

## **Structure communicative du discours : étude pour la génération automatique de textes**

Laurence Delort

LaTTICE-TALaNa, Université Paris 7  
UFR de Linguistique, case 7003  
2, place Jussieu  
75251 Paris Cedex 05  
laurence.delort@linguist.jussieu.fr  
Fin de thèse : décembre 2005

### **Mots-clefs – Keywords**

Structure communicative (ou structure informationnelle), progression thématique, structure du discours (relations de discours), structuration de document, génération automatique de textes

Communicative structure (or information structure), thematic progression, discourse structure (discourse relations), document structuring, natural language generation

### **Résumé – Abstract**

Dans cet article, nous montrons que la cohérence d'un discours dépend de la relation entre la structure communicative des phrases et la structure du discours. Du point de vue de la synthèse, la visée communicative contrôle la structure du discours, et la structure du discours contraint le choix des structures communicatives phrastiques : nous proposons de reproduire ce processus dans un système de génération de textes. Nous montrons de quelle manière la structure communicative intervient lors de la phase de structuration de document pour permettre la génération de discours cohérents et répondant à des visées communicatives particulières.

In this paper, we show that discourse coherence depends on the relationship between communicative structure of sentences and discourse structure. From the synthesis point of view, communicative goal controls discourse structure, and discourse structure constrains choice of communicative structures of sentences : we propose to reproduce this process in a natural language generation system. We show how communicative structure can be used during document structuring task so as to produce discourses which are coherent and corresponding to particular communicative goals.

# 1 Introduction

Des études linguistiques, dont (Charolles, 1978), ont démontré que la cohérence d'un discours est tributaire de la structure communicative des phrases qui composent le texte (de quoi parle-t-on et qu'en dit-on ?) et des relations de discours en jeu entre les éventualités (événements, états) décrites dans le texte. Selon (Charolles, 1978), pour qu'un texte soit cohérent, "il faut qu'il comporte dans son développement linéaire des éléments à récurrence stricte" (méta-règle de répétition) et "que son développement s'accompagne d'un apport sémantique constamment renouvelé" (méta-règle de progression). En d'autres termes, pour être cohérent, un texte doit avoir une continuité thématique, i.e. avoir un seul propos, et une progression rhématique, i.e. apporter des informations nouvelles sur ce propos. Par ailleurs, pour qu'un texte soit cohérent, "il faut que les faits qu'il dénote dans le monde représenté soient reliés" (méta-règle de relation), ce qui signifie qu'il doit exister des relations qui permettent de lier les éventualités décrites et de structurer le discours : ces relations sont appelées relations rhétoriques ou discursives.

La présente étude a pour objectif de démontrer que structure communicative et structure du discours sont fortement liées : à une relation de discours correspond un enchaînement de structures communicatives, appelé "progression thématique", privilégié. Cette étude a également pour objectif de montrer que des informations communicatives, telles que des règles sur la corrélation entre une relation de discours et une progression thématique, peuvent être intégrées dans des systèmes de Traitement Automatique des Langues et notamment dans un générateur de textes. Nous présentons les notions de structure communicative et de structure de discours dans la section 2. Dans la section 3, l'analyse d'un texte va permettre de rendre compte du lien entre une relation de discours et une progression thématique. Enfin, dans la section 4, nous montrons dans quelle mesure des informations communicatives peuvent intervenir dans un système de génération de textes.

## 2 Structure communicative et structure du discours

Pour qu'un discours soit jugé cohérent, il faut que son organisation communicative intra-phrastique (la partition thème~rhème) et inter-phrastique (les enchaînements des structures communicatives) soit en adéquation avec son organisation textuelle (les relations de discours). Avant d'éclaircir cette conception de la cohérence textuelle, nous présentons ce que nous entendons par thème, rhème, progression thématique et relation de discours.

