes (voir les numéros des SDRSs).

- la SDRS (13) peut donner lieu à :
 (37) Paul était triste.
 La raison de la tristesse de Paul est que Marie a dû partir tôt.
 comme dans le discours (7).

● Deuxième exemple : ‘Paul’ est le topique discursif

Nous avons vu plus haut que le discours (31) fonctionnait bien car le topique ‘Paul’ était constant dans tout le discours (on n’a pas changé pas de topique en cours de route comme dans (30)).

À partir de la forme logique (1) et tel que ‘Paul’ soit le topique, la SDRS (9) convient et donnerait lieu à (38) :

- (38) a. Paul a passé la soirée avec des amis.
 b. Il a revu Marie.
 c. Ils ont discuté.
 d. Paul était triste parce que Marie a dû partir tôt.

En revanche, (39) n’est pas acceptable :

- (39) a. Paul a passé la soirée avec des amis.
 b. Il a revu Marie.
 c. Ils ont discuté.
 d. La raison de la tristesse de Paul est que Marie a dû partir tôt.

Le discours (39) ne fonctionne pas car dans la phrase (39d), *la tristesse de Paul* est un élément donné comme nous l’avons expliqué dans la partie 2.1.3.1.. Ceci implique qu’il doit y avoir dans le contexte antérieur une phrase qui dénote le fait ‘Paul est triste’ comme dans le discours :

- (40) a. Paul a passé la soirée avec des amis.
 b. Il a revu Marie.
 c. Ils ont discuté.
 d. Paul était triste.
 e. La raison de la tristesse de Paul est que Marie a dû partir tôt.

Le discours (40) est plus acceptable que (39) car l’expression *la raison de la tristesse de Paul* impose que le discours ait traité auparavant de ‘la tristesse de Paul’ et pas seulement de ‘Paul’ (comme dans (39)).

Les SDRSs telles que ‘Paul’ est le topique sont les suivantes :

- la SDRS (9) peut donner lieu à la réalisation :
 (41) Paul était triste.
 Marie a dû partir tôt.
 comme dans le discours (2) et (38).
- la SDRS (13) peut donner lieu à :
 (42) Paul était triste.
 La raison de la tristesse de Paul est que Marie a dû partir tôt.
 comme dans les discours (7) et (40).

La notion de topique mérite indiscutablement une étude plus approfondie et plus sérieuse que celle que nous avons présentée ici et dans la partie 1.2.2.. Il serait essentiel de pouvoir déterminer, à l'aide de critères linguistiques et/ou de tests, l'entité ou l'événement qui constitue le topique d'un discours. Dans le cadre de notre travail, en génération, nous savons de quoi nous voulons parler donc le problème de la détermination du topique ne se pose pas *a priori*. En revanche, il serait nécessaire de savoir plus précisément quelles sont les conséquences syntaxiques lorsqu'un topique est déterminé (usage des pronoms personnels, par exemple).

2.2.2. D'une SDRS à la structure communicative des phrases

Une fois la ou les SDRS(s) choisie(s), il faut spécifier l'organisation des phrases qui constituent le discours à générer.

Ici, nous allons proposer pour chaque relation de discours décrite en SDRT la ou les progression(s) thématique(s) privilégiée(s) observée(s) en corpus. Nous tenons à préciser qu'il s'agit d'enchaînements *privilégiés* car pour chaque relation, il existe différentes progressions possibles.

Par souci de clarté, nous utilisons une légende (couleurs) pour référer aux éléments dont nous rappelons les définitions :

- le **thème** est le propos de l'énoncé ;
- le **rhème** est ce qui est communiqué (quant au thème ou non) ;
- le **focus rhématique** est l'élément le plus saillant au niveau informationnel dans le rhème ;
- le **spécifieur modal** est la marque de la subjectivité du message (présence du locuteur) ;
- le **spécifieur circonstant** indique les circonstances de l'événement ou de l'énonciation ;
- le spécifieur externe peut être un connecteur traduisant une relation de discours.

Un segment phrastique correspond à un énoncé et plus précisément à la réalisation syntaxique d'un sens propositionnel (que l'on retrouve dans une DRS, par exemple). Il est représenté dans une « bulle ».

Pour chaque segment, i.e. dans chaque bulle, l'organisation communicative est indiquée par le biais de « boîtes » qui peuvent représenter le thème (en bleu), le rhème (en rouge), le focus rhématique (compris dans le rhème), le spécifieur modal (en vert) ou le spécifieur circonstant (en mauve).

Entre chaque segment, la progression thématique est signalée : la flèche entre deux bulles indique quel élément (thème, rhème, etc.) d'un segment est repris dans un élément d'un autre segment. Il peut arriver que ce soit l'événement entier décrit dans un segment qui soit lié à un élément d'un autre segment : la flèche part alors de la bulle vers un élément de l'autre segment, et non d'un élément particulier de la bulle vers un autre. Mais voyons maintenant cela plus concrètement.

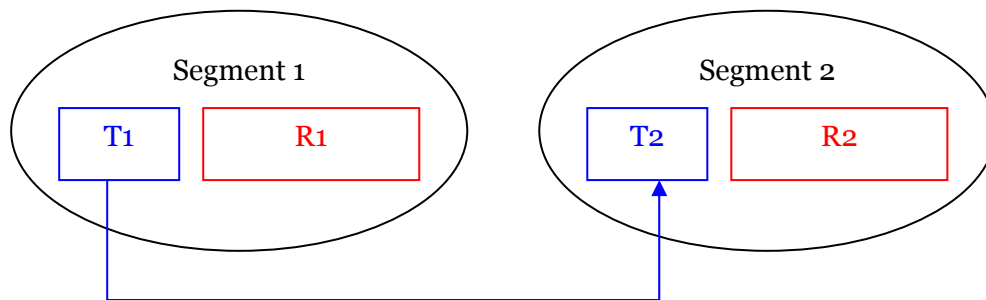
2.2.2.1. La relation Narration

S'il existe une relation Narration entre deux événements décrits dans deux segments phrastiques S1 et S2, les rhèmes dénotent des événements successifs et le thème du premier segment (S1) peut être repris en thème dans le deuxième (S2).

Par exemple :

S1 : Marie est rentrée dans le salon.

S2 : Puis, elle a dit bonjour à tout le monde.



On peut également trouver une relation Narration dans le discours :

S1 : Mon collègue d'Elk Island a mentionné brièvement que les chiffres ne reflètent pas fidèlement la réalité.

S2 : Dans les minutes qu'il me reste, je vais repasser ces chiffres rapidement.

Cette progression dite « à thème constant », est surtout contrainte par le topique discursif, comme nous l'avons vu dans la partie 2.2.1., car nous pouvons trouver des discours comme

S1 : Marie est rentrée dans le salon.

S2 : Paul est allé l'embrasser.

dans lequel il y a une relation Narration en jeu. Le thème n'est pas constant mais on retrouve 'Marie' sous la forme d'un élément donné dans S2 i.e. le pronom personnel *la (l')*. Ce type de progression est possible mais il est préférable de trouver le même thème pour exprimer une relation Narration entre deux événements.

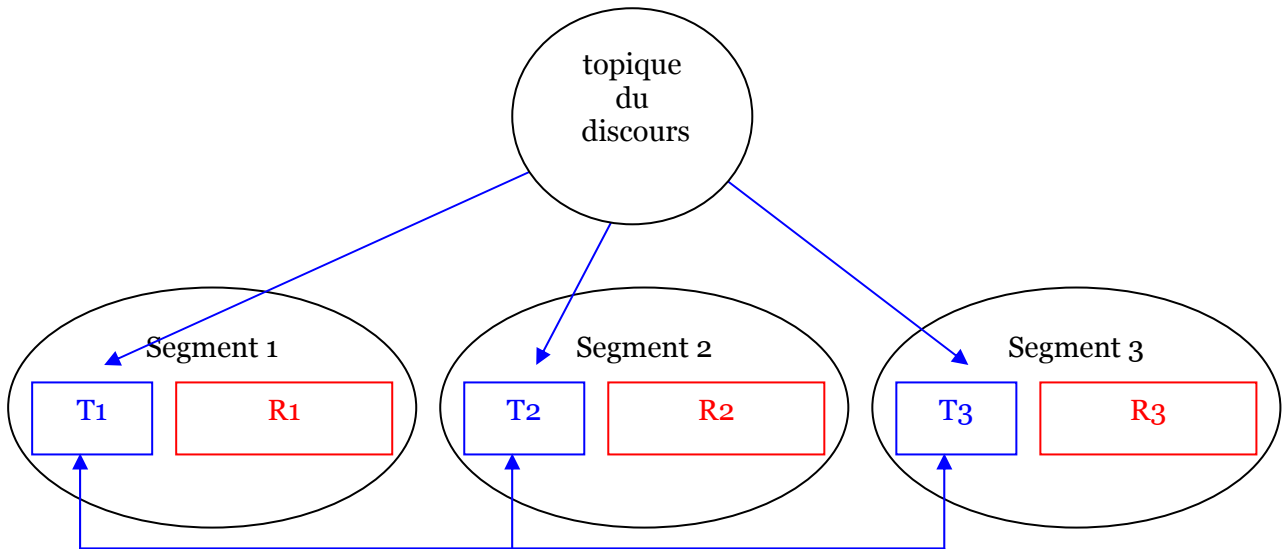
2.2.2.2. La relation Parallèle

Les événements décrivent le même référent de discours (le thème). On peut comparer la relation Parallèle à celle de Narration sans succession : ici, les événements sont simultanés (e1 O e2). Il s'agit d'une élaboration du topique (explicitée plus loin en 2.2.2.4.) sur plusieurs segments phrastiques.

Le thème (souvent introduit en début de discours dans un segment rhématique) peut être repris d'un segment à l'autre.

Par exemple :

- So : Je voudrais parler de Marie.
- S1 : Marie est partie tôt hier soir.
- S2 : Elle n'a dit bonsoir à personne.
- S3 : Elle avait l'air furieuse.



La relation Parallèle rejoint Narration d'un point de vue communicatif dans le sens où ces deux relations imposent que le thème soit repris dans chaque segment, comme par exemple dans les discours suivants :

- So : Il faut aussi souligner (...) la performance de Simon Gamache
- S1 : qui a récolté au moins un point dans chacun des 21 matchs éliminatoires pour compléter la série avec un record de 57 points.
- S2 : Il a également remporté le trophée Guy Lafleur.

- S1 : Le projet de loi augmente le montant sur lequel le crédit d'impôt pour déficience mentale ou physique est basé.
- S2 : Il élargit la liste des proches auxquels le crédit peut être transféré....
- S3 : De plus, le projet de loi accroît la déduction maximale pour frais de garde d'enfants....
- S4 : Il ajoute aussi à la liste des dépenses donnant droit au crédit d'impôt pour frais médicaux....

2.2.2.3. La relation Commentaire

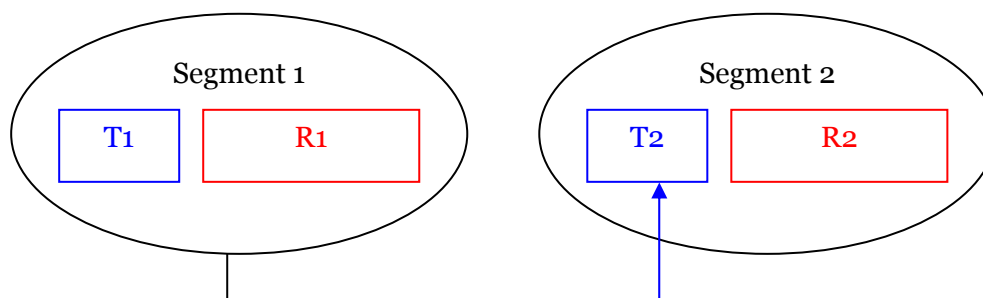
L'événement décrit dans S2 est un commentaire, une glose sur l'événement décrit dans S1.

Les événements décrits dans le(s) segment(s) précédent(s) peuvent être repris en thème dans S2 sous la forme du pronom démonstratif *cela* ou d'un substantif déverbal (co-référence).

Par exemple :

S1 : Marie est partie tôt hier soir.

S2 : Cela a rendu Paul triste.



Ce type de progression est courant en discours : on parle d'un événement et on en dit quelque chose par la suite. La partition thème~rhème des segments qui décrivent l'événement repris dans S2 n'a aucune importance. Voici quelques illustrations de ce phénomène :

1 événement { Le crime organisé a des conséquences sur la pauvreté.
On s'attaque aux populations les plus vulnérables.
On les utilise plus facilement.

S2 : Il faut être bien conscients que cela a des coûts sociaux et économiques très importants pour notre société.

S1 : Les crayons, les stylos, le papier, l'approvisionnement en eau, les ordinateurs, la main-d'oeuvre, presque tout vient du Québec.

S2 : Cela témoigne du fait que l'économie québécoise est très bien diversifiée.

S1 : Le projet de loi contient de nombreuses modifications de forme pour moderniser le libellé de la loi afin de suivre l'évolution de notre système judiciaire et de permettre aux sociétés de faire une meilleure utilisation des communications électroniques.

S2 : Cela ne pose aucun problème.

S1 : Ce n'est pas normal qu'il y ait à peine plus de 40 p. 100 des personnes au chômage qui puissent bénéficier du régime d'assurance-emploi.

S2 : C'est parce que le régime est trop restreint....

2.2.2.4. La relation Élaboration

Nous distinguons ici, comme nous l'avons fait lors de la définition des relations de discours (dans la partie 1.1.3.), une Élaboration objet et une Élaboration événement.

a. Élaboration objet

Le premier segment phrastique introduit un nouveau référent de discours. Ce référent peut être repris dans le deuxième segment dans lequel on en dit quelque chose.

Le nouveau référent peut être présenté par le biais d'une phrase matrice (avec un verbe performatif) correspondant à un spécifieur modal (façon qu'a le locuteur de présenter le message) ou par le biais d'un présentatif comme *il y a*.

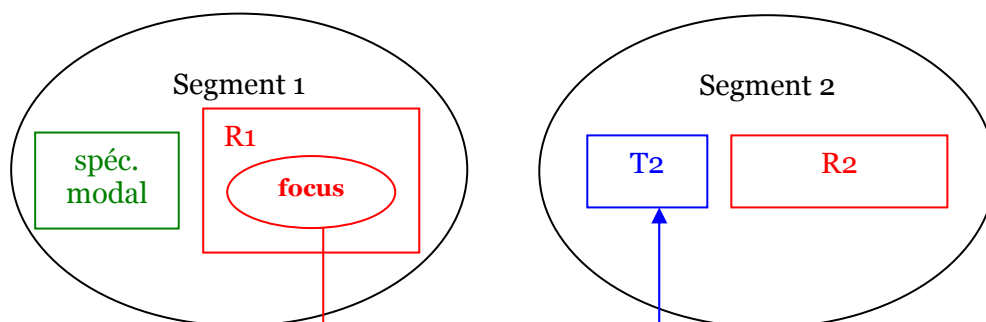
S2 élabore, fournit des informations sur l'objet introduit dans S1.

- Le focus rhématique de S1 peut être repris en thème dans S2.

Par exemple :

S1 : Je voulais te parler de Marie.

