Université de Montréal

Structuration et mise en jeu procédurale d'un modèle linguistique déclaratif dans un cadre de génération de texte

par

Alain Polguère

Département de linguistique et philologie

Faculté des arts et des sciences

Thèse présentée à la Faculté des études supérieures
en vue de l'obtention du grade de
Philosophiæ Doctor (Ph.D.)
en linguistique

janvier 1990

© Alain Polguère, 1990

Université de Montréal
Faculté des études supérieures

Cette thèse intitulée:

"Structuration et mise en jeu procédurale d'un modèle linguistique déclaratif dans un cadre de génération de texte"

présentée par:

Alain Polguère

a été évaluée par un jury composé des personnes suivantes:

Thèse acceptée le:.....

SOMMAIRE

Cette recherche relève de la linguistique informatique. Elle vise à déterminer les procédures d'activation des règles déclaratives des modèles linguistiques de type Sens-Texte dans un cadre de génération de texte. La génération de texte est comprise ici dans un sens très limité puisque l'on s'intéresse uniquement aux opérations linguistiques impliquées dans la synthèse des énoncés phrastiques.

Plus précisément, nous allons étudier les liens entre la structuration sémantique et communicative d'une phrase et sa structuration syntaxique. On considère une acception bien spéciale du terme "Sens" qui signifie l'invariant de paraphrase langagière et l'on considère un niveau de représentation syntaxique "profonde", qui ne met en jeu que les lexèmes de la phrase ayant véritablement un Sens.

Nous partons d'expérimentations sur deux systèmes de génération de texte ainsi que d'une étude en profondeur de la composante sémantique des modèles Sens-Texte. Cette composante est celle prise en compte dans la transition entre la représentation du Sens et la représentation syntaxique profonde d'un énoncé: c'est-à-dire, la transition sémantique.

Les deux apports notionnels majeurs de cette thèse sont les suivants:

1) Au niveau de la Théorie Sens-Texte elle-même, nous introduisons la notion de DEPENDANCE COMMUNICATIVE, qui

cohabite avec celle de dépendance sémantique. Elle permet la construction de véritables règles déclaratives de correspondance entre les représentations sémantiques et les représentations syntaxiques profondes — sur le modèle des autres règles de correspondance Sens-Texte.

2) Au niveau des procédures d'activation, nous introduisons la notion de PARCOURS SEMANTIQUE, qui permet de mettre en œuvre les opérations constituant la transition sémantique.

Notre recherche n'aborde pas la transition sémantique de façon exhaustive. On verra notamment que les problèmes de paraphrase lexicale et de grammaticalisation (réalisation de certaines significations par des moyens grammaticaux) ne reçoivent pas de traitement approfondi. L'emphase est mise sur le problème de l'interprétation de la structure communicative d'une représentation sémantique (grosso modo, la structure thème/rhème) afin de réaliser l'arborisation, qui est la construction d'une hiérarchie syntaxique.

La structure de la thèse est la suivante:

- L'Introduction situe plus précisément la recherche dans son contexte expérimental et théorique. Elle présente aussi un plan détaillé de la thèse en résumant le contenu de chaque chapitre.
- La Partie **1** fait une présentation approfondie de la composante sémantique des modèles Sens-Texte. Elle examine le système complet de règles sémantiques déclaratives en introduisant les notions clés qui serviront à la définition du système d'activation de ces règles.
- La Partie **2** traite de la transition sémantique proprement dite. Elle propose des règles d'activation de la composante sémantique tout en approfondissant la notion de structure communicative, qui joue un rôle central dans le contrôle du processus étudié.

TABLE DES MATIERES

Sommaire iii

Table des matières v

Liste des figures.....xi

Liste des abréviations et symboles.....xvi

Conventions typographiques xvii

Remerciements xviii

Dédicace xix

INTRODUCTION.....1

1 Présentation de la recherche.....1

1.1 Motivations et buts 1

1.2 Génération de texte et synthèse linguistique.....6

1.2.1 Les modèles linguistiques en TALN 6

1.2.2 Le "véritable" titre de cette thèse 10

1.2.3 Les systèmes GOSSIP et FoG.....12

2 Fondements théoriques.....17

2.1 Comparaison de deux approches: linguistique Sens-Texte *vs.* linguistique systémique 18

2.2 Mises au point terminologiques 24
3 Structure de la thèse 26
3.1 Partie 1: "La Composante Sémantique d'un MST" 27
3.2 Partie 2: "La Transition Sémantique" 30