### 2.1 Structure communicative de la phrase

Toute phrase possède une structure communicative (ou 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 et donc également d'un point de vue syntaxique (Lambrecht, 1994), (Mel'čuk, 2001). Au niveau communicatif, la caractéristique de tout message est que l'on dit quelque chose (le rhème) à propos de quelque chose (le thème) :

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

*Structure communicative du discours : étude pour la génération automatique de textes*

- Le thème est l'entité au sujet de laquelle la proposition (plus précisément le rhème) est. Il correspond au propos du message, à ce dont on parle. Il est un élément facultatif : il existe des phrases sans thème, appelées rhématiques (ou thétiques).

Par exemple, dans le discours (1)<sup>1</sup> :

- (1) a. IL FAUT SOULIGNER LA VICTOIRE DE L'ÉQUIPE DES FOREURS.  
b. Ils ont remporté LES SÉRIES ÉLIMINATOIRES POUR LA DEUXIÈME FOIS DE LEUR HISTOIRE.

l'événement 'la victoire de l'équipe des Foreurs'<sup>2</sup> est introduit dans (1a) par le biais de la formule présentative *il faut souligner*. Il s'agit d'une phrase rhématique : il n'y a aucun propos particulier, tout l'énoncé constitue une information nouvelle. Dans (1b), il est question de 'la victoire de l'équipe des Foreurs' (le thème, lexicalisé par *ils ont remporté*) et il y a des informations nouvelles quant à cette victoire (le rhème) : ce sont *les séries éliminatoires* qui ont été remportées et *pour la deuxième fois de leur histoire*. La question sous-jacente à (1b) pourrait être : que peut-on dire à propos de la victoire de l'équipe des Foreurs ?

Dans l'énoncé (2) :

- (2) C'est L'ÉQUIPE DES FOREURS qui a remporté les séries éliminatoires.

il est question d'une entité qui *a remporté les séries éliminatoires* (le thème) et on en dit que c'est *l'équipe des Foreurs* (le rhème). L'énoncé (2) peut être la réponse à la question : qui a remporté les séries éliminatoires ?

Dans un discours, les structures communicatives s'enchaînent d'une certaine façon : c'est ce qui est appelé "progression thématique". Les premières études sur la progression thématique de Daneš sont reprises dans (Combettes, 1988), où sont présentées trois progressions thématiques canoniques. Dans une progression "linéaire", le thème de la phrase est issu du rhème de la phrase précédente. Par exemple, dans (3) :

- (3) a. LA PERFORMANCE DE [MAXIME DAIGNEAULT] EST À NOTER.  
b. IL DEVIENT LE PREMIER GARDIEN DE BUT RECRUE DE 16 ANS À CONDUIRE SA FORMATION À LA CONQUÊTE DE LA COUPE DU PRÉSIDENT.

le thème de (3b), *il*, est issu du focus rhématique *Maxime Daigneault*<sup>3</sup>, partie du rhème la plus informationnelle de (3a), donnée par une intonation particulière (Marandin *et al.*, 2002).

Dans une progression "à thème constant", le thème de la phrase est issu du thème de la phrase précédente. Par exemple :

- (4) a. Simon Gamache A RÉCOLTÉ AU MOINS UN POINT DANS CHACUN DES 21 MATCHS.  
b. IL A ÉGALEMENT REMPORTE LE TROPHÉE GUY LAFLEUR.

Dans une progression "à thème dérivé", le thème de la phrase est issu d'un "hyper-thème". Par exemple :

<sup>1</sup>Nous marquons le thème en minuscules et le RHÈME en majuscules.

<sup>2</sup>Les sèmes (ou sémantèmes) sont marqués entre guillemets simples (ex : 'victoire') et les réalisations de surface de ces sèmes en italique (ex : *ont remporté, sont victorieux*, etc.).

<sup>3</sup>Dans le RHÈME, le [FOCUS RHÉMATIQUE] est marqué entre crochets.