S2 : Elle est partie hier soir en claquant la porte.



Les discours suivants illustrent cette progression thématique particulière (appelée *chaining* dans (Mel'čuk, 2001)) et fréquemment utilisée :

S1 : Il faut aussi souligner (...) la performance de Simon Gamache

S2 : qui a récolté au moins un point dans chacun des 21 matchs éliminatoires pour compléter la série avec un record de 57 points.

S1 : La performance de Maxime Daigneault est aussi à noter.

S2 : Il devient le premier gardien de but recrue de 16 ans....

S1 : Je voudrais citer un passage de l'article....

S2 : Cet article a été écrit le 3 octobre 2000....

S1 : Il est intéressant de souligner le déficit financier de la Colombie-Britannique en termes d'achat fédéraux de biens et de services sur son territoire.

S2 : Ce déficit est proportionnellement plus élevé que pour toutes les autres provinces et est également plus élevé en chiffres absolus.

S1 : Il y a une nouvelle crainte généralisée selon laquelle les normes de salubrité environnementale ont chuté.

S2 : Cette crainte est très réelle [parce que deux choses se sont produites simultanément.]

S1 : Je vais maintenant citer un document.

S2 : Il a été publié le 23 juin 1993 par le leader parlementaire d'alors et par le whip d'alors.

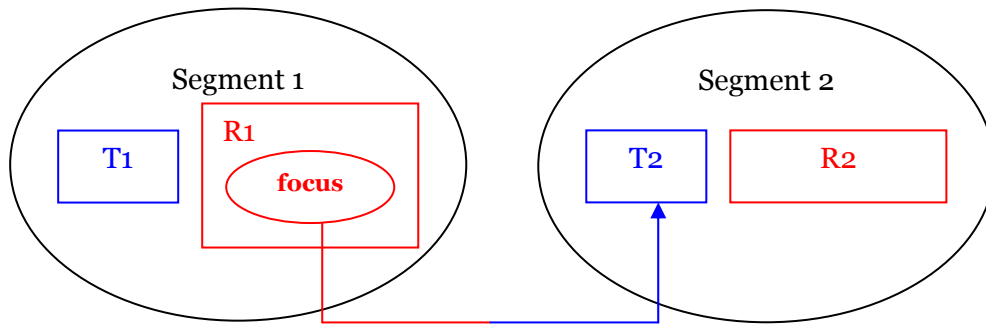
S3 : Ils signalaient très clairement....

Dans le dernier discours, lors du passage de S2 à S3, le focus rhématique (correspondant à deux éléments de S2) n'est pas introduit par un segment rhématique, il fait juste partie du rhème dans S2 dont l'organisation est une partition thème~rhème classique. Il s'agit ici d'une autre façon de traduire une Élaboration objet.

Par exemple :

S1 : Jean a vu Marie ce matin.

S2 : Elle n'avait pas l'air contente.



De manière générale, le focus rhématique est repris en thème dans la phrase suivante (S2) quelle que soit la structure communicative de S1. On peut même dire que tout focus rhématique est un thème potentiel.

- On peut aussi trouver le focus rhématique de S1 repris sous la forme de thèmes dérivés dans plusieurs segments successifs.

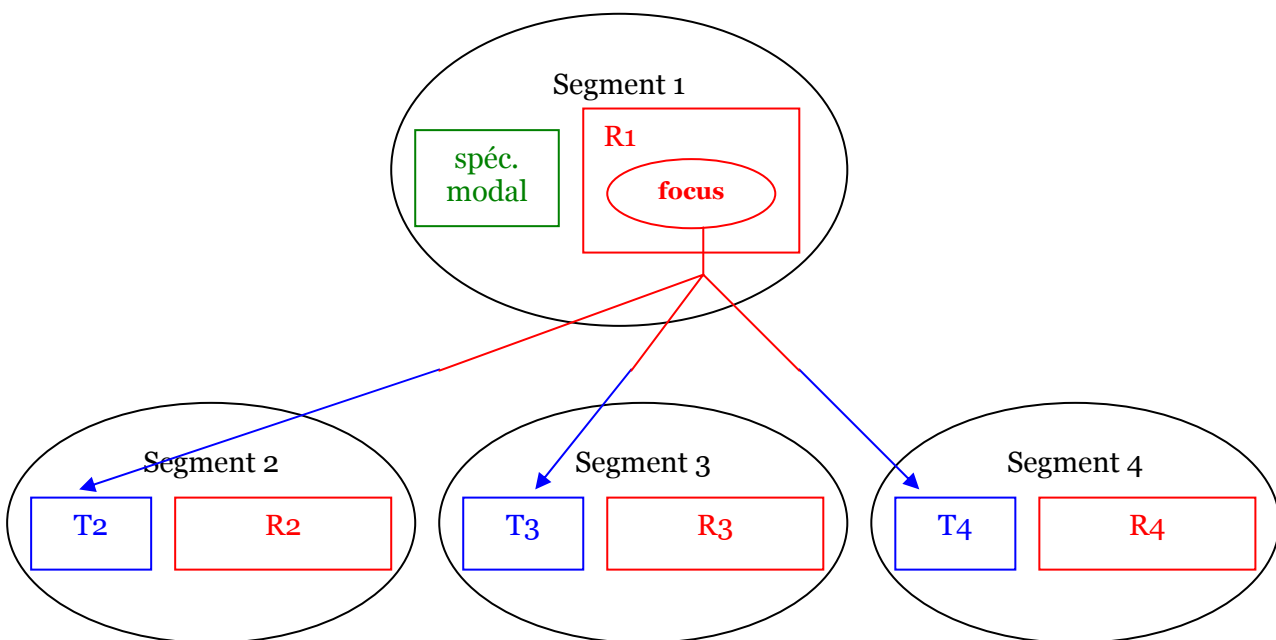
Par exemple :

S1 : Avant de terminer, j'aborderai brièvement trois autres mesures.

S2 : La première instaurerait des changements à l'imposition des fiducies et de leurs bénéficiaires....

S3 : La mesure suivante permettrait au gouvernement du Canada de continuer d'imposer aux immigrants des taxes....

S4 : Une autre mesure rendrait intégralement déductibles les frais de publicité dans les périodiques....



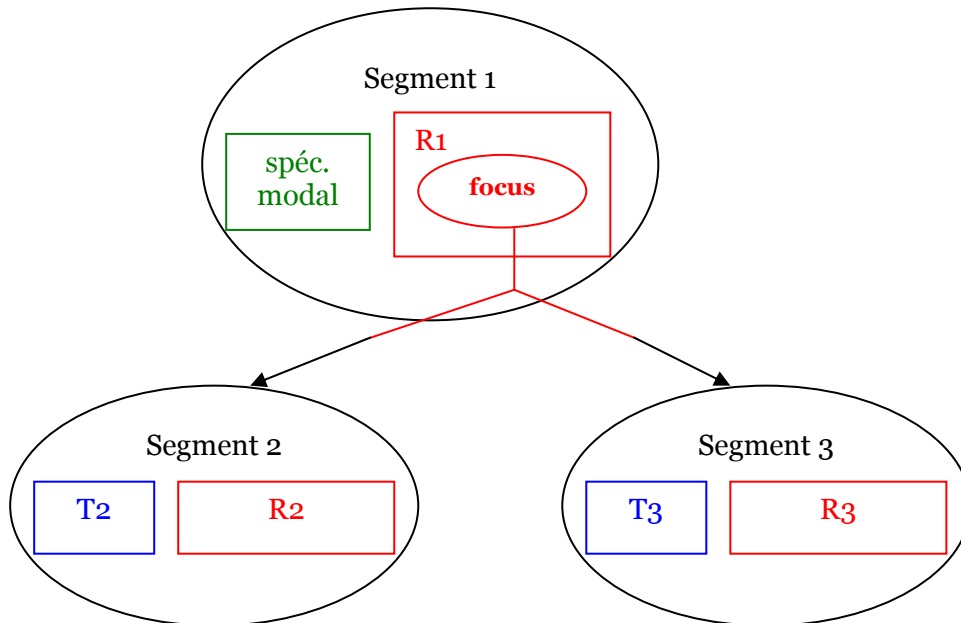
- Nous pouvons également trouver le focus rhématique de S1 repris sous la forme d'événements décrits dans les segments suivants.

Par exemple :

S1 : Je voulais vous dire deux choses :

S2 : Marie est partie tôt hier soir.

S3 : Elle n'a dit bonsoir à personne.



Cette progression thématique permet d'introduire plusieurs événements dans le discours, comme par exemple dans les discours suivants :

S1 : [Cette crainte est très réelle parce que] deux choses se sont produites simultanément.

S2 : Tout d'abord, il y a plus de menaces à la santé.

S3 : Deuxièmement, il y a moins de protection en matière de santé.

S1 : Le communiqué de presse... disait ceci :

S2 : Cette loi établira des normes nationales applicables en matière de santé....

S3 : Cette mesure contribuera beaucoup à maintenir et à améliorer l'état de santé des Canadiens.

S1 : Les modifications contenues dans le projet de loi mettent à jour quatre éléments fondamentaux de la loi actuelle.

S2 : Premièrement, le projet de loi accroît les droits des actionnaires de communiquer entre eux....

S3 : Deuxièmement, le projet de loi aidera à éliminer les obstacles à la concurrence....

S4 : Troisièmement, le projet de loi S-11 définit de façon plus raisonnable les responsabilités des sociétés....

S5 : Enfin, le projet de loi élimine le chevauchement réglementaire.

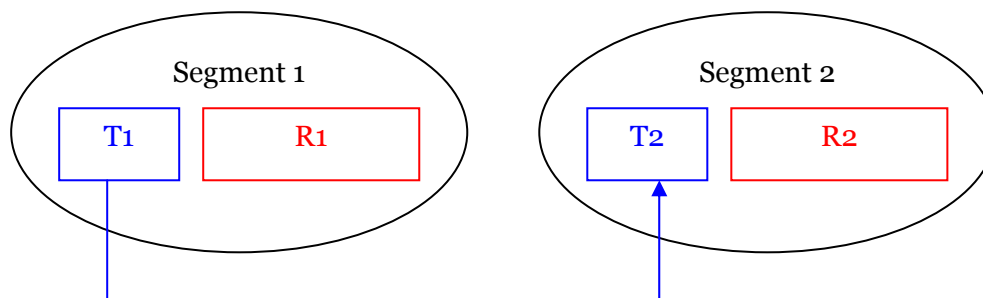
b. Élaboration événement

Le deuxième événement est une partie de l'événement décrit dans le premier segment. L'événement de S2 élabore l'événement de S1. Le thème de S1 peut être repris dans S2 .

Par exemple :

S1 : Marie est partie hier soir.

S2 : Elle a pris sa voiture.



Comme la relation Narration, la reprise du thème est privilégiée quand il y a une relation Élaboration événement. Ce qui distingue les deux relations tient aux événements dénotés dans les segments phrastiques : pour Narration, les événements sont successifs, pour Élaboration, le deuxième événement est un sous-événement du premier.

D'autres discours illustrent cette progression thématique :

S1 : Personnellement, j'étais là lors de l'adoption du projet de loi C-20.

S2 : J'ai voté contre ce projet de loi.

S1 : Quelques publications de cette période ... constituent des études approfondies sur la famille canadienne.

S2 : Dans ces publications, il est question du soin des personnes âgées, du soin des parents, du divorce et du mariage.

2.2.2.5. La relation Particularisation

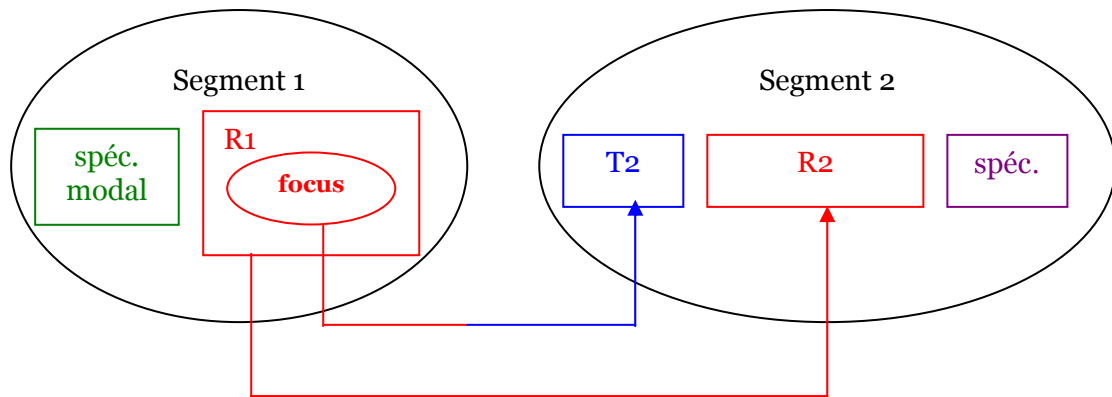
Il s'agit ici d'un cas de co-référence événementielle : l'événement décrit dans le premier segment phrastique est repris dans le deuxième avec un ajout d'informations. Les informations supplémentaires peuvent correspondre à des circonstants (adjonction) ou à une spécification (du rhème, en l'occurrence).

- Le focus rhématique de S1 peut être repris en thème et le rhème de S1 repris dans S2 avec des circonstants.

Par exemple :

S1 : Je voudrais parler du départ de Marie.

S2 : Elle est partie hier soir en claquant la porte.



Ce type d'enchaînement est très utilisé pour commencer un discours : le locuteur introduit un événement, et le développe par la suite, comme dans le discours :

S1 : Monsieur le Président, j'aimerais souligner la victoire hier de l'équipe des Foreurs de Val-d'Or.

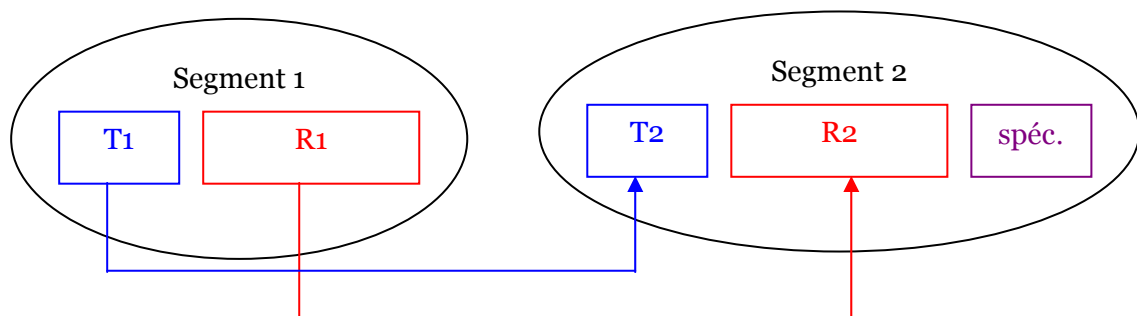
S2 : À leur huitième saison dans la Ligue junior majeure du Québec, les Foreurs ont remporté pour la deuxième fois de leur histoire les séries éliminatoires.

- On peut également trouver le thème et le rhème de S1 repris dans S2 avec des circonstants.

Par exemple :

S1 : Marie est partie.

S2 : Elle est partie hier soir en claquant la porte.