PARTIE 1
La Composante Sémantique d'un Modèle Sens-Texte

CHAPITRE I **Caractérisation formelle des RSém et des RSyntP** 36
1 Rappel des principes de base de la sémantique selon la TST 37
1.1 Sens et paraphrase 37
1.2 Théorie sans texte 40
2 Quelques particularités formelles des réseaux sémantiques 43
2.1 Généralités 43
2.2 Représentation linéarisée des réseaux sémantiques 46
2.3 Acyclicité 50
3 Autres approches utilisant les réseaux sémantiques 52
4 Les arbres syntaxiques profonds 57
4.1 Remarques sur le recours à une syntaxe de dépendance 57
4.2 Formalisme des arbres syntaxiques profonds 60
5 Justification théorique des RSém et des RSyntP 63

CHAPITRE II **Règles établissant les correspondances RSém ⇔ RSyntP** 70
1 Les règles de bonne formation 71
2 Les règles de paraphrase sémantique 75
2.1 Formalisation des règles de paraphrase sémantique 75

2.2 Application facultative des règles de paraphrase sémantique 89
2.3 Le definiens et la notion de représentation sémantique réduite 90
3 Les règles grammatico-sémantiques 93
3.1 Examen d'un type de signification pris en compte 94
3.2 Le cas des articles en français 97
3.2.1 Signification de l'article 98
3.2.2 Traitement grammatical des articles du français 104
3.3 Formalisation des grammèmes et intervention dans la transition sémantique 106
4 Les règles de correspondance sémantique proprement dites 113
4.1 Formalisation des règles de correspondance sémantique 114
4.1.1 Les règles de correspondance dans SSE 115
4.1.2 Les règles de correspondance sémantique et la notion de dépendance communicative 118
4.1.3 Quelques exemples 123
4.2 Le cas des relations **APPEND** et **COORD** 127
4.3 Introduction de lexèmes par les règles de correspondance sémantique 130
4.3.1 Quantification des actions 131
4.3.2 Sommet de l'arbre syntaxique profond 132

ANNEXE DU CHAPITRE II **Ébauches de règles de correspondance sémantique mettant en jeu la relation ATTR** 134
1 Corpus tiré de Apresjan *et al.* 1984 135
2 Règles RSém ⇔ RSyntP mettant en jeu la relation syntaxique profonde **ATTR** 139

PARTIE 2
La Transition Sémantique

CHAPITRE III **Principes généraux de transition sémantique**..... 148
1 Exposé du problème à traiter..... 149
2 Opérations de transition impliquées par la composante sémantique d'un MST..... 153
2.1 Répertoire des opérations mises en jeu 153
2.1.1 La lexicalisation (au sens propre)..... 154
2.1.2 La grammaticalisation 155
2.1.3 L'arborisation..... 156
2.1.4 La hiérarchisation..... 157
2.2 Dans quel ordre effectuer les opérations de transition sémantique?..... 158
3 Méthode proposée et notion de RSém Réduite 162

CHAPITRE IV **Hiérarchisation: calcul du nœud dominant des réseaux sémantiques**..... 172
1 La notion de nœud dominant..... 172
2 Nœud dominant et structure communicative..... 179
3 Règles de calcul du nœud dominant..... 190
3.1 Quand calculer le ND?..... 190
3.2 Comment calculer le ND? 191
3.2.1 Contrainte C_1 sur le choix du ND..... 191
3.2.2 Contrainte C_2 sur le choix du ND..... 192
3.2.3 Règle R_1 de choix du ND 194
3.2.4 Règle R_2 de choix du ND 194
3.2.5 Règle R_3 de choix du ND 195
3.2.6 Règle R_4 de choix du ND 195
3.2.7 Règle R_5 de choix du ND 196
4 Conséquences pour la bonne formation des RSém 198
4.1 Structure communicative et bonne formation des RSém..... 198

4.2 Examen des différentes configurations communicatives possibles 201
4.2.1 Absence de thème 201
4.2.2 Absence de rhème 203
4.2.3 Inclusion du rhème dans le thème 205
4.2.4 Inclusion du thème dans le rhème 208
4.2.5 Intersection du thème et du rhème 209
4.2.6 Autres problèmes topographiques 210