- (5) a. Il faut souligner LA PERFORMANCE DE [TOUS CES JEUNES].  
 b. Simon Gamache A RÉCOLTÉ AU MOINS UN POINT DANS CHACUN DES 21 MATCHS.  
 c. Maxime Daigneault DEVIENT LE PREMIER GARDIEN DE BUT RECRUE DE 16 ANS À CONDUIRE SA FORMATION À LA CONQUÊTE DE LA COUPE DU PRÉSIDENT.

*Simon Gamache* dans (5b) et *Maxime Daigneault* dans (5c) sont des éléments thématiques issus de *tous ces jeunes* introduit dans (5a).

Dans les textes, nous pouvons trouver différentes combinaisons de ces trois progressions thématiques. Il est également possible de trouver d'autres enchaînements thématiques. En effet, il existe une progression dans laquelle le thème n'est issu ni du thème ni du rhème de la phrase précédente (section 3.1).

## 2.2 Structure du discours

La structure d'un discours est mise au jour par des relations dites rhétoriques ou discursives. Ces relations de discours rendent compte des relations entre les constituants discursifs (éventualités (événements ou états), faits ou propositions). Nous allons définir quelques-unes de ces relations telles qu'elles sont envisagées dans le cadre de la *Segmented Discourse Representation Theory*, abrégée en SDRT (Asher, 1993), (Asher & Lascarides, 2003).

Soient deux éventualités  $e_1$  et  $e_2$  décrites respectivement dans les phrases (a) et (b). Les éventualités  $e_1$  et  $e_2$  sont liées par une relation :

- Élaboration, lorsque  $e_2$  explicite, donne des informations supplémentaires sur  $e_1$ . Par exemple :
  - (6) a. La performance de Maxime Daigneault est à noter.
  - b. Il devient le premier gardien de but recrue de 16 ans à conduire sa formation à la conquête de la Coupe du président.
- Parallèle, lorsque  $e_1$  et  $e_2$  portent sur le même "thème". Par exemple :
  - (7) a. Simon Gamache a récolté au moins un point dans chacun des 21 matchs.
  - b. Il a également remporté le trophée Guy Lafleur.
- Commentaire, lorsque  $e_2$  est un commentaire sur  $e_1$ . Par exemple :
  - (8) a. L'équipe des Foreurs a remporté les séries éliminatoires.
  - b. C'est une belle victoire.
- Résultat, lorsque  $e_2$  est causée par  $e_1$ . Par exemple :
  - (9) a. Les Foreurs ont remporté pour la deuxième fois de leur histoire les séries éliminatoires.
  - b. Je suis fier de cette équipe.
- Explication, lorsque  $e_2$  cause (ou explique)  $e_1$ . Par exemple :
  - (10) a. Je suis fier de l'équipe des Foreurs.
  - b. Ils ont remporté pour la deuxième fois de leur histoire les séries éliminatoires.

### 3 Mise au jour du lien entre structure communicative et structure du discours : analyse d'un texte

Notre analyse se fonde sur la structure communicative des phrases et sur les relations discursives qu'elles entretiennent dans le texte suivant<sup>4</sup>:

- (11) a. Albert Jambon, géologue à Paris-VI, est tout content de la pierre Al-Mala'ika.
- b. Le petit caillou qu'il vient d'étudier, avec des collègues de Brest, de Lyon et d'Angers, serait un roc de Mars.
- c. C'est ce qu'il assure dans son communiqué de déclaration à la Meteoritical Society.

Nous procédons en deux temps : nous analysons tout d'abord la structure communicative de chaque phrase, pour déduire des progressions thématiques et, ensuite, nous analysons les relations de discours établies entre les constituants (11a), (11b) et (11c). Cette analyse va permettre de mettre au jour la corrélation entre une relation de discours et une progression thématique.

#### 3.1 Analyse de la structure communicative des phrases

Nous proposons pour chaque phrase du texte (ce qui nous semble être) sa structure communicative. Pour simplifier la démonstration, nous n'analysons pas les éléments mis entre parenthèses. Pour l'énoncé (11a), la structure communicative est la suivante :

(11a) Albert Jambon, (géologue à Paris-VI,) EST TOUT CONTENT DE [LA PIERRE AL-MALA'IKA].