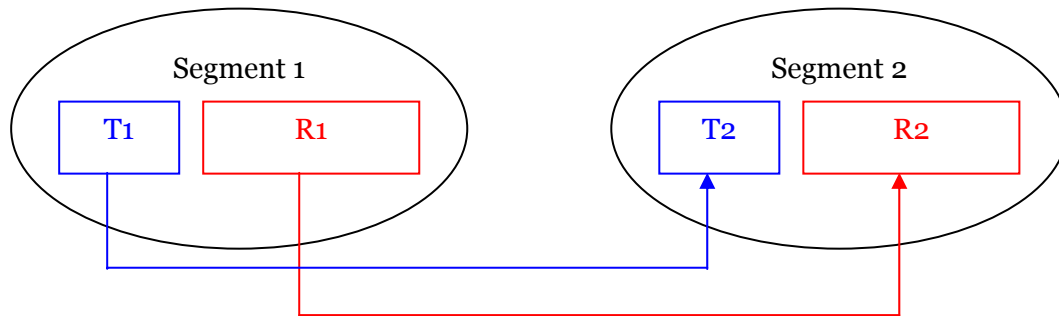


- Lorsque le deuxième segment n'apporte pas d'informations supplémentaires sur l'événement décrit dans le premier segment, il s'agit d'une paraphrase. Les événements décrits sont exactement identiques : co-référence événementielle stricte. Le thème et le rhème de S1 sont alors repris dans S2.

Par exemple :

S1 : Marie n'était pas contente.

S2 : Elle était malheureuse.



D'autres discours illustrent ce genre de progression :

S1 : (...) le régime est trop restreint, le régime est trop fermé.

S2 : Ce régime exclut trop de personnes....

S1 : Des milliers de Canadiens veulent toujours faire confiance à leurs gouvernements.

S2 : Ils veulent pouvoir compter sur leurs représentants politiques.

S1 : Trudeau est le parrain de la politique économique interventionniste du gouvernement fédéral.

S2 : C'est Trudeau qui a été le fer de lance de la société....

Dans le dernier exemple, la focalisation du rhème (i.e. *Trudeau*) dans S2 permet une structure du discours particulière qui accentue l'effet de paraphrase : le thème de S1 est repris dans S2 sous la forme d'une proposition et le rhème de S1 est repris dans S2 sous la forme du nom *Trudeau*.

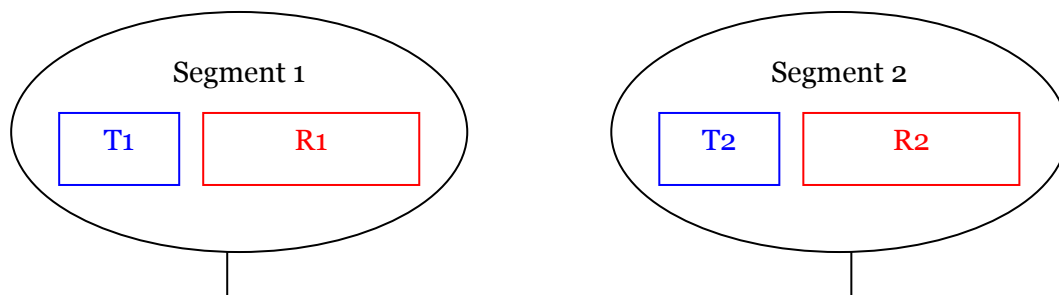
2.2.2.6. La relation Explication

L'événement décrit dans le deuxième segment phrastique correspond à l'explication de l'événement décrit dans le premier segment.

Par exemple :

S1 : Paul est triste.

S2 : Marie a dû partir tôt hier soir.



La structure communicative des phrases ne rentrent pas en jeu lorsqu'il y a une relation Explication. Voici d'autres exemples de discours impliquant une Explication :

S1 : J'en suis également fier
S2 : parce que mon fils Samuel joue pour cette équipe.

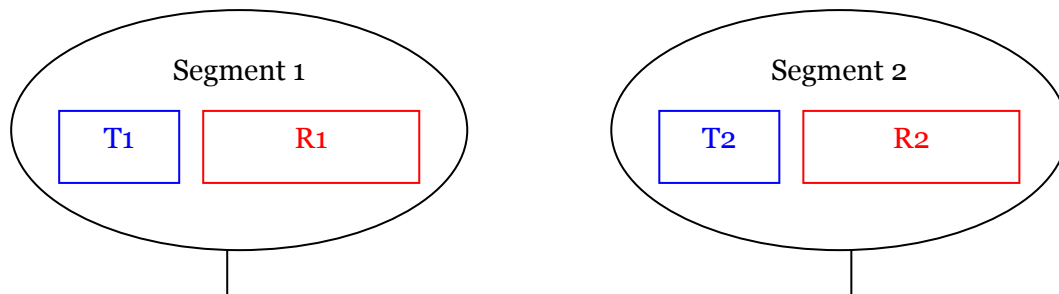
S1 : Cette crainte est très réelle
S2 : parce que deux choses se sont produites simultanément.

2.2.2.7. La relation Résultat

L'événement décrit dans le deuxième segment phrastique correspond à la conséquence de l'événement décrit dans le premier segment.

Par exemple :

S1 : Marie a dû partir tôt hier soir.
S2 : Paul est triste.



Tout comme la relation Explication, le relation Résultat n'impose pas de structures communicatives particulières, comme dans les discours :

S1 : À leur huitième saison dans la Ligue junior majeure du Québec, les Foreurs ont remporté pour la deuxième fois de leur histoire les séries éliminatoires.
S2 : Je suis fier de cette équipe.

S1 : Il n'y a pratiquement plus personne qui soit capable de remplir sa propre déclaration de revenus.
S2 : Les gens doivent faire appel à des comptables ou des spécialistes.

Pour les relations Explication et Résultat, la structure communicative n'est pas pertinente. Ces relations dépendent surtout de la sémantique des énoncés, non de leurs organisations communicatives internes. En effet, il n'y a aucune progression thématique régulière qui puisse exprimer une relation causale. La pertinence de ces relations intervient lors du choix du topique dans une étape antérieure dans le processus de génération : nous avons vu qu'en fonction d'un topique, l'une ou l'autre de ces relations était privilégiée (partie 2.2.1.).

Mais outre le cas d'Explication et de Résultat, nous pouvons dire qu'à une relation de discours peut correspondre une structure communicative particulière pour chaque segment phrastique. Bien que l'on puisse trouver des progressions thématiques différentes pour une même relation de discours, il existe quand même des structures communicatives privilégiées pour exprimer une relation entre deux événements. Notons que l'étude présentée ici n'est pas bornée et qu'elle mériterait qu'on l'approfondisse.

Conclusion

Nous pouvons à ce stade tirer deux grandes conclusions. D'une part, il existe un lien entre la cohérence du discours (liée à un topique) et le choix d'un plan de document. Et d'autre part, il existe une corrélation entre une relation de discours et la structure communicative des phrases (mis à part le cas des relations Explication et Résultat). Il y a bel et bien interaction entre la structure du discours (mis au jour par les relations de discours) et l'organisation communicative des segments phrastiques qui constituent le discours.

Conclusion deuxième partie

Nous avons vu dans cette deuxième partie que la structure communicative pouvait intervenir à différents niveaux : lors du passage de la forme logique à une SDRS et lors du passage d'une SDRS à la réalisation syntaxique des énoncés qui constituent le discours à générer. Les informations communicatives que nous souhaitons intégrer dans le processus de génération semblent pertinentes et nécessaires pour produire des textes cohérents. La SDRT a apporté de nombreuses solutions quant à la structuration du texte mais nous pouvons encore améliorer l'organisation du discours grâce à des informations communicatives.

Nous allons voir dans une troisième partie comment et à quel niveau ces informations interviennent concrètement.

3. Génération d'un texte avec des informations communicatives

Dans la partie précédente, nous avons détaillé pour de nombreuses configurations discursives quelles étaient les possibilités de plans de document et de structures communicatives. Nous allons ici appliquer sur un texte issu d'un corpus le processus de génération détaillé auparavant.

Nous allons tout d'abord présenter le texte que l'on veut générer. Ce texte sera segmenté en propositions (ou segments phrastiques) puis nous proposerons deux segments discursifs à générer.

Ensuite, nous passerons à la génération proprement dite des deux segments discursifs pour illustrer la pertinence de l'intégration d'informations communicatives en génération de textes.

3.1. Présentation du texte à générer

Pour illustrer le processus de génération d'un texte avec des informations communicatives, nous avons choisi de produire (théoriquement) quelques segments phrastiques du texte suivant :

M. Claude Duplain (Portneuf, Lib.) :

Monsieur le Président, j'aimerais souligner la victoire hier de l'équipe des Foreurs de Val-d'Or.

À leur huitième saison dans la Ligue junior majeure du Québec, les Foreurs ont remporté pour la deuxième fois de leur histoire les séries éliminatoires.

Je suis fier de cette équipe d'autant plus que cette formation a terminé en dernière position l'an dernier.

Une autre raison d'en être fier est que mon fils Samuel joue pour cette équipe.

Il faut aussi souligner la performance de tous ces jeunes et peut-être plus particulièrement celle de Simon Gamache, qui a récolté au moins un point dans chacun des 21 matchs éliminatoires pour compléter la série avec un record de 57 points.

Il a également remporté le trophée Guy Lafleur.

La performance de Maxime Daigneault est aussi à noter.

Il devient le premier gardien de but recrue de 16 ans à conduire sa formation à la conquête de la Coupe du président.

Félicitations à l'équipe, aux entraîneurs et à tous ceux qui ont participé à cette victoire.

Il s'agit d'un extrait de déclaration de députés à la Chambre des Communes à Montréal (Hansard révisé n° 58, le 9 mai 2001). La source de ce corpus est le concordancier bilingue TransSearch¹¹.

La première étape consiste à découper le discours en segments phrastiques. Ce que nous appelons segment phrastique correspond à la réalisation syntaxique d'un sens propositionnel. Chaque segment phrastique aura une représentation en DRS.

Voici le découpage proposé :

¹¹ TransSearch est accessible par <http://www.tsrali.com>

<segment 1> Monsieur le Président, j'aimerais souligner la victoire hier de l'équipe des Foreurs de Val-d'Or.

<segment 2> À leur huitième saison dans la Ligue junior majeure du Québec, les Foreurs ont remporté pour la deuxième fois de leur histoire les séries éliminatoires.

<segment 3> Je suis fier de cette équipe

<segment 4> d'autant plus que cette formation a terminé en dernière position l'an dernier.

<segment 5> Une autre raison d'en être fier est que

<segment 6> mon fils Samuel joue pour cette équipe.

<segment 7> Il faut aussi souligner la performance de tous ces jeunes

<segment 8> et peut-être plus particulièrement celle de Simon Gamache,

<segment 9> qui a récolté au moins un point dans chacun des 21 matchs éliminatoires pour compléter la série avec un record de 57 points.

<segment 10> Il a également remporté le trophée Guy Lafleur.

<segment 11> La performance de Maxime Daigneault est aussi à noter.

<segment 12> Il devient le premier gardien de but recrue de 16 ans à conduire sa formation à la conquête de la Coupe du président.

<segment 13> Félicitations à l'équipe, aux entraîneurs et à tous ceux qui ont participé à cette victoire.

Nous partons du principe que chaque segment phrastique a une DRS sous-jacente et qu'entre chaque DRS, il y a des relations de discours en jeu.

La deuxième étape consiste à choisir quels segments discursifs nous allons proposer pour illustrer la génération du texte. Nous allons simuler le processus de génération des discours suivants :

<segment discursif 1>

Monsieur le Président, j'aimerais souligner la victoire hier de l'équipe des Foreurs de Val-d'Or.

À leur huitième saison dans la Ligue junior majeure du Québec, les Foreurs ont remporté pour la deuxième fois de leur histoire les séries éliminatoires.

Je suis fier de cette équipe d'autant plus que cette formation a terminé en dernière position l'an dernier.

Une autre raison d'en être fier est que mon fils Samuel joue pour cette équipe.

<segment discursif 2>

Il faut aussi souligner la performance de tous ces jeunes et peut-être plus particulièrement celle de Simon Gamache, qui a récolté au moins un point dans chacun des 21 matchs éliminatoires pour compléter la série avec un record de 57 points.

Il a également remporté le trophée Guy Lafleur.

L'étude séparée de ces deux discours ne signifie pas qu'ils sont indépendants l'un de l'autre. Nous verrons que le texte entier forme un tout cohérent et nous en proposons la structure générale en Annexe 6.

3.2. Processus de génération du texte

Pour générer les discours 1 et 2, nous allons nous baser sur les études réalisées dans la partie 2. Pour chaque discours, nous allons tout d'abord procéder au choix d'un topique, puis, en fonction de ce topique, définir une SDRS sous-jacente. Ensuite, en fonction de la relation de discours en jeu (dans la SDRS), nous allons proposer les structures communicatives privilégiées pour chaque segment phrastique. Ce processus, bien que théorique, devrait aboutir à la formation des deux discours originaux.

3.2.1. Génération du segment discursif 1

Nous rappelons ici le segment discursif 1 (avec le découpage en segments phrastiques) :

<segment 1> Monsieur le Président, j'aimerais souligner la victoire hier de l'équipe des Foreurs de Val-d'Or.

<segment 2> À leur huitième saison dans la Ligue junior majeure du Québec, les Foreurs ont remporté pour la deuxième fois de leur histoire les séries éliminatoires.

<segment 3> Je suis fier de cette équipe

<segment 4> d'autant plus que cette formation a terminé en dernière position l'an dernier.

<segment 5> Une autre raison d'en être fier est que

<segment 6> mon fils Samuel joue pour cette équipe.

La première lecture de ce segment laisse apparaître deux problèmes linguistiques : le connecteur *d'autant plus que* et l'expression *une autre raison de*. Nous renvoyons le lecteur à (Névéal, 2002) pour l'étude détaillée du connecteur *d'autant plus que*. Nous expliquerons ici les grandes lignes de l'analyse nécessaires à la génération des segments 3 et 4. Quant à l'expression *une autre raison de*, nous en avons déjà fait l'analyse dans la partie 2.1.3.2..

Nous allons maintenant procéder à la génération du segment discursif 1 étape par étape : de la forme logique aux SDRSs (partie 3.2.1.1.) et des SDRSs aux structures communicatives des segments phrastiques (partie 3.2.1.2.).