CHAPITRE V **Hiérarchisation: parcours des réseaux sémantiques** 214
1 Caractérisation de la notion de parcours appliquée aux RSém 214
2 Parcours descendant 218
3 Parcours ascendant 224
4 Contrôle du parcours sémantique 228
4.1 Dépendance communicative et dépendance syntaxique 229
4.2 Conséquences d'une mauvaise hiérarchisation 233
4.2.1 Présentation d'un exemple 234
4.2.2 La notion de présupposition 236
4.2.3 Ébauche de définition topographique de la présupposition de lexème 239
4.2.4 Deux cas d'échec dans la hiérarchisation 249
4.3 Imperméabilité des sous-réseaux **thème** et **rhème** 258
5 Parcours des graphes conceptuels proposé dans Sowa 1984 260

CHAPITRE VI **Arborisation: activation des règles de correspondance sémantique proprement dites** .. 268
1 Utilisation de l'algorithme de hiérarchisation dans le contrôle de l'arborisation 269
1.1 Justification de la méthode proposée 269

1.2 Fonctionnement des procédures.....272
1.3 Structure communicative de la RSyntP.....277
2 Traitement d'un exemple: *"Ces cinq ou six promenades, faites au bras de celui qu'elle idolâtrait en secret,..."* 279

CONCLUSION.....289

INDEX 298

BIBLIOGRAPHIE.....302

LISTE DES FIGURES

Chaque numéro de figure est préfixé par le numéro du chapitre dans lequel la figure apparaît. Seules les figures de l'Introduction possèdent une numérotation simple.

Figure 1 Génération/compréhension *vs.* synthèse/analyse 7
Figure 2 Exemple de *système* 21
Figure 3 Exemple de *réseau de systèmes*..... 21

Figure I-1 RSém avec cycle..... 51
Figure I-2 RSém de *"Ping regarde Pong regarder Ping."* 51
Figure I-3 Exemple de graphe conceptuel selon Sowa 1984..... 55
Figure I-4 Dichotomie linguistique/non-linguistique dans les systèmes typiques de génération 66
Figure I-5 Dichotomie linguistique/non-linguistique permise par le recours à un MST 67
Figure I-6 Décomposition du processus de traduction selon Sowa (à paraître)..... 69

Figure II-1 Définition de ACHETER**1** sous forme de réseau 79
Figure II-2 Définition de ÉCOLE**I.1b** sous forme de réseau 80
Figure II-3 RSém de (3') 82
Figure II-4 RSém de (3'')..... 82
Figure II-5 Les articles français..... 98
Figure II-6 RSém de (10) 110

Figure II-7 RSyntP de (10) 110
Figure II-8 RSyntS de (10) 111
Figure II-9 Une règle RSyntS ⇔ RMorphP pour la relation parenthétique 115
Figure II-10 Reformulation de la règle de la figure II-9 117
Figure II-11 Correspondance du type 1 pour le premier actant sémantique 124
Figure II-12 Correspondance du type 2 pour le premier actant sémantique 125
Figure II-13 Correspondance du type 3 pour le premier actant sémantique 126
Figure II-14 Exemple de règle de correspondance sémantique impliquant la relation **APPEND** 129
Figure II-15 Exemple de règle de correspondance sémantique impliquant la relation **COORD** 129
Figure II-16 Exemple de règle impliquant la relation **COORD** à partir de dépendances sémantique et communicative inverses 130
Figure II-17 Correspondance sémantique impliquant FOIS**I.1** 132
Figure II-18 Correspondance sémantique impliquant ETRE^{1}**II** 133

Figure III-1 Exemple de règle syntaxique RSyntP ⇔ RSyntS 150
Figure III-2 Tableau récapitulatif des quatre opérations de transition sémantique 159
Figure III-3 Enchaînement des opérations de transition RSém ⇒ RSyntP 161
Figure III-4 Exemple de graphe orienté 163
Figure III-5 Rôle de la RSémR dans la transition RSém ⇒ RSyntP 166
Figure III-6 Deux approches de la dichotomie profond/de surface pour le niveau sémantique 168