Le locuteur parle d'*Albert Jambon* et il en dit qu'il *est tout content de la pierre Al-Mala'ika*. La *pierre Al-Mala'ika* est le focus rhématique et il est repris en thème dans la phrase suivante<sup>5</sup> :

(11b) Le petit caillou qu'il vient d'étudier, (avec des collègues de Brest, de Lyon et d'Angers,) SERAIT UN ROC DE MARS.

*Le petit caillou qu'il vient d'étudier* est issu du focus rhématique de (11a) : la progression thématique est linéaire. Et le locuteur dit de ce *petit caillou* qu'il *serait un roc de Mars*.

Le lien entre les structures communicatives de (11b) et de (11c) n'est pas évident à première vue car le thème de (11c) n'est issu ni du thème ni du rhème de (11b) :

(11c) C'EST CE QU'IL ASSURE DANS SON COMMUNIQUÉ DE DÉCLARATION À LA METEORITICAL SOCIETY.

Le locuteur dit de *quelque chose* que *c'est ce qu'il assure dans son communiqué de déclaration à la Meteoritical Society*. Ce *quelque chose* est le fait qui est décrit dans (11b) soit 'le petit caillou qu'il vient d'étudier serait un roc de Mars'. Ce fait est repris sous la forme du pronom démonstratif *cela* (*c'*) dans (11c). En d'autres termes, le fait que 'le petit caillou qu'il vient d'étudier serait un roc de Mars' (le thème) est 'ce qu'il assure dans son communiqué de déclaration à la Meteoritical Society' (le rhème).

<sup>4</sup>Extrait, légèrement modifié, d'un article paru dans *Libération* (édition du 23 janvier 2003 (p.22) ou <http://www.liberation.fr/page.php?Article=83169>).

<sup>5</sup>Dans le texte original, la phrase (11a) est la suivante : *Albert Jambon, géologue à Paris-VI, est tout content de sa "dernière martienne"*, où *sa "dernière martienne"* est une cataphore événementielle (co-référence avec le fait décrit dans (11b)). Nous préférons ne pas tenir compte de cette version de (11a) car elle demande des explications qui ne font pas l'objet de la présente étude.

### 3.2 Analyse de la structure du discours

Entre les constituants discursifs (11a) et (11b), il y a une relation *Élaboration*. Le fait décrit dans (11b) explicite une partie de l'éventualité décrite dans (11a). En l'occurrence, le locuteur introduit une entité *la pierre Al-Mala'ika* dans (11a) et l'élabore dans la phrase suivante (11b) :

- (11) a. Albert Jambon, géologue à Paris-VI, est tout content de la pierre Al-Mala'ika.  
    ↓ *Élaboration*
- b. Le petit caillou qu'il vient d'étudier, avec des collègues de Brest, de Lyon et d'Angers, serait un roc de Mars.

Il y a une relation *Commentaire* entre les constituants (11b) et (11c). En effet, l'éventualité décrite dans (11c) est un commentaire, une glose sur le fait décrit dans la phrase précédente (11b) :

- (11) b. Le petit caillou qu'il vient d'étudier, avec des collègues de Brest, de Lyon et d'Angers, serait un roc de Mars.  
    ↓ *Commentaire*
- c. C'est ce qu'il assure dans son communiqué de déclaration à la Meteoritical Society.

### 3.3 Observations

L'analyse des progressions thématiques et des relations de discours dans le texte (11) nous amène au constat suivant : la cohérence d'un discours dépend non seulement de la progression thématique mais également de son lien avec une relation de discours.