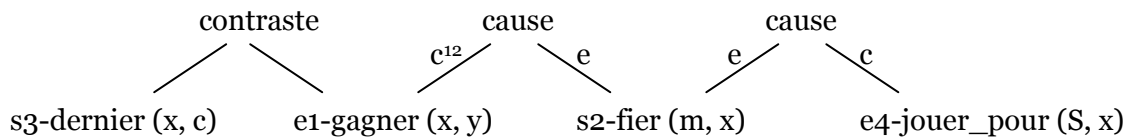
3.2.1.1. De la forme logique aux SDRSs

Au départ, nous avons la forme logique (1) suivante :

(1) $\exists e1, s2, s3, e4, x, y, m, S, c, t1, t2, t3, t4$ (

$e1\text{-gagner}(x, y) \wedge$	$t1 = 8^{\text{e}} \text{ saison de la Ligue} \wedge$
$s2\text{-fier}(m, x) \wedge$	$t2 = \text{pour la } 2^{\text{e}} \text{ fois de leur histoire} \wedge$
$s3\text{-dernier}(x, c) \wedge$	$t3 = \text{l'an dernier} \wedge$
$e4\text{-jouer_pour}(S, x) \wedge$	$t4 = \text{toujours} \wedge$
$\text{cause}(e1, s2) \wedge$	$e1 < \text{now} \wedge$
$\text{contraste}(e1, s3) \wedge$	$e1 \geq t1 \wedge$
$\text{cause}(e4, s2) \wedge$	$e1 \text{ O } t2 \wedge$
	$s2 \text{ O } \text{now} \wedge$
$x = \text{équipe des Foreurs de Val-d'Or} \wedge$	$s3 \text{ O } t3 \wedge$
$y = \text{séries éliminatoires} \wedge$	$e4 \text{ O } t4$
$m = \text{moi} \wedge$)
$S = \text{Samuel} \wedge$	
$c = \text{championnat} \wedge$	

Une vue d'ensemble de cette forme logique peut être la suivante :



Grâce à cette représentation, il apparaît de suite que l'état $s2$ ($s2\text{-fier}(m, x)$) est causé par deux événements : $e1\text{-gagner}(x, y)$ et $e4\text{-jouer_pour}(S, x)$.

La construction des SDRSs va se faire récursivement à partir des conditions contenues dans une DRS issue de la forme logique (1), soit la DRS suivante :

¹² 'c' représente la cause et 'e' représente l'effet.

	x, y, m, S, c, e1, s2, s3, e4, t1, t2, t3, t4
(2)	cond1 : x = équipe des Foreurs de Val-d'Or cond2 : y = séries éliminatoires cond3 : m = moi cond4 : S = Samuel cond5 : c = championnat cond6 : e1-gagner (x, y) cond7 : e1 < now cond8 : e1 ≥ t1 cond9 : e1 O t2 cond10 : s2-fier (m, x) cond11 : s2 O now cond12 : s3-dernier (x, c) cond13 : s3 O t3 cond14 : e4-jouer_pour (S, x) cond15 : e4 O t4 cond16 : cause (e1, s2) cond17 : contraste (e1, s3) cond18 : cause (e4, s2) cond19 : t1 = 8 ^e saison de la Ligue cond20 : t2 = pour la 2 ^e fois de leur histoire cond21 : t3 = l'an dernier cond22 : t4 = toujours

La DRS (2) contient tous les éléments de la forme logique (1). Nous avons numéroté les conditions pour y faire plus facilement référence par la suite.

Nous tenons à avertir le lecteur que nous ne prenons pas en considération les présuppositions dans les représentations en (S)DRS. Elles seront uniquement signalées dans les gloses sur les (S)DRSs.

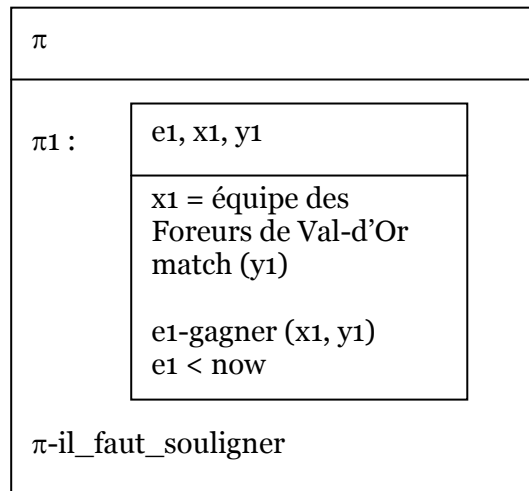
Comme il s'agit d'un début de discours, on peut recourir à un acte de langage, comme par exemple l'expression *j'aimerais souligner*, qui permet d'introduire le topique discursif et de commencer le discours. Nous proposons de représenter cet acte de langage dans la DRS sous la forme de π qui indique les conditions d'énonciation de la proposition principale représentée dans la DRS π_1 ¹³.

Maintenant, prenons comme topique 'la victoire de l'équipe des Foreurs de Val-d'Or', soit l'événement e1. En principe, il faudrait mettre dans la première DRS tout ce qui a trait à e1 or, comme nous commençons un discours, nous allons choisir de mettre dans la DRS π_1 uniquement les conditions cond6 (correspondant à *e1-gagner (x, y)*), cond7, cond1 (pour 'l'équipe des Foreurs de Val-d'Or') et un hyperonyme de 'séries éliminatoires', soit un nouveau référent de discours y1 qui correspond à 'un match'. Nous allons réserver les informations temporelles (comme t1 et t2) et 'les séries éliminatoires' (c'est-à-dire la spécification de y1) pour la suite du discours, en l'occurrence, dans la DRS π_2 . Cette démarche en deux temps (i.e. représenter un même événement dans deux DRSs distinctes)

¹³ À ce sujet, nous tenons à remercier Laurent Roussarie pour ses explications claires et sa pédagogie.

permet d'introduire le topique (dans π_1) et de rajouter des informations sur le topique par la suite (dans π_2). On anticipe une relation Particularisation entre π_1 et π_2 .

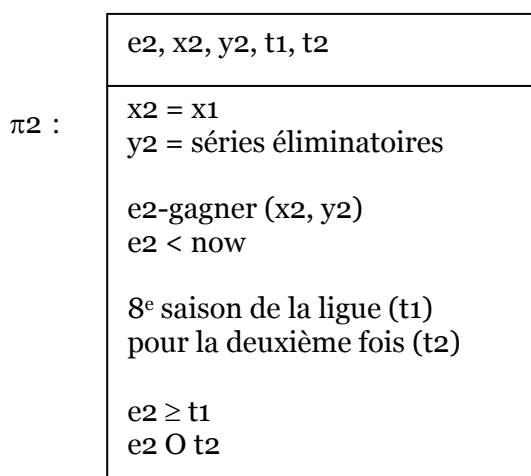
La DRS π_1 va représenter le topique du discours 'la victoire de l'équipe des Foreurs de Val-d'Or'. Il va être introduit par un acte de langage représenté par π 'il faut souligner' qui spécifie π_1 . Nous avons donc la DRS π_1 suivante :



Cette DRS représente l'événement 'équipe des Foreurs de Val-d'Or a gagné un match'. C'est dans la suite du texte, i.e. dans la DRS π_2 , que l'on va avoir des informations supplémentaires sur cet événement : de quel match s'agit-il ? Quand cette victoire s'est-elle produite ? etc.

Comme toutes les conditions relatives à e_1 dans la DRS (2) ne sont pas épuisées, on va les mettre dans une DRS étiquetée π_2 : il s'agit des conditions $\text{cond}_2, \text{cond}_8, \text{cond}_9, \text{cond}_{19}$ et cond_{20} . Cette DRS va donc représenter l'événement e_1 avec des informations complémentaires : l'événement 'équipe des Foreurs de Val-d'Or a remporté les séries éliminatoires, lors de la huitième saison de la ligue, pour la deuxième fois de leur histoire'. L'expression *pour la deuxième fois de leur histoire* présuppose que l'équipe a déjà gagné une fois les séries éliminatoires, nous ne le mentionnons pas dans la DRS.

Nous avons donc la DRS π_2 suivante :



Lors de la construction des DRSs, comme la DRS π_2 , il est nécessaire de créer une nouvelle variable pour référer à un référent de discours déjà présent dans une autre DRS. C'est ici le

cas pour le référent x_1 'l'équipe des Foreurs de Val-d'Or' que nous avons déjà dans la DRS π_1 . On ne peut pas le remettre dans la DRS π_2 : on crée alors x_2 , et on ajoute la condition $x_2 = x_1$.

Il y a une co-référence événementielle entre π_1 et π_2 : $e_1 = e_2$. Dans la DRS π_2 , il y a un ajout d'informations par rapport à la DRS π_1 : une spécification ('match' \Rightarrow 'séries éliminatoires') et des adjonctions correspondant à des circonstants ('8^e saison de la ligue' et 'pour la deuxième fois'). La relation Particularisation intervient donc entre π_1 et π_2 .

Par ailleurs, l'événement e_1 est lié à l'état s_2 par une relation causale *cause* (e_1, s_2) correspondant à la condition *cond16* qui n'est pas encore exprimée. Nous avons déjà une SDRS (avec Particularisation (π_1, π_2)) représentant e_1 , nous avons donc le choix entre deux relations de discours pour traduire la relation causale : la relation Résultat ou la relation Commentaire. Ces deux relations impliquent que la SDRS représentant e_1 est avant la représentation de l'état s_2 i.e. l'énoncé décrivant e_1 (le topique) sera devant celui décrivant s_2 dans le texte final.

Pour représenter l'état s_2 , nous proposons la DRS π_3 suivante :

π_3 :	s3, m, x3
	m = moi x3 = x1 s3-fier (m, x3) s3 O now

Toutes les conditions concernant s_2 sont dans la DRS π_3 (*cond1*, *cond3*, *cond10* et *cond11*) sauf la relation conceptuelle *cause* (e_1, s_2) (*cond16*). Pour traduire cette relation, on peut mettre la relation Résultat entre [π_1 - π_2] et π_3 .

Avec une SDRS avec une relation Résultat sous-jacente, nous aurons, par exemple, la réalisation syntaxique suivante :

- (3) À leur huitième saison dans la Ligue junior majeure du Québec, les Foreurs ont remporté pour la deuxième fois de leur histoire les séries éliminatoires.
Je suis fier de cette équipe.

Il est également possible de construire une SDRS avec la relation Commentaire : la DRS π_3 comprend alors les conditions supplémentaires *cause* (e, s_3) et l'équation anaphorique $e = e_1$ ¹⁴. La réalisation finale serait par exemple :

- (4) À leur huitième saison dans la Ligue junior majeure du Québec, les Foreurs ont remporté pour la deuxième fois de leur histoire les séries éliminatoires.
À cause de cela, je suis fier de cette équipe.

ou encore :

- (5) À leur huitième saison dans la Ligue junior majeure du Québec, les Foreurs ont remporté pour la deuxième fois de leur histoire les séries éliminatoires.
Cela me rend fier de cette équipe.

avec comme conditions supplémentaires dans π_3 : *f-cause* (e, s_3) et l'équation anaphorique $e = e_1$.

Si nous avons choisi 'je suis fier de l'équipe des Foreurs de Val-D'or' comme topique, nous aurions l'état s_2 dans une DRS π_1 et l'événement e_1 'l'équipe des Foreurs de Val-d'Or' a

¹⁴ Nous avons détaillé ce processus dans la partie 2.1.2..

remporté les séries éliminatoires, lors de la huitième saison de la ligue, pour la deuxième fois de leur histoire' dans une DRS π_2 : on aurait alors une relation Explication entre π_1 et π_2 avec 'je suis fier de l'équipe des Foreurs de Val-D'or' en première position dans le texte final. De plus, on aurait une autre relation Explication traduisant la relation *cause* (e_4, s_2) entre la DRS π_1 représentant 'je suis fier de l'équipe des Foreurs de Val-D'or' et une DRS π_3 représentant 'mon fils Samuel joue pour cette équipe'. Le texte final ressemblerait donc à :

(6) Je suis fier de l'équipe des Foreurs de Val-D'or.

À leur huitième saison dans la Ligue junior majeure du Québec, les Foreurs ont remporté pour la deuxième fois de leur histoire les séries éliminatoires.

Une autre raison d'en être fier est que mon fils Samuel joue pour cette équipe.

si 'je suis fier de l'équipe des Foreurs de Val-D'or' était le topique du discours.

Or, nous avons décidé que 'la victoire de l'équipe des Foreurs de Val-d'Or' serait le topique et nous choisissons de produire le discours (3).

L'événement e_1 est également lié à l'état s_3 par le biais de la relation *contraste* (e_1, s_3) correspondant à la condition non traduite *cond17* de la DRS (2). Cette relation va de pair avec la relation *cause* (e_1, s_2). En effet, ces deux relations indiquent la présence d'une argumentation contrastée. La conjugaison de ces relations conceptuelles est une relation assez particulière qui ne peut être traduite que par le biais de la relation de discours *D'autant_plus_que*, étudiée par Aurélie Névéol (dans (Névéol, 2002)). Elle propose la validation de la relation *D'autant_plus_que* suivante :

Soit π_r et π_p deux étiquettes de SDRS et k une DRS,

- Si $\pi_r: k_r, \pi_p: k_p, me(k) = e, me(k_r) = e_r, me(k_p) = e_p$ et
- Si *cause*(e, e_p) et *contraste*(e, e_r)

Alors si π_k étiquette k *D'autant_plus_que* (π_p, π_k) est valide.

La relation de discours *D'autant_plus_que* paraît donc la plus appropriée pour traduire la relation *contraste* (e_1, s_3) en accord avec la relation *cause* (e_1, s_2) et. Nous aurons donc une relation *D'autant_plus_que* entre la DRS π_3 et la DRS π_4 suivante :

$\pi_4 :$	<p>s_4, x_4, c, t_4</p> <hr/> <p>$x_4 = x_1$ $c = \text{championnat}$ l'an dernier (t_4)</p> <p>$s_4\text{-dernier}(x_4, c)$ $s_4 \text{ O } t_4$</p>
-----------	--

Dans cette DRS, nous avons mis tout ce qui concerne l'état s_3 , c'est-à-dire les conditions *cond1*, *cond5*, *cond12*, *cond13* et *cond21*.

La réalisation syntaxique de la relation *D'autant_plus_que* est le connecteur *d'autant plus que*. À partir du discours (3), nous obtenons donc le discours suivant :

(7) À leur huitième saison dans la Ligue junior majeure du Québec, les Foreurs ont remporté pour la deuxième fois de leur histoire les séries éliminatoires.

Je suis fier de cette équipe d'autant plus que cette formation a terminé en dernière position l'an dernier.

Nous devons revenir sur l'état s_2 car il est non seulement lié à e_1 mais il est également lié à l'événement e_4 par le biais d'une relation causale *cause* (e_4, s_2), condition *cond18* non traduite.

Pour représenter e_4 , il est possible de regrouper les conditions *cond1*, *cond4*, *cond14*, *cond15* et *cond22* pour construire la DRS suivante :

π_5' :	$e_5', s_5', S', x_5', t_5'$ <hr/> $S' = \text{mon fils Samuel}$ $x_5' = x_1$ $e_5'\text{-jouer_pour}(S', x_5')$ <i>cause</i> (e_5', s_5') $s_5' = s_3$ <i>toujours</i> (t_5') $e_5' \text{ O } t_5'$
------------	---

Nous avons mis dans la DRS π_5' tout ce qui a trait à l'événement e_4 dans la DRS (2), y compris la condition *cond18* *cause* (e_4, s_2). Il a fallu rajouter l'équation $s_5' = s_3$ car le référent de discours s_3 est déjà présent dans la DRS π_3 . Rappelons que l'état s_2 est causé par deux événements : e_1 et e_4 .

La DRS π_5' pourrait donner lieu au discours :

- (8) Je suis fier de l'équipe des Foreurs de Val-D'or aussi parce que mon fils Samuel joue pour cette équipe.