Figure IV-1 Métaphore physique de la hiérarchisation d'un réseau 175

Figure IV-2 Différentes structures communicatives possibles dans une même RSém 179
Figure IV-3 Structuration communicative complexe d'un énoncé (proposée par J.Firbas) 186
Figure IV-4 Exemple de chemin communicatif unissant deux nœuds n_1 et n_2 193
Figure IV-5 RSém avec chemin communicatif..... 197
Figure IV-6 RSém avec structure communicative non valide..... 198
Figure IV-7 RSém avec rhème zéro 204
Figure IV-8 RSém présentant une inclusion du rhème dans le thème..... 205
Figure IV-9 RSém sans chemin communicatif..... 206
Figure IV-10 RSém avec inclusion du thème dans le rhème 208
Figure IV-11 RSém avec intersection du thème et du rhème 209
Figure IV-12 Représentation de la conjonction de phrases au niveau syntaxique profond 211
Figure IV-13 RSém avec sous-réseaux conjoints 212
Figure IV-14 RSémR de (15) 213

Figure V-1 Structure sémantique à parcourir..... 216
Figure V-2 Résultat du parcours descendant réalisé à partir du ND 219
Figure V-3 Hiérarchie communicative résultant d'un parcours descendant..... 220
Figure V-4 RSém de *"Agricol épouse la Marquise bien que Dagobert le désapprouve."* 222
Figure V-5 Hiérarchie communicative établie sur la RSém de *"Agricol épouse la Marquise bien que Dagobert le désapprouve."* 223
Figure V-6 Résultat du parcours ascendant effectué à partir du nœud 'promenade**1**' 226
Figure V-7 RSém entièrement hiérarchisée 227
Figure V-8 Hiérarchie communicative de la figure V-7 229
Figure V-9 Hiérarchie communicative modifiée pour créer une structure arborescente..... 230

Figure V-10 Hiérarchie communicative associée à la définition de ACHETER**1** 232
Figure V-11 RSém de *"The police persecute pilots who possess a Polish plane."* 234
Figure V-12 Définition du lexème anglais PERSECUTE**1** 235
Figure V-13 Représentation graphique de la composante assertive des définitions de PRIER et ORDONNER 243
Figure V-14 Représentation graphique de la composante présuppositionnelle des définitions de PRIER et ORDONNER 244
Figure V-15 Hiérarchisation communicative invalide de la RSém V-11 établie à partir du ND '(have) power**I.5**' 250
Figure V-16 Hiérarchisation communicative invalide de la RSém V-11 établie à partir du ND 'possess**1**' 253
Figure V-17 Interprétations dénotationnelles ensemblistes de (3) et (16) 255
Figure V-18 Graphe conceptuel avec chemin énonciatif (d'après Sowa 1984) 263

Figure VI-1 Construction progressive d'une RSyntP 274
Figure VI-2 Arborisation produisant une structure communicative syntaxique profonde non valide..... 277
Figure VI-3 Arborisation produisant une structure communicative syntaxique profonde valide 278
Figure VI-4 (rappel figure V-7) RSémR de (5) 280
Figure VI-5 RSyntP de *"Ces promenades étaient des jours de bonheur."* 282
Figure VI-6 RSyntP de *"cinq ou six promenades"* 283
Figure VI-7 RSyntP de *"des promenades faites en tenant son bras"* 284
Figure VI-8 Exemple de règle d'ajustement de la RSyntP..... 285
Figure VI-9 RSyntP de *"celui qu'elle idolâtrait en secret"* 286

Figure VI-10 RSyntP de *"les seuls jours qu'elle eût jamais connus"*..... 287
Figure VI-11 RSyntP complète de (5) 287