Voici le texte (11) où les structures communicatives et les relations de discours sont indiquées :

- (11) a. Albert Jambon, (géologue à Paris-VI,) EST TOUT CONTENT DE [LA PIERRE AL-MALA'IKA].  
    ↓ *Élaboration*
- b. Le petit caillou qu'il vient d'étudier, (avec des collègues de Brest, de Lyon et d'Angers,) SERAIT UN ROC DE MARS.  
    ↓ *Commentaire*
- c. C'EST CE QU'IL ASSURE DANS SON COMMUNIQUÉ DE DÉCLARATION À LA METEORITICAL SOCIETY.

Nous observons qu'à une relation de discours peut être associée une progression thématique privilégiée. S'il y a une relation *Élaboration* entre deux constituants discursifs, la progression thématique est de préférence linéaire (le focus rhématique est repris en thème). Et si nous avons une relation *Commentaire*, le thème reprend l'éventualité ou le fait décrit dans la phrase précédente.

Nous pouvons ainsi tenter d'associer une progression thématique à chaque relation de discours décrite en SDRT. Des règles sur cette corrélation ont pu être dégagées à partir de l'observation de corpus (Delort, 2002).

## 4 Application à la génération automatique de textes

Notre travail se situe dans le cadre de la génération automatique de textes (abrégée en GAT), système de Traitement Automatique des Langues qui requiert des informations linguistiques pointues et également, nous allons le voir, communicatives<sup>6</sup>.

Un générateur de textes a pour but de produire un discours cohérent et répondant à une visée communicative à partir de données conceptuelles. Pour atteindre ce double objectif, il est nécessaire de prendre en considération, d'une part, l'interaction entre la structure du discours (i.e. les relations de discours) et l'organisation communicative des phrases (i.e. la progression thématique), et d'autre part, les intentions du locuteur.

Intégrer des informations communicatives dans un système de GAT permettrait d'assurer la cohérence du discours (par la prise en compte de la corrélation entre une progression thématique et une relation de discours) et de limiter la génération de paraphrases à celles qui répondent à une visée communicative particulière.

### 4.1 Structuration du discours

L'entrée d'un générateur de textes peut être une représentation logique du contenu sémantique du texte à produire. C'est à partir de cette représentation logique que l'on va organiser le texte : c'est la phase dite de "structuration de document" (Danlos *et al.*, 2001). Le module de structuration de document doit proposer un plan de document au module de lexicalisation, où s'opère le choix des constructions syntaxiques. Afin d'illustrer le processus de structuration de document, prenons la forme logique suivante :

$$(12) \exists x, x_1, y, y_2, z, w, s_1, e_2, f_2, e_3 ($$

Albert Jambon ( $x$ ) $\wedge$	géologue_Paris-VI ( $x_1$ ) $\wedge$	$x = x_1 \wedge$
Al-Mala'ika ( $y$ ) $\wedge$	pierre ( $y_2$ ) $\wedge$	$y = y_2 \wedge$
roc_de_Mars ( $z$ ) $\wedge$	Meteoritical Society ( $w$ ) $\wedge$	
être_content ( $s_1, x, y$ ) $\wedge$	$s_1$ O now $\wedge$	
étudier ( $e_2, x, y_2$ ) $\wedge$	$e_2$ O now $\wedge$	
être ( $f_2, y_2, z$ ) $\wedge$	$f_2$ O now $\wedge$	
déclarer ( $e_3, x, w, f_2$ ) $\wedge$	$e_3$ O now )	

Cette forme logique, qui représente la sémantique du discours à générer (proche du discours (11)), comprend six entités ('Albert Jambon', 'géologue à Paris-VI', 'Al-Mala'ika', 'pierre', 'roc de Mars' et 'Meteoritical Society'), un état ( $s_1$ ), un fait ( $f_2$ ) et deux événements ( $e_2$  et  $e_3$ ). Il y a également deux équations ( $x = x_1$  et  $y = y_2$ ) qui indiquent une co-référence entre deux entités. L'objectif pour le module de structuration de document est de permettre la génération de tous les para-textes possibles à partir de la forme logique (12). À la fin du processus de génération, nous pourrions obtenir, par exemple, les para-textes (13) et (14) suivants :

- (13) a. Albert Jambon, géologue à Paris-VI, a fait une déclaration à la Meteoritical Society.  
b. La pierre qu'il vient d'étudier serait un roc de Mars.  
c. Il est content de la pierre Al-Mala'ika.