Une autre solution est de procéder à la réification de la cause, comme dans la DRS π_5'' suivante :

π_5'' :	$e_5'', s_5'', S'', x_5'', t_5'', f$ <hr/> $S'' = \text{mon fils Samuel}$ $x_5'' = x_1$ $e_5''\text{-jouer_pour}(S'', x_5'')$ <i>f-cause</i> (e_5'', s_5'') $s_5'' = s_3$ <i>toujours</i> (t_5'') $e_5'' \text{ O } t_5''$
-------------	--

Nous pourrions avoir, à partir de π_5'' , les discours suivants :

- (9) Que mon fils Samuel joue pour cette équipe est une autre raison de ma fierté.
(10) Que mon fils Samuel joue pour cette équipe cause aussi ma fierté.

Or, nous voulons générer le discours :

- (11) Une autre raison d'être fier de l'équipe des Foreurs de Val-d'Or est que mon fils Samuel joue pour cette équipe.

dans lequel l'événement dénotant l'effet est exprimé avant celui dénotant la cause, où l'antécédent de *raison* n'est pas encore exprimé et dans lequel une présupposition intervient au travers de l'adjectif *autre* (le même état s_2 est causé par deux événements : e_1 'la victoire

de l'équipe des Foreurs de Val-D'or' dans un premier temps, puis e4 'mon fils Samuel joue pour l'équipe des Foreurs de Val-D'or'). Nous avons étudié ce cas particulier dans la partie 2.1.3.2., et pour représenter ce phénomène, nous proposons la SDRS suivante :

π_5, π_6									
<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 20px;"></td> <td style="padding: 5px;">s_5, m_5, x_5, f</td> </tr> <tr> <td style="width: 20px; vertical-align: middle;">$\pi_5 :$</td> <td style="padding: 5px;"> $m_5 = m_3$ $x_5 = x_1$ $s_5\text{-fier}(m_5, x_5)$ $s_5 \text{ O now}$ $s_5 = s_3$ $f\text{-cause}(e, s_5)$ $e = ?$ </td> </tr> </table>		s_5, m_5, x_5, f	$\pi_5 :$	$m_5 = m_3$ $x_5 = x_1$ $s_5\text{-fier}(m_5, x_5)$ $s_5 \text{ O now}$ $s_5 = s_3$ $f\text{-cause}(e, s_5)$ $e = ?$	<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 20px;"></td> <td style="padding: 5px;">e_6, S, x_6, t_6</td> </tr> <tr> <td style="width: 20px; vertical-align: middle;">$\pi_6 :$</td> <td style="padding: 5px;"> $S = \text{mon fils Samuel}$ $x_6 = x_1$ $e_6\text{-jouer_pour}(S, x_6)$ $\text{toujours}(t_6)$ $e_6 \text{ O } t_6$ $e_6 = e$ </td> </tr> </table>		e_6, S, x_6, t_6	$\pi_6 :$	$S = \text{mon fils Samuel}$ $x_6 = x_1$ $e_6\text{-jouer_pour}(S, x_6)$ $\text{toujours}(t_6)$ $e_6 \text{ O } t_6$ $e_6 = e$
	s_5, m_5, x_5, f								
$\pi_5 :$	$m_5 = m_3$ $x_5 = x_1$ $s_5\text{-fier}(m_5, x_5)$ $s_5 \text{ O now}$ $s_5 = s_3$ $f\text{-cause}(e, s_5)$ $e = ?$								
	e_6, S, x_6, t_6								
$\pi_6 :$	$S = \text{mon fils Samuel}$ $x_6 = x_1$ $e_6\text{-jouer_pour}(S, x_6)$ $\text{toujours}(t_6)$ $e_6 \text{ O } t_6$ $e_6 = e$								
Commentaire-Égalité (π_5, π_6)									

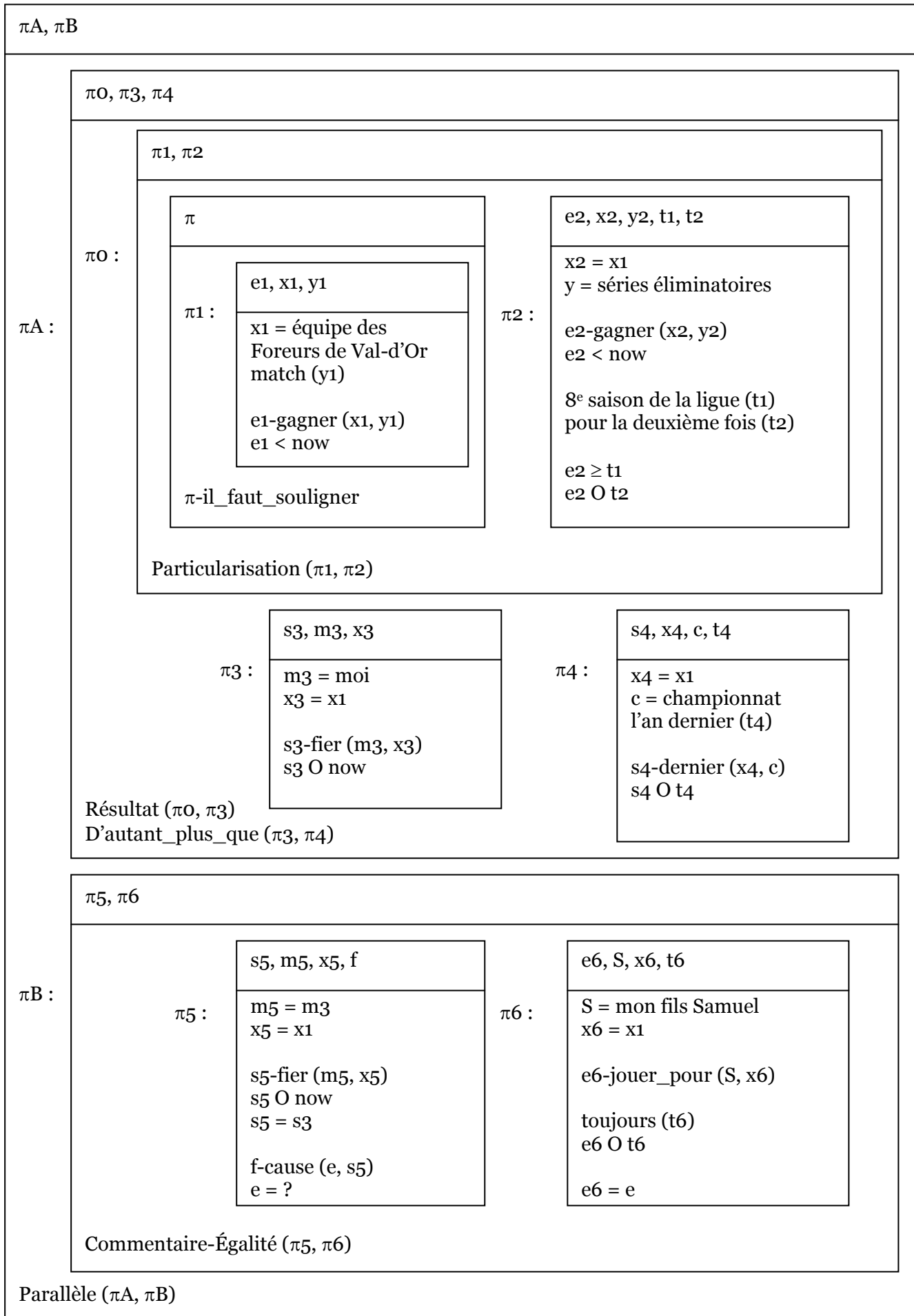
Cette SDRS résulte de la répartition des conditions de la DRS π_5 : nous avons mis toutes les conditions concernant l'événement e4 dans la DRS π_6 (cond1, cond4, cond14, cond15 et cond22) et tout le reste dans la DRS π_5 : la relation causale (cond18), la condition $s_5 = s_3$ et tout ce qui concerne l'état s2 'je suis fier de l'équipe des Foreurs de Val-D'or', représenté ici par le référent de discours s5.

Nous notons par le biais de la condition *f-cause* (e, s_5) qu'il existe un événement qui cause l'état s5. Cet événement n'est pas encore exprimé, il est en fait exprimé dans le groupe nominal cataphorique *la raison de* (condition $e = ?$). C'est dans la suite du texte, i.e. dans l'énoncé représenté dans π_6 , que la raison est exprimée linguistiquement (équation anaphorique $e_6 = e$).

La relation Commentaire entre π_5 et π_6 est justifiée car la condition cond18 apparaît dans la DRS π_5 (sous la forme de *f-cause* (e, s_5)) et elle est exprimée dans un groupe nominal (*la raison de*) et non dans un connecteur. D'autre part, la relation Égalité entre en jeu car elle met en relation deux DRSs représentant un même événement : l'événement qui cause la fierté du locuteur *est, correspond à, équivaut* à l'événement 'mon fils Samuel joue pour l'équipe des Foreurs de Val-D'or'.

Nous n'avons pas représenté dans la SDRS la présupposition qui intervient dans ce discours : l'adjectif *autre* présuppose qu'il y a déjà eu une raison pour laquelle 'je suis fier de l'équipe des Foreurs de Val-D'or'. Ce présupposé est tout de même déchiffrable au travers de la relation Parallèle entre les SDRSs π_A (regroupant π_1, π_2, π_3 et π_4) et π_B (regroupant π_5 et π_6), et, grâce à l'équation anaphorique $s_5 = s_3$ dans la DRS π_5 .

Toutes les conditions de la DRS (2) sont maintenant épuisées. Au final, nous obtenons pour le segment discursif 1 la SDRS suivante :



3.2.1.2. Des SDRSs aux structures communicatives

Maintenant que nous avons dégagé les SDRSs et les relations de discours de la forme logique, nous allons pouvoir nous intéresser aux segments phrastiques et à leurs structures communicatives.

Pour récapituler, nous avons :

- entre les segments 1 et 2 une relation Particularisation
- entre les segments [1-2] et 3 une relation Résultat
- entre les segments 3 et 4 une relation D'autant_plus_que
- entre les segments 5 et 6 une relation Commentaire-Égalité
- entre les segments [1-2-3-4] et [5-6] une relation Parallèle

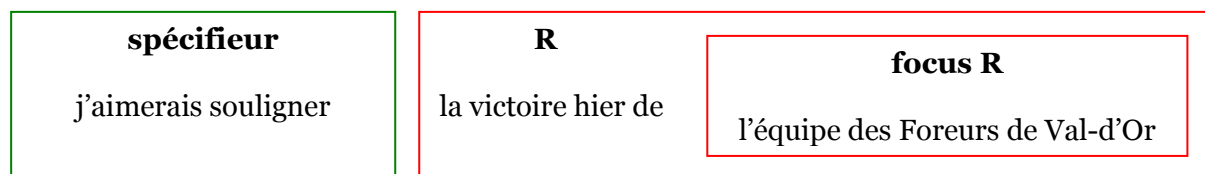
C'est à partir des SDRSs et des relations de discours déterminées précédemment que nous allons attribuer la structure communicative adéquate à chaque segment phrastique.

Nous pouvons remarquer que les DRSs contraignent déjà les réalisations syntaxiques des segments. La structure communicative d'une proposition est donc déjà guidée par sa représentation en DRT. Nous allons tout de même tenter de proposer des segments qu'il est possible et impossible de produire à partir d'une DRS et en fonction d'une structure communicative particulière.

Le segment phrastique 1 est assez particulier car, nous l'avons vu, il a pour fonction d'introduire le topique discursif. La DRS π_1 représentant la sémantique de ce segment comporte un acte de langage qui va donner lieu à un spécifieur modal.

Par ailleurs, nous avons vu qu'après un spécifieur modal, il est préférable d'avoir un segment rhématique comme 'la victoire hier de l'équipe des Foreurs de Val-d'Or' plutôt qu'une phrase simple avec une partition thème-rhème comme 'l'équipe des Foreurs de Val-d'Or a gagné hier'.

Nous proposons donc la structure communicative suivante pour le segment 1 :



● *j'aimerais souligner* est un spécifieur modal factuel : il donne une indication sur l'attitude du locuteur par rapport à ce qu'il dit, par rapport au contenu du message qu'il veut faire passer.

Il permet d'introduire le message dans le discours, il présente un nouvel événement 'la victoire de l'équipe des Foreurs de Val-d'Or' (le verbe *souligner* est un verbe performatif) qui correspond au topique du discours.

● *la victoire hier de l'équipe des Foreurs de Val-d'Or* est un segment rhématique. Ce segment est sémantiquement (mais non communicativement) équivalent à la phrase :

L'équipe des Foreurs de Val-d'Or a gagné hier.

● *l'équipe des Foreurs de Val-d'Or* est le focus rhématique.

Pour le segment 1, nous allons préférer générer la phrase

(12) J'aimerais souligner la victoire hier de l'équipe des Foreurs de Val-d'Or.

plutôt que la phrase :

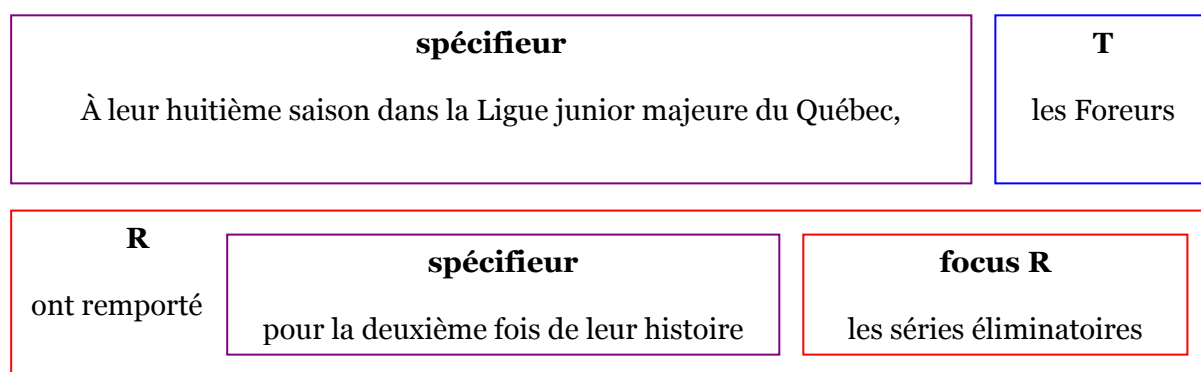
(13) J'aimerais souligner que l'équipe des Foreurs de Val-d'Or a gagné hier.

Puis, comme cela est décrit dans la partie 2.2.2.5., si la relation Particularisation entre en jeu, alors le focus rhématique du segment 1 (i.e. 'l'équipe des Foreurs de Val-d'Or') est repris en thème et le rhème (i.e. 'la victoire') est repris en rhème dans le segment suivant avec des circonstants. Cette règle nous empêche de générer un segment phrastique avec 'les séries éliminatoires' comme thème c'est-à-dire que l'on ne va pas générer le discours suivant :

- (14) a. J'aimerais souligner la victoire hier de l'équipe des Foreurs de Val-d'Or.
 b. À leur huitième saison dans la Ligue junior majeure du Québec, les séries éliminatoires ont été remportées par les Foreurs pour la deuxième fois de leur histoire.