---

<sup>6</sup>Voir notamment (Iordanskaja, 1992) et (Wanner *et al.*, 2003) pour des études similaires dans le cadre de la Théorie Sens-Texte.

Dans le texte (13), 'la déclaration à la Meteoritical Society' semble être le propos central du discours, appelé "topique"<sup>7</sup>. La suite du discours serait l'élaboration de ce topique : qu'est-ce qu'a déclaré Albert Jambon à la Meteoritical Society ?

Quant au texte (14), il paraît traiter d'une seule entité 'Albert Jambon' :

- (14) a. Albert Jambon, géologue à Paris-VI, est content de la pierre Al-Mala'ika.
- b. Il vient d'étudier une pierre qui serait un roc de Mars.
- c. Il déclare cela à la Meteoritical Society.

La position thématique d'"Albert Jambon' dans chaque phrase est un indice (non suffisant) pour dire qu'il s'agit du topique de (14).

Le contenu propositionnel des para-textes (13) et (14) est identique. Ce qui les distingue est l'organisation des phrases qui les composent et leurs visées communicatives. Selon nous, la visée communicative guide la structure du discours : en fonction du propos du locuteur et de ce qu'il veut en dire, nous aurons une organisation textuelle (et des relations de discours) spécifique. Par exemple, dans (13), le locuteur veut parler de 'la déclaration à la Meteoritical Society' et l'expliciter : nous aurons donc une relation Élaboration entre l'éventualité de (13a) (qui introduit le topique) et les constituants (13b) et (13c).

Donc, en fonction d'une visée communicative, nous allons avoir un plan de document particulier. En l'occurrence, nous voulons que le texte traite de 'la pierre Al-Mala'ika' et nous voulons donner des informations sur (i.e. élaborer) cette entité dans le discours. Suivant le processus de construction d'une SDRS (pour *segmented discourse representation structure*) décrit dans (Danlos *et al.*, 2001), nous obtenons le plan de document illustré dans la figure 1<sup>8</sup>.