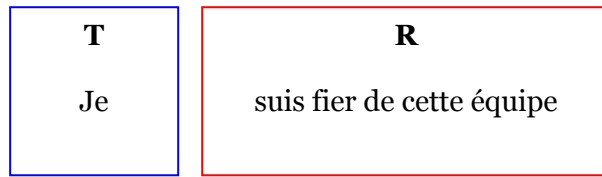
qui est très maladroit.

Nous aurons donc la structure communicative suivante pour le segment 2 :



- *À leur huitième saison dans la Ligue junior majeure du Québec* est un spécifieur circonstant factuel : il indique les circonstances (informations temporelles) de l'événement 'la victoire des Foreurs' décrit dans la phrase.
- *les Foreurs* est le thème : on parle de l'équipe des Foreurs et on en dit quelque chose. C'est aussi un élément donné, car il a déjà été énoncé dans le segment 1 sous la forme du syntagme nominal *l'équipe des Foreurs de Val-d'Or*.
- *ont remporté les séries éliminatoires* correspond au rhème de la phrase : c'est ce qu'on dit à propos des Foreurs. Il s'agit en partie d'un élément donné car on a appris dans le segment 1 qu'il y avait 'une victoire', cet événement est repris ici sous la forme du syntagme verbal *ont remporté*.
- *les séries éliminatoires* est le focus rhématique. On sait que 'l'équipe a gagné quelque chose' (*les Foreurs ont remporté* sont des éléments donnés) et on apprend ici de quoi il s'agit. Il s'agit d'une information nouvelle au même titre que les deux circonstants factuels.
- *pour la deuxième fois de leur histoire* est un spécifieur circonstant factuel : il donne des informations sur les circonstances de l'événement 'la victoire des Foreurs' (la fréquence des victoires de l'équipe). En particulier, il présuppose que l'équipe a déjà gagné une fois (nous ne nous en préoccupons pas ici : nous avons noté ce circonstant comme une information temporelle sur l'événement dans la DRS)

La relation Résultat n'impose aucune contrainte quant à la structure communicative des segments phrastiques mis en relation. Pour le segment 3, nous proposons la structure communicative suivante (guidée par la DRS π_3) :



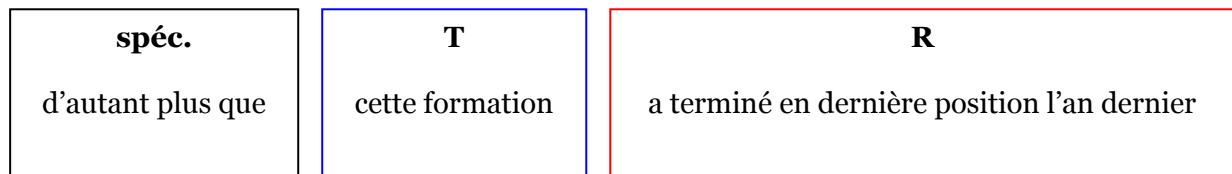
- *Je* est le thème : on parle du locuteur 'moi' (le locuteur parle de lui). C'est un élément donné.

Q : qu'y a-t-il à propos de moi ?

R : (au sujet de moi, il est dit que) 'moi' est fier de l'équipe.

- *suis fier de cette équipe* est le rhème et il comporte un élément donné 'cette équipe' (sous-topique discursif).

Nous n'avons pas étudié la relation D'autant_plus_que sous l'angle de la structure communicative donc nous allons mettre une structure thème-rhème classique pour le segment 4 (d'autant plus que la DRS π_4 ne permet pas d'autres structures communicatives possibles) :

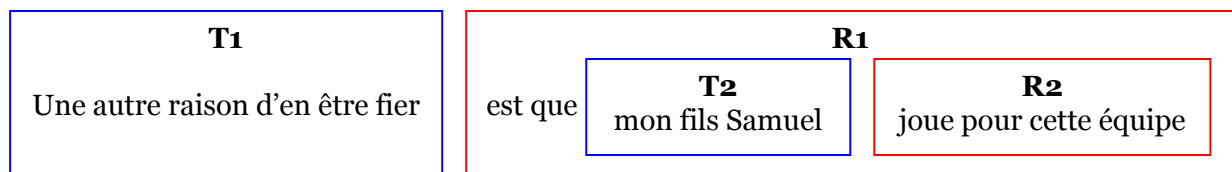


- *d'autant plus que* est un connecteur discursif qui traduit la relation D'autant_plus_que.

- *cette formation* est le thème de la phrase : le locuteur parle de l'équipe des Foreurs.

- *a terminé en dernière position l'an dernier* est ce qui est à propos du thème.

Entre les segments 5 et 6, il y a une relation Commentaire-Égalité : cela impose une structure communicative particulière. En effet, la cause est explicitée dans le segment 6, plus précisément dans la partie rhématique du segment 5 en tant que partition secondaire. Concrètement, voici la structure communicative proposée :



- *une autre raison d'en être fier* est le thème : on parle d'un événement qui est la raison, la cause de la fierté du locuteur pour l'équipe ('je suis fier de l'équipe' est un élément donné). De plus, l'adjectif *autre* présuppose qu'il a déjà eu une raison exprimée antérieurement ce qui est exprimé par la relation Parallèle entre les segments [1-2-3-4] et [5-6].

- *mon fils Samuel* est le thème secondaire de la phrase : on parle du fils du locuteur. Notons que le locuteur présuppose que l'interlocuteur connaisse l'existence de Samuel ou lui apprend par la même occasion.

- *joue pour cette équipe* constitue le rhème secondaire de la phrase, on dit de Samuel qu'il joue dans l'équipe des Foreurs. *Cette équipe* est un élément donné : c'est le sous-topique discursif de tout le texte.

3.2.2. Génération du segment discursif 2

Nous allons à présent tenter de générer le segment discursif 2 rappelé ici :

<segment 7> Il faut aussi souligner la performance de tous ces jeunes
 <segment 8> et peut-être plus particulièrement celle de Simon Gamache,
 <segment 9> qui a récolté au moins un point dans chacun des 21 matchs éliminatoires pour compléter la série avec un record de 57 points.
 <segment 10> Il a également remporté le trophée Guy Lafleur.

Le segment discursif 2 comporte des particularités par rapport au segment discursif 1, notamment, nous allons le voir, des relations Élaboration qui imposent des structures communicatives particulières.

3.2.2.1. De la forme logique aux SDRSs

Soit la forme logique (15) :

$$(15) \quad \exists e1, e2, e3, e4, x, g, p, t \quad ($$

$$e1\text{-performance}(x) \wedge$$

$$e2\text{-performance}(g) \wedge$$

$$e3\text{-récolter}(g, p) \wedge$$

$$e4\text{-remporter}(g, t) \wedge$$

$$x = \text{équipe des Foreurs de Val-d'Or} \wedge$$

$$g = \text{Simon Gamache} \wedge$$

$$p = \text{points} \wedge$$

$$t = \text{trophée Guy Lafleur} \wedge$$

$$e1 < \text{now} \wedge$$

$$e2 < \text{now} \wedge$$

$$e3 < \text{now} \wedge$$

$$e4 < \text{now} \quad)$$

Nous allons construire les SDRSs de façon récursive à partir des conditions contenues dans une DRS issue de la forme logique (15), soit la DRS suivante :

	x, g, p, t, e1, e2, e3, e4
(16)	cond1 : x = équipe des Foreurs de Val-d'Or cond2 : g = Simon Gamache cond3 : p = points cond4 : t = trophée Guy Lafleur cond5 : e1-performance (x) cond6 : e2-performance (g) cond7 : e3-récolter (g, p) cond8 : e4-remporter (g, t) cond9 : e1 < now cond10 : e2 < now cond11 : e3 < now cond12 : e4 < now

Rappelons que nous avons choisi comme topique discursif 'la victoire de l'équipe des Foreurs de Val-d'Or'. Dans la DRS (16), nous trouvons la condition cond1 correspondant à l'entité 'l'équipe des Foreurs de Val-d'Or' mais nous n'avons pas d'événement 'victoire'. Pour conserver une certaine progression thématique, nous pouvons choisir de représenter l'événement 'la performance de l'équipe des Foreurs de Val-d'Or' impliquant l'entité 'l'équipe des Foreurs de Val-d'Or' : il va s'agir d'un « sous-topique ». Ce nouveau topique peut être introduit par le biais d'un acte de langage que nous représentons par π 'il faut souligner' pour spécifier π_7^{15} . Nous avons donc la DRS π_7 suivante :

π					
$\pi_7 :$	<table border="1"> <tr> <td>e7, x7</td> </tr> <tr> <td>x7 = x1</td> </tr> <tr> <td>e7-performance (x7)</td> </tr> <tr> <td>e7 < now</td> </tr> </table>	e7, x7	x7 = x1	e7-performance (x7)	e7 < now
e7, x7					
x7 = x1					
e7-performance (x7)					
e7 < now					
π -il_faut_souligner					

Nous avons mis dans cette DRS toutes les informations relatives à e1 présentes dans la DRS (16) soit les conditions cond1, cond5 et cond9.

Dans la DRS (16), on trouve un autre événement dénotant une autre performance : la performance d'un des joueurs en particulier. L'événement e2 constitue un autre sous-topique, le locuteur parle de la performance l'équipe puis de celle d'un des joueurs. Nous allons procéder de la même manière que pour e1. Pour représenter e2, nous aurons donc la DRS π_8 :

¹⁵ La numérotation des DRSs et des variables correspond à la suite de la description du segment 1.

π	
$\pi 8 :$	$e 8, g$
	$g = \text{Simon Gamache}$ $e 8\text{-performance}(g)$ $e 8 < \text{now}$
$\pi\text{-il_faut_souligner}$	

Nous retrouvons dans cette DRS les conditions cond2, cond6 et cond10 relatives à l'événement e2.

Entre les DRSs $\pi 7$ et $\pi 8$, il y a une relation Élaboration car $\pi 8$ représente un sous-événement de celui décrit dans $\pi 7$: le locuteur parle de la performance de l'équipe puis de celle d'un des membres de l'équipe.

Concernant le référent de discours g 'Simon Gamache' dans la DRS (16), on trouve un autre événement : e3. Cet événement sera représenté par la DRS $\pi 9$ suivante :

$\pi 9 :$	$e 9, g 2, p$
	$g 2 = g$ $p = \text{points}$ $e 9\text{-récolter}(g 2, p)$ $e 9 < \text{now}$

De la même manière que précédemment, nous mettons tout ce qui a trait à l'événement e3 de la DRS (16) dans la DRS $\pi 9$ (i.e. les conditions cond2, cond3, cond7 et cond11). Cet événement est un sous-événement de e8 (de la DRS $\pi 8$) : il explicite, donne des informations sur 'la performance de Simon Gamache' traduite dans la DRS $\pi 8$. Nous pouvons donc mettre une relation Élaboration entre $\pi 8$ et $\pi 9$.

Un autre événement mettant en jeu 'Simon Gamache' est l'événement e4 que nous pouvons représenter par la DRS $\pi 10$:

$\pi 10 :$	$e 10, g 3, t$
	$g 3 = g$ $t = \text{trophée Guy Lafleur}$ $e 10\text{-gagner}(g 3, t)$ $e 10 < \text{now}$

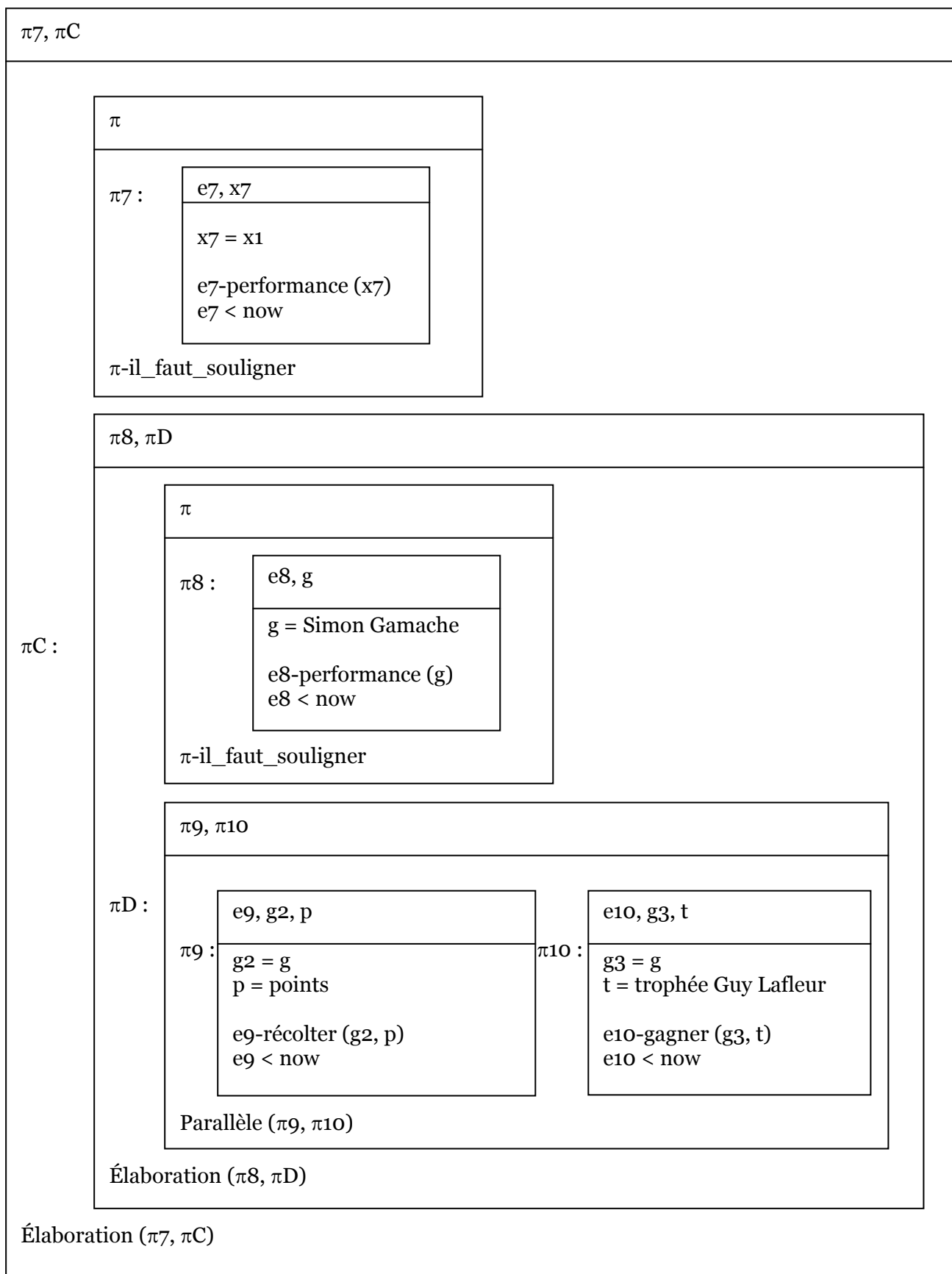
Ici, nous retrouvons les conditions cond2, cond4, cond8 et cond12.

L'événement e10 est un sous-événement de l'événement e8 'la performance de Simon Gamache'. Nous aurons donc ici une relation Élaboration entre $\pi 8$ et $\pi 10$.

Par ailleurs, comme $\pi 9$ et $\pi 10$ décrivent le même référent i.e. 'Simon Gamache' on pourra mettre une relation Parallèle entre les deux DRSs : Simon Gamache a récolté des points et il a

également remporté le trophée Guy Lafleur. Cet ensemble d'événements concernant 'Simon Gamache' constitue l'élaboration de 'la performance de Simon Gamache' décrit dans $\pi 8$.

Au final, nous obtenons pour le segment discursif 2 la SDRS suivante :



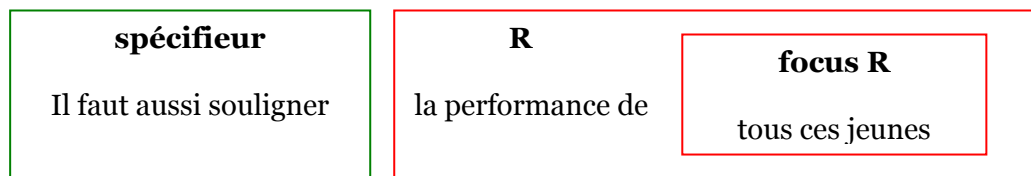
3.2.2.2. Des SDRS aux structures communicatives

Maintenant que nous avons construit toutes les DRSs nécessaires, nous allons pouvoir déterminer les structures communicatives adéquates pour assurer la cohérence du segment discursif 2.

Pour résumer, nous avons les relations de discours suivantes :

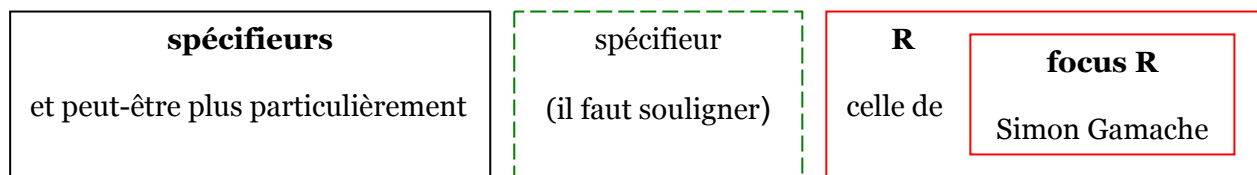
- entre les segments 7 et 8 une relation Élaboration ;
- entre les segments 9 et 10 une relation Parallèle ;
- entre les segments 8 et [9-10] une relation Élaboration.

Le segment phrastique 7 est, comme le segment 1, particulier car il a pour fonction d'introduire un (sous-) topique discursif. Pour les mêmes raisons qui nous ont fait proposer la structure communicative du segment 1, nous proposons l'organisation suivante pour le segment 7 :



- *Il faut aussi souligner* permet d'introduire un nouvel événement impliquant un référent connu. Il s'agit d'un spécifieur modal factuel qui indique l'attitude du locuteur par rapport à ce qu'il dit : ici, il veut mettre en avant un nouvel événement.
- *la performance de tous ces jeunes* est un segment rhématique. Il introduit un nouvel événement 'performance' au sujet d'un élément connu 'tous ces jeunes' i.e. 'l'équipe'.
- *tous ces jeunes* constitue le focus rhématique.

La relation Élaboration entre la DRS π_7 et la DRS π_8 est une Élaboration objet : 'Simon Gamache' est une élaboration de 'tous ces jeunes'. On devrait donc retrouver 'Simon Gamache' en thème dans le segment 8. Or, il faut prendre en compte le fait que l'on parle de 'la performance de Simon Gamache', il s'agit d'un nouveau topique. On aura donc la même structure communicative que pour le segment 7, soit :



- *et peut-être plus particulièrement* constitue un spécifieur communicatif externe c'est-à-dire qu'il s'agit d'un connecteur factuel qui fonctionne entre le segment 6 et le segment 7. Il lexicalise la relation Élaboration entre les deux segments (*plus particulièrement*) et les coordonne (*et*).
- ce segment est elliptique car il manque le spécifieur modal *il faut souligner* qui permet d'introduire le segment rhématique 'la performance de Simon Gamache'.
- *celle de* est une reprise anaphorique de *la performance de*. 'La performance de Simon Gamache' est le rhème de la phrase : il s'agit d'un segment rhématique par le biais duquel on présente un nouveau référent 'Simon Gamache', pour en parler par la suite.
- *Simon Gamache* est le focus rhématique.

Les segments 9 et 10 élaborent l'événement décrit dans le segment 8, soit 'la performance de Simon Gamache'. On va donc retrouver dans ces deux segments le même thème, 'Simon Gamache' correspondant au focus rhématique du segment 8. Nous aurons donc les deux structures communicatives suivantes :

T	R
qui	a récolté au moins un point dans chacun des 21 matchs éliminatoires pour compléter la série avec un record de 57 points.

- *qui* est le thème et il est connu : il est la reprise anaphorique du référent 'Simon Gamache'.
- ce qu'on dit du thème est le rhème *a récolté au moins un point dans chacun des 21 matchs éliminatoires pour compléter la série avec un record de 57 points*.

T	R
Il	a également remporté le trophée Guy Lafleur.

- *il* est le thème et il est un élément connu car il s'agit d'un pronom anaphorique pour 'Simon Gamache'.
- on dit de 'Simon Gamache' qu'*il a remporté le trophée Guy Lafleur* : le rhème de la phrase.

La relation Parallèle entre les segments 9 et 10 justifie l'adverbe *également* dans le segment 10. En effet, quand l'événement 'Simon a remporté le trophée Guy Lafleur' est énoncé, il est connu que Simon a déjà fait quelque chose, en l'occurrence, l'événement décrit dans le segment 9 'Simon a récolté des points'.

La règle sur la relation Élaboration interdit la génération de discours comme :

- (17) a. Il faut signaler la performance de Simon Gamache.
 b. Au moins un point dans chacun des 21 matchs éliminatoires a été récolté par lui.
 c. Il a également remporté le trophée Guy Lafleur.

Pour produire un discours le plus cohérent possible, il faut que le thème, introduit en focus rhématique dans (17a), soit repris dans l'énoncé (17b). Ainsi, la progression thématique respectera les conditions imposées par la relation Élaboration objet.

Par ailleurs, le discours (17) n'est pas cohérent car on passe de 'Simon Gamache' dans (17a) à des 'points récoltés' dans (17b) pour revenir à 'Simon Gamache' dans (17c). Et la transition (17b)-(17c) est maladroitement parce que l'on ne voit pas clairement le parallèle (traduit par l'adverbe *également* dans (17c)) entre les deux énoncés. Pour respecter la relation Parallèle, il faut que le thème soit identique dans les deux segments phrastiques i.e. avoir le discours :

- (18) a. Il faut signaler la performance de Simon Gamache.
 b. Il a récolté au moins un point dans chacun des 21 matchs éliminatoires pour compléter la série avec un record de 57 points.
 c. Il a également remporté le trophée Guy Lafleur.

qui correspond au discours original.

Conclusion troisième partie

Nous avons pu observer ici que le choix du topique est important car c'est lui qui guide la structuration du discours et donc également la structure communicative des segments phrastiques. Du choix du propos du discours à générer dépendent la structure du discours et la structure informationnelle des énoncés. Le topique discursif implique donc le respect de certaines règles que nous avons énoncées dans la partie 2.

À travers deux illustrations, nous avons pu constater que des informations sur la structure communicative permettaient de produire des discours plus cohérents et d'éviter des discours bancals. L'intégration de telles informations dans un système de génération de textes est donc tout à fait pertinente si l'on veut garantir la cohérence du discours que l'on veut générer.

Conclusion générale

La génération automatique de textes est un processus très complexe et si l'on veut aboutir à des résultats satisfaisants, i.e. à des discours cohérents, il faut enrichir le processus d'informations communicatives.

Notre travail avait un double objectif : regarder dans quelle mesure une relation de discours pouvait imposer une progression thématique particulière dans un discours cohérent et voir comment des informations communicatives (comme la notion de topique et les contraintes dégagées précédemment) pouvaient intervenir dans un système de génération de textes et particulièrement comment elles pouvaient enrichir la structuration de document afin de générer des discours cohérents.

Nous avons proposé certaines règles qui permettent la prise en compte de la structure informationnelle, comme le topique discursif ou la partition thème~rhème, dans le processus de génération et plus précisément lors de la planification de texte. Nous avons également démontré que grâce à des renseignements sur la structure communicative (et encore nous n'avons pas exploité toutes les oppositions communicatives possibles pour définir la structure informationnelle d'une phrase) on pouvait générer des discours cohérents et rejeter des discours maladroits.

Il ne s'agit ici que d'un travail préliminaire car nous avons abordé beaucoup de problèmes et soulevé de nombreuses questions, dont : comment déterminer rigoureusement un topique de discours ? Peut-on vraiment associer à une relation de discours une, et une seule, progression thématique particulière ?

Il s'agit également d'un travail théorique et il faut se demander dans quelle mesure les règles et/ou les informations communicatives peuvent être implémentées dans un système de génération. Une des questions qui se posent est précisément : comment insérer des informations communicatives dans un formalisme comme une SDRS ?

Par ailleurs, il reste de nombreux points théoriques sur la structure communicative à éclaircir et à explorer. Il s'agit d'un sujet très vaste qui a donné lieu à de nombreux travaux mais les auteurs restent flous quant à une définition précise de la structure communicative. Et si I. Mel'čuk fait une description minutieuse de la structure communicative de la phrase dans son dernier ouvrage (Mel'čuk, 2001), il ne rend pas compte du fait que celle-ci dépend du contexte discursif dans lequel la phrase se trouve comme le fait K. Lambrecht dans (Lambrecht, 1994). Une véritable étude sur la structure communicative serait, selon nous, une étude combinant l'approche de I. Mel'čuk et celle de K. Lambrecht.

L'étude approfondie de la structure communicative peut fournir de nombreuses informations linguistiques qui permettraient d'améliorer certains systèmes de Traitement Automatique des Langues et notamment, nous l'avons vu dans le présent travail, des systèmes de génération de textes. La structure communicative constitue à elle seule des ressources linguistiques non négligeables que nous n'avons pas pu aborder dans le cadre de notre travail.

Annexe 1

Exemples de corpus illustrant le cas n°1 : l'effet est nouveau

- la cause est en rouge
- l'effet est en bleu

Elle lui apportera un peu d'argent. C'est pourquoi, la veille de Noël, en s'en allant à la ville, il se disait qu'il n'était pas plus malheureux qu'un autre.

Mme Laizy se fit, cette année-là, voler par son fermier. Elle jura qu'elle ne mangerait plus que du pain et du fromage : c'est pourquoi on lui vint en aide.

Mon frère Gédéon naquit l'année où commença la guerre avec les Français. C'est pourquoi père voulut le nommer Gédéon, qui est un nom guerrier.

Le passé n'est que le présent devenu invisible et muet ; c'est pourquoi ses lueurs et ses murmures sont infiniment précieux.

On a avantage, au point de vue du rendement, à éviter les pertes de chaleur par les parois de tels fours ; c'est la raison pour laquelle beaucoup d'appareils sont à doubles parois

Couthon envoyait en masse les prévenus à la guillotine en caressant un lapin blanc dont il ne pouvait se dessaisir. C'est la raison pour laquelle il est volontiers végétarien.

Nous voyons par là que le personnel d'exécution nécessaire à la production n'a nullement besoin d'être nombreux : c'est la raison pour laquelle on voit si peu de personnel dans les usines modernes de produits alimentaires

Naturellement, j'aimerais te dire ce qui m'arrive, mais je ne pourrais te parler que de choses indifférentes. C'est la raison pour laquelle nous devons renoncer à nous voir.

L'industrie des pâtes consomme surtout des résineux ; c'est la raison pour laquelle elle fait un appel à l'importation

dans le cas d'un accident grave entraînant un danger d'irradiation radioactive qui heureusement ne s'est jamais produit jusqu'à maintenant. C'est la raison pour laquelle les grands réacteurs atomiques sont encore aujourd'hui construits dans des régions de faible densité de population

Ce dernier, en effet, ne donne qu'un seul spectre dans lequel est concentrée toute l'énergie lumineuse. C'est la raison pour laquelle, en astronomie, on préfère fréquemment utiliser un spectroscopie à prisme.

En tous cas, il est certain que la plupart des algues supportent mal une lumière trop intense ; c'est la raison pour laquelle, en Méditerranée, la flore médiolittorale et celle des niveaux élevés de l'infralittoral sont constituées surtout d'espèces à développement hivernal

Annexe 2

Exemples de corpus illustrant le cas n°2 : l'effet est donné dans le contexte linguistique antérieur

- la cause est en rouge
- l'effet est en bleu
- l'antécédent de l'effet est souligné

Il était triste et m'a confié la raison de cette tristesse : sa mère est voyante.

il a suffi d'examiner dans le détail [...] les principes que ce sentiment invoque et la raison de ces principes, pour faire une découverte singulière : c'est que nous ne prenons aujourd'hui contact avec les lettres et le langage même

le désespoir de Mesnard, qui en a été malade pendant un an, sans soupçonner la raison de sa maladie.

J'arrivai en larmes au rendez-vous et Robin, sans connaître la raison de mon désarroi, parvint à me calmer.

La Mousson pleure rarement et ne connaît jamais la raison de ses larmes.

Reine a cru devoir mettre fin à ses jours, sans expliquer la raison de son geste.

Il y avait des partants, des arrivants, des revenants et j'étais curieux de la raison de ce flux et reflux dont virginie ne s'inquiétait pas

j'ai vu cette randonnée par les montagnes [...] et les gens dans leurs étoffes noires ont peut-être compris la raison de cet équipage

en me demandant quel était ce grand bruit. [...] et je cherchai du regard la cause de ce vacarme

Annexe 3

Exemples de corpus illustrant le cas n°3 : l'effet fait référence à la situation d'énonciation

- la cause est en rouge
- l'effet est en bleu

J'ai trouvé la raison de **mon extrême douleur** : **son corps est brouillé avec ses rêves**.

Telle était la raison de **ces regards haineux**, de **ces mouvements de recul**, de **ces chuchotements**

C'est la raison de **ma visite**.

Je connus bientôt la raison de **ce calme relatif**.

Il respire profondément, jusqu'à oublier la raison de **sa venue**

Papa semblait deviner la raison de **ma présence** sans oser se l'avouer trop ouvertement.

Puis il s'en est allé du côté du village, sans m'avoir éclairé sur la raison de **sa visite**.

elle se refuse à me donner la raison de **cette porte close du matin**.

Je lui demandai avec impatience la raison de **ces pleurs** ;

Madame Rollin me prit avec embarras sur ses genoux et me demanda la raison de **mes larmes**

On se demandait la raison de **cette exubérance**, le pourquoi de **ces confidences qu'elles se glissaient à l'oreille**, la cause de **ces regards vifs jetés de côté** avec un mouvement de tête qui agitait leurs boucles blondes et brunes.

Annexe 4

Exemples de corpus illustrant le cas n°4 : l'effet fait référence à des connaissances communes

- la cause est en rouge
- l'effet est en bleu

Je dois à mon éducation catholique **un sur-moi fortement développé** : c'est là la raison de **mon puritanisme et de la déficience de mon narcissisme**.