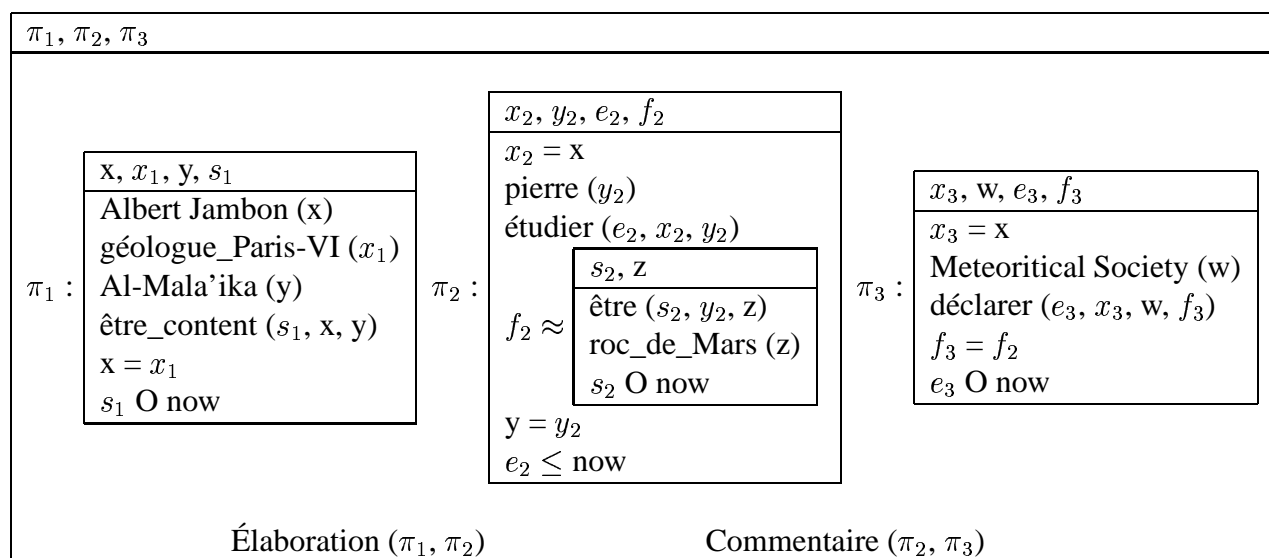


Figure 1: Plan de document possible construit à partir de la forme logique (12)

La DRS (pour *discourse representation structure*) étiquetée  $\pi_1$  permet l'introduction du référent 'la pierre Al-Mala'ika'. La DRS  $\pi_2$  élabore 'la pierre Al-Mala'ika', nous avons donc une relation Élaboration entre  $\pi_1$  et  $\pi_2$  (Élaboration ( $\pi_1, \pi_2$ )). Et la DRS  $\pi_3$  commente ce qui a été dit sur 'Al-Mala'ika' : une relation Commentaire est établie entre  $\pi_2$  et  $\pi_3$  (Commentaire ( $\pi_2, \pi_3$ )). Ce

<sup>7</sup>Pour une discussion sur la notion de topique, voir (Marandin, 1988).

<sup>8</sup>Faute de place, nous ne pouvons pas rentrer dans les détails.



plan de document, qui rend compte de l'organisation des énoncés et des relations de discours en jeu dans le texte, est obtenu en fonction d'une visée communicative particulière.

## 4.2 Structuration communicative des phrases

Une fois la structure du discours établie et les relations de discours mises au jour, il faut attribuer à chaque phrase la structure communicative appropriée. Pour qu'un discours soit cohérent, il doit observer certaines progressions thématiques liées à des relations de discours : si une relation Élaboration est établie entre deux éventualités  $e_1$  et  $e_2$ , alors le focus rhématique de  $P_1$  (phrase exprimant  $e_1$ ) peut être repris en thème dans  $P_2$  (phrase exprimant  $e_2$ ) ; et si une relation Commentaire est établie entre deux éventualités  $e_1$  et  $e_2$ , alors  $e_1$  peut être repris en thème dans  $P_2$ .

La prise en compte des règles sur la corrélation entre une relation de discours et une progression thématique va permettre la génération du discours suivant :

- (15) a. Albert Jambon, géologue à Paris-VI, est content de la pierre Al-Mala'ika.
- b. La pierre qu'il vient d'étudier serait un roc de Mars.
- c. C'est ce qu'il déclare à la Meteoritical Society.

Ce discours respecte la visée communicative telle que l'on parle de 'la pierre Al-Mala'ika' et que l'on l'élabore. Par ailleurs, ce discours est cohérent car les règles sur la corrélation entre une relation de discours et une progression thématique sont respectées.

Sans but communicatif particulier, le générateur aurait produit tous les para-textes possibles à partir de la forme logique (12) (comme (13) et (14)) et, sans informations sur les enchaînements thématiques, il aurait produit des textes maladroits comme le suivant :

- (16) a. Albert Jambon, géologue à Paris-VI, est content de la pierre Al-Mala'ika.
- b. Il vient d'étudier une pierre qui serait un roc de Mars.
- c. C'est ce qu'il déclare à la Meteoritical Society.