Je suis sûr qu'**il en a souffert** et c'est sans doute la raison de **ses ardeurs politiques et de sa froide austérité**.

Mais **on m'entend, on m'entend...** C'est la raison de **ma mauvaise réputation...**

Annexe 5

Discours exprimant deux événements causant le même effet

Notons que certains extraits de corpus sont très réduits et ne permettent donc pas une étude approfondie du contexte linguistique, notamment de trouver l'effet.

- la cause 1 est en rouge
- la cause 2 est en vert
- l'effet est en bleu

ce qui **le retient d'y entrer**, c'est qu'**il est riche**. Mais j'ai découvert une autre raison, qui peut-être est la principale : c'est qu'**il manque d'imagination**.

-pourquoi ?

Il répondit simplement, directement :

-parce que **tu courrais à ton malheur avec lui**.

-oh, fit-elle comme amusée, et à demi incrédule.

-il y a une autre raison, dit-il, lui saisissant la main.

-qui que c'est ?

-tu n'es pas ma sœur, heureusement, et tu me dis beaucoup.

Or, **Carmen était devenu mon meilleur camarade**. Non seulement à cause de **notre proche voisinage et de ses nombreuses confidences**, mais une autre raison devait encore y aider.

Une autre raison, outre **la jeunesse même de l'industrie de l'azote**, explique que **la production des engrais azotés reste encore aujourd'hui à un niveau nettement inférieur**.

Est-ce qu'il y a une autre raison, que celle de **rejoindre ses parents**, qui **fait du Havre, pour elle, cette obsession insensée** ?

Annexe 6

Structure globale du texte

Afin d'avoir une vue d'ensemble de la structure générale du texte des « Foreurs de Val-D'or », nous proposons ici un schéma représentant les DRSs et les relations de discours en jeu dans le texte.

Nous rappelons le texte ci-dessous :

M. Claude Duplain (Portneuf, Lib.) :

Monsieur le Président, j'aimerais souligner la victoire hier de l'équipe des Foreurs de Val-d'Or.

À leur huitième saison dans la Ligue junior majeure du Québec, les Foreurs ont remporté pour la deuxième fois de leur histoire les séries éliminatoires.

Je suis fier de cette équipe d'autant plus que cette formation a terminé en dernière position l'an dernier.

Une autre raison d'en être fier est que mon fils Samuel joue pour cette équipe.

Il faut aussi souligner la performance de tous ces jeunes et peut-être plus particulièrement celle de Simon Gamache, qui a récolté au moins un point dans chacun des 21 matchs éliminatoires pour compléter la série avec un record de 57 points.

Il a également remporté le trophée Guy Lafleur.

La performance de Maxime Daigneault est aussi à noter.

Il devient le premier gardien de but recrue de 16 ans à conduire sa formation à la conquête de la Coupe du président.

Félicitations à l'équipe, aux entraîneurs et à tous ceux qui ont participé à cette victoire.

Les relations de discours sont représentées par les flèches suivantes :

Explication 

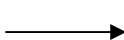
Résultat 

Élaboration 

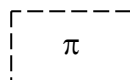
Particularisation 

Parallèle 

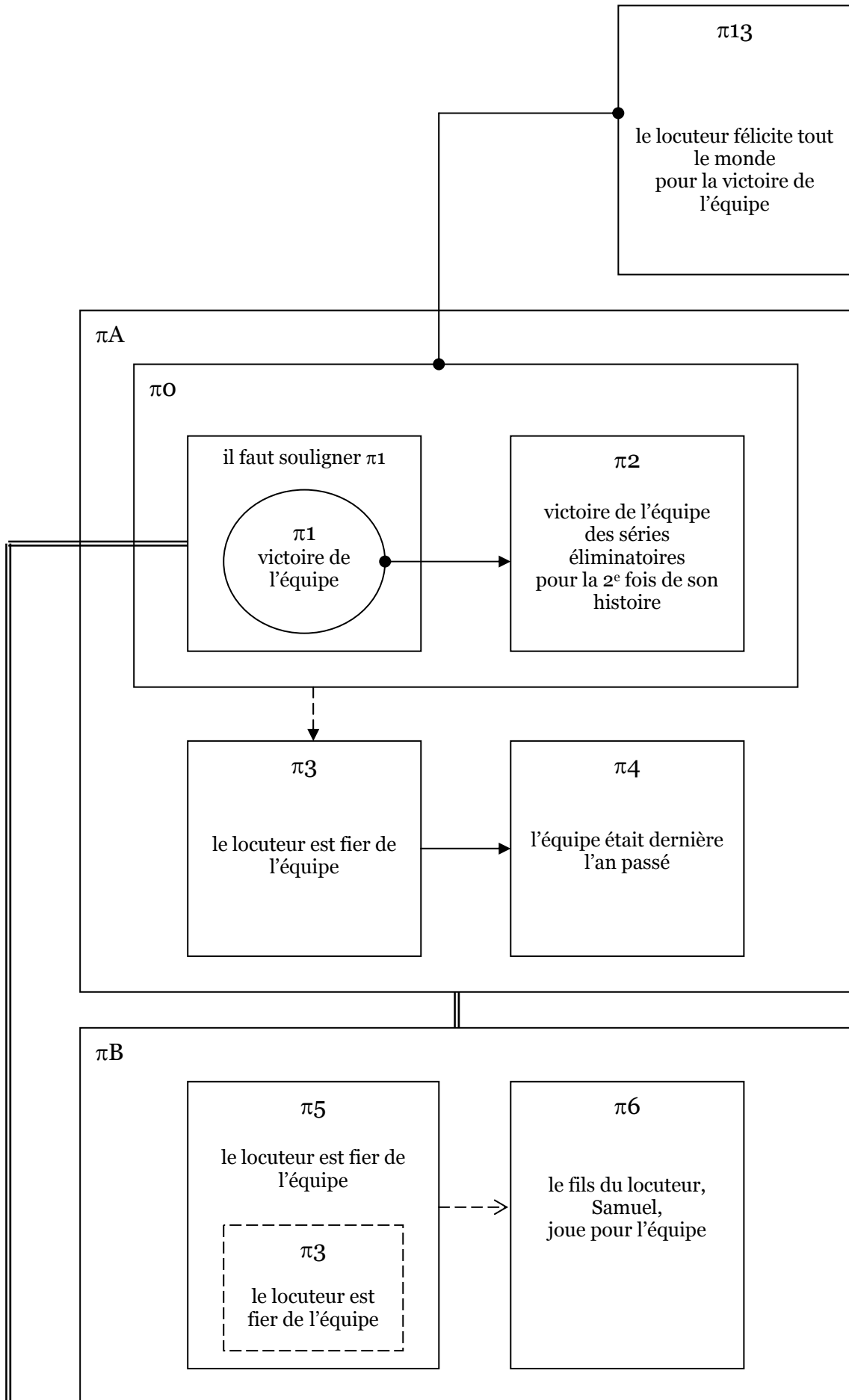
Commentaire 

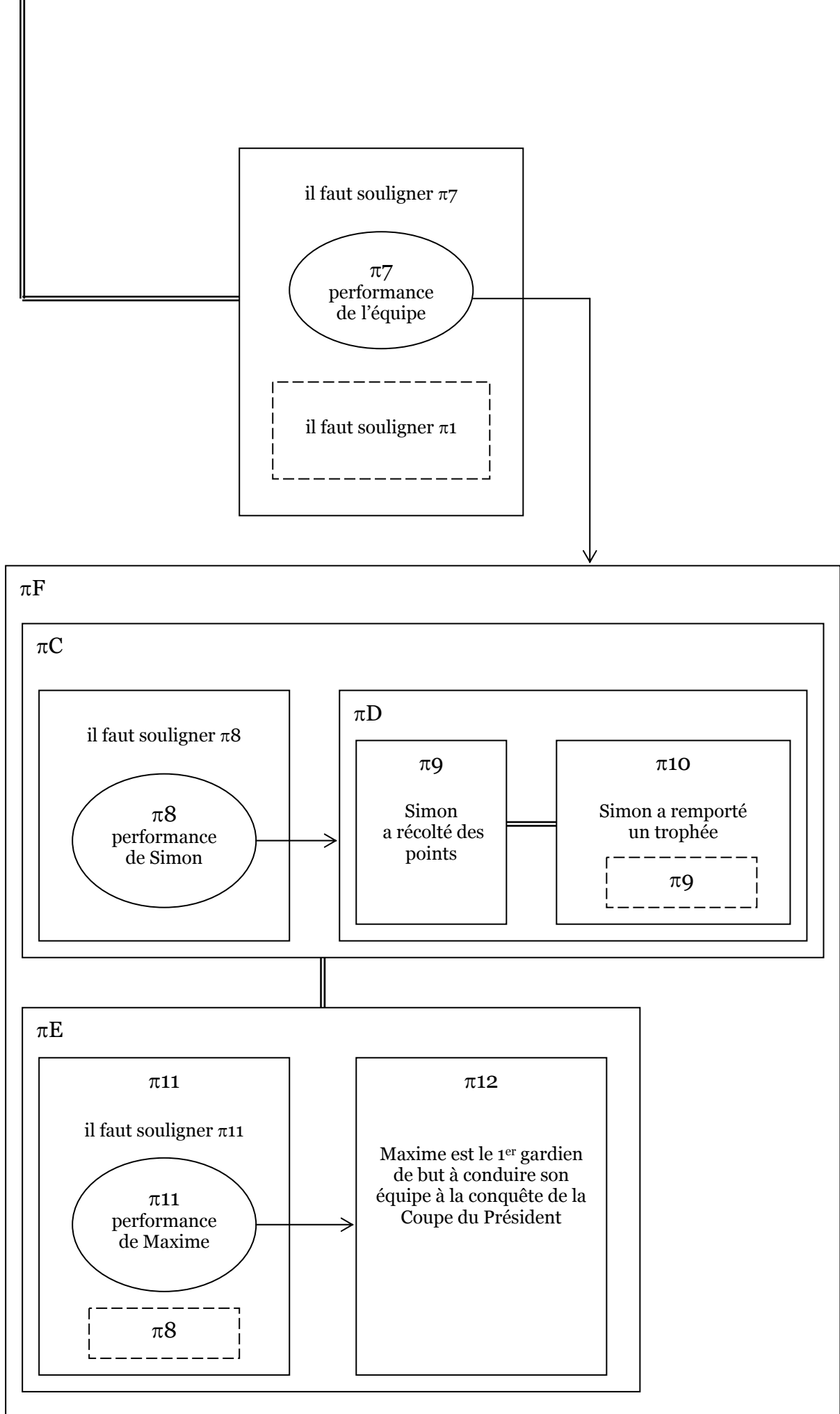
D'autant_plus_que 

Les boîtes en pointillés



correspondent à des DRSs présupposées.





Références bibliographiques

- Asher, Nicholas (1993). *Reference to Abstract Objects in Discourse*, Kluwer Academic Publishers, Dordrecht.
- Asher, Nicholas (1996). « L'interface pragmatique-sémantique et l'interprétation du discours », in *Langages* 123, pp. 30-50.
- Asher, Nicholas et Lascarides, Alex (1994). « Intentions and information in discourse », in *Proceedings of the 32nd annual meeting of the Association for Computational Linguistics*, Las Cruces, Mexico.
- Asher, Nicholas et Lascarides, Alex (2002). *Logics of Conversation*, CUP.
- Corblin, Francis (2002). *Représentation du discours et sémantique formelle*, P.U.F..
- Danlos, Laurence (1988). « Connecteurs et relations causales », in *Langue Française* 77, pp. 92-127.
- Danlos, Laurence (1998). « G-TAG : un formalisme lexicalisé pour la génération de textes inspiré de TAG », in *TAL* 39 (2), pp. 7-33.
- Danlos, Laurence (2000). « G-TAG : A lexicalized formalism for text generation inspired by Tree Adjoining Grammar », in Anne Abeillé et Owen Rambow (éd.), *Tree Adjoining Grammar : formalisms, linguistics analysis and processing*, CSLI Publications, Stanford, pp. 343-370.
- Danlos, Laurence et El Ghali, Adil (2002). « A complete integrated NLG system using AI and NLU tools », in *Proceedings of COLING'02*, Taipei, Taïwan, pp. 211-217.
- Danlos, Laurence et Roussarie, Laurent (2000). « La génération automatique de textes », in J.-M. Pierrel (éd.), *Ingénierie des Langues*, Hermès, Paris.
- Danlos, Laurence, Gaiffe, Bertrand et Roussarie, Laurent (2001). « Document Structuring à la SDRT », in *Proceedings of the 8th European Workshop on Natural Language Generation (ACL'01)*, Toulouse, pp. 11-20.
- Grosz, Barbara J., Joshi, Aravind K. et Weinstein, Scott (1995). « Centering : A framework for modeling the local coherence of discourse », in *Computational Linguistics* 21 (2), pp. 203-225.
- Kahane, Sylvain et Mel'čuk, Igor (1999). « La synthèse automatique ou la correspondance entre graphes sémantiques et arbres syntaxiques. Le cas des phrases à extraction en français contemporain », in *TAL* 40 (2), pp. 25-85.
- Kahane, Sylvain (2001). « Grammaires de dépendance et théorie Sens-Texte », in *Actes de TALN* (version revue et corrigée), Tours.
- Kahane, Sylvain (2002). *Grammaire d' Unification Sens-Texte : Vers un modèle mathématique articulé de la langue*, document de synthèse pour l' Habilitation à Diriger des Recherches, École Doctorale Sciences du Langage, Université Denis Diderot, Paris 7.

- Lambrecht, Knud (1994). *Information Structure and Sentence Form : Topic, focus and the mental representations of discourse referents*, Cambridge Studies in Linguistics 71, Cambridge University Press.
- Lascarides, Alex et Asher, Nicholas (1991). « Discourse relations and defeasible knowledge », in *Proceedings of the 29th annual meeting of the Association for Computational Linguistics*, University California, Berkeley.
- Mel'čuk, Igor (2001). *Communicative Organization in Natural Language : The Semantic-communicative Structure of Sentences*, Benjamins, Amsterdam.
- Névél, Aurélie (2002). *Étude du connecteur « d'autant plus que » dans une optique de génération automatique. Intégration au générateur CLEF*, Mémoire de DEA en Linguistique Théorique, Descriptive et Automatique, mention Linguistique Informatique, École Doctorale Sciences du Langage, Université Denis Diderot, Paris 7.
- Raynal, Céline (2002). *Calcul effectif de la présupposition. Seul, même, aussi*, Mémoire de DEA en Linguistique Théorique, Descriptive et Automatique, mention Linguistique Informatique, École Doctorale Sciences du Langage, Université Denis Diderot, Paris 7.
- Roussarie, Laurent (1998). « Le problème de la structuration et de la représentation du discours vu sous l'angle de la génération automatique », in *TAL* 39 (2), pp. 35-55.
- Roussarie, Laurent (2000). *Un modèle théorique d'inférences de structures sémantiques et discursives dans le cadre de la génération automatique de textes*, Thèse de Doctorat en Linguistique, École Doctorale Sciences du Langage, Université Denis Diderot, Paris 7.