D'une part, l'interprétation du discours formé par les phrases (16a) et (16b) est difficile car *la pierre Al-Mala'ika*, introduit dans la première phrase, demande une explicitation (qu'est-ce que *la pierre Al-Mala'ika* et pourquoi Albert Jambon en est-il content ?) et il est question dans la phrase suivante, non pas d'une élaboration de *la pierre Al-Mala'ika* mais d'un nouvel événement qui est 'l'étude d'une pierre'. D'autre part, dans (16c), le pronom démonstratif *c'* peut être interprété comme étant la reprise de l'événement 'l'étude d'une pierre' (décrit dans (16b)) alors que (16c) est un commentaire sur le fait 'la pierre est un roc de Mars' (également décrit dans (16b)).

C'est le respect des progressions thématiques et de leur corrélation avec une relation de discours qui va permettre de générer des discours les plus cohérents possible et d'éviter la génération de discours maladroits.

## 5 Conclusion

La cohérence d'un discours dépend de la façon dont les informations sont organisées et des relations entre les événements dénotés. Nous avons montré que structure communicative phrastique

et structure du discours entretiennent un rapport privilégié puisque à une relation de discours, nous pouvons associer une progression thématique. Cette interaction a un intérêt linguistique, car elle est liée à la notion de cohérence textuelle, et un intérêt applicatif. En GAT, la visée communicative guide la construction du plan de document. Le plan de document, à son tour, contraint les structures communicatives des phrases i.e. les futures constructions syntaxiques. L'intégration d'informations sur la structure communicative dans un système de GAT va permettre de :

- réduire le nombre de paraphrases obtenues à la fin du processus de génération ;
- produire un discours répondant à une visée communicative particulière ;
- produire un discours cohérent par la prise en compte des progressions thématiques et des relations de discours.

L'étude approfondie de la structure communicative du discours peut fournir de nombreuses informations linguistiques qui permettraient d'améliorer certains systèmes de Traitement Automatique des Langues comme, nous l'avons vu dans la présente étude, des systèmes de génération de textes.

## Références

- ASHER N. (1993). *Reference to Abstract Objects in Discourse*, Kluwer Academic Publishers, Dordrecht.
- ASHER N. & LASCARIDES A. (2003). *Logics of Conversation*, Cambridge University Press, à paraître.
- CHAROLLES M. (1978). Introduction aux problèmes de la cohérence des textes, in *Langue Française* 38, pp. 7-41.
- COMBETTES B. (1988). *Pour une grammaire textuelle : la progression thématique*, De Boeck-Duculot, Bruxelles.
- DANLOS L., GAIFFE B. & ROUSSARIE L. (2001). Document Structuring à la SDRT, in *Proceedings of the 8th European Workshop on Natural Language Generation (ACL'01)*, Toulouse, pp. 11-20.
- DELORT L. (2002). *Intégration d'informations sur la structure communicative dans un système de génération de textes*, Mémoire de DEA en Linguistique Informatique, École Doctorale Sciences du Langage, Université Denis Diderot, Paris 7.
- IORDANSKAJA L. (1992). Communicative Structure and its Use during Text Generation, in *International Forum on Information and Documentation* 17 (2), pp. 15-27.
- LAMBRECHT K. (1994). *Information Structure and Sentence Form : Topic, focus and the mental representations of discourse referents*, Cambridge Studies in Linguistics 71, Cambridge University Press.
- MARANDIN J.-M. (1988). À propos de la notion de thème en discours. Éléments d'analyse dans le récit, in *Langue Française* 78, pp. 67-87.
- MARANDIN J.-M., BEYSSADE C., DELAIS-ROUSSARIE E. & RIALLAND A. (2002). Discourse marking in French : C accents and discourse moves, in *Proceedings of the Speech Prosody Conference*, Aix-en-Provence, pp. 471-474.
- MEL'ČUK M. (2001). *Communicative Organization in Natural Language : The Semantic-communicative Structure of Sentences*, Benjamins, Amsterdam.
- WANNER L., BOHNET B. & GIERETH M. (2003). Deriving the Communicative Structure in Applied NLG, in *Proceedings of the 9th European Workshop on Natural Language Generation (EACL'03)*, Budapest, Hongrie, pp. 111-118.