LISTE DES ABRÉVIATIONS ET SYMBOLES

BNF	(**B**ackus-**N**aur **F**orm)
FoG	(**Fo**recast **G**enerator)
GOSSIP	(**G**enerating **O**perating **S**ystem **S**ummaries in **P**rolog)
IA	(**I**ntelligence **A**rtificielle)
MST	(**M**odèle **S**ens-**T**exte)
ND	(**N**œud **D**ominant)
RMorphP	(**R**eprésentation **Morph**ologique **P**rofonde)
RMorphS	(**R**eprésentation **Morph**ologique de **S**urface)
RPhonP	(**R**eprésentation **Phon**ologique **P**rofonde)
RPhonS	(**R**eprésentation **Phon**ologique de **S**urface)
RSém	(**R**eprésentation **Sém**antique)
RSémR	(**R**eprésentation **Sém**antique **R**éduite)
RSyntP	(**R**eprésentation **Synt**axique **P**rofonde)
RSyntS	(**R**eprésentation **Synt**axique de **S**urface)
SSE	(***S**urface **S**yntax of **E**nglish* = Mel'čuk & Pertsov 1987)
TALN	(**T**raitement **A**utomatique de la **L**angue **N**aturelle)
TST	(**T**héorie **S**ens-**T**exte)
$\Leftrightarrow$	(correspondance linguistique)
$\Rightarrow$	(transition linguistique synthétique)
$\leftarrow$	(association *énoncé-réponse* $\leftarrow$ 'question hypothétique')
$\oplus$	(méta-opération d'union linguistique)

CONVENTIONS TYPOGRAPHIQUES

Le soulignement sert à mettre en évidence une partie de citation ou d'exemple.

LES PETITES MAJUSCULES SERVENT A METTRE EN EVIDENCE UNE PARTIE DU TEXTE.

Les termes indexés apparaissent en italiques.

> Les textes en retrait et en simple interligne sont des citations ou exemples référencés.

"Les textes en italiques entre guillemets sont des citations ou exemples courts."

Les "boites" mettent en évidence certains points importants ou nouveaux énoncés dans le texte.

Les noms de lexèmes sont écrits dans le texte en grandes majuscules et suivis d'une numérotation en gras renvoyant aux entrées du dictionnaire *Petit Robert*: cf. ÉCOLE**1**. Les Sens de lexèmes possèdent la même numérotation mais sont écrits en minuscules entre guillemets simples: cf. 'école**1**'. Les noms de phrasèmes sont écrits en grandes majuscules entre ⌜...⌝: cf. ⌜PASSER À TABAC⌝.

Cette thèse a été dirigée par Richard Kittredge, co-dirigée par Igor Mel'čuk et est basée sur une recherche effectuée en très étroite collaboration avec Lida Iordanskaja.

À tous trois, je voudrais maintenant exprimer mon plus profond respect et ma reconnaissance. Je leur dois d'avoir effectivement entrepris et achevé ce travail. Je leur dois encore et surtout le plaisir de pouvoir travailler sur ce que j'aime.

Je tiens à souligner qu'en plus d'avoir effectué une lecture complète de ce travail, Lida Iordanskaja a consacré beaucoup de temps à me présenter le contenu des textes russes cités dans la bibliographie.

Finalement, je remercie ma collègue et amie Corinne Fournier pour sa lecture attentive du manuscrit et les nombreuses améliorations qu'elle m'a permis d'y apporter.

Ma venue au Québec ainsi que ma scolarité de doctorat ont été financées par une bourse de l'Entraide Universitaire Mondiale du Canada.

Ce travail est dédié à Monsieur Pierre Ledrux.

INTRODUCTION

1 PRESENTATION DE LA RECHERCHE

1.1 Motivations et buts

La *linguistique informatique* est une branche de la linguistique formelle traitant de la manipulation des modèles de la langue. Cette discipline devrait plus exactement être appelée *linguistique procédurale* puisqu'elle étudie les procédures d'activation des modèles linguistiques. Aujourd'hui, ces procédures sont la plupart du temps étudiées et validées au moyen d'implémentations[1] informatiques; il n'est donc pas vraiment nécessaire d'effectuer ici une distinction entre linguistique informatique et procédurale. En revanche, ce domaine de recherche doit être très nettement distingué de ce qu'il conviendrait d'appeler l'*informatique linguistique*. Ce dernier terme réfère à l'écriture de programmes informatiques qui modélisent, dans leur exécution, certains comportements linguistiques. L'informatique linguistique peut aussi désigner toutes les recherches en informatique qui visent à mieux stocker, rendre accessible ou, plus généralement, manipuler de l'information sur la langue.

---

[1] Dans cette thèse, nous adopterons certains anglicismes, fréquents dans le domaine de la linguistique informatique, lorsque les termes officiels proposés sont d'un emploi trop marginal. On trouvera donc utilisé (avec modération) dans ce travail des termes comme: *implémentation*, *computationnel* ou *matcher*.

Pour référer à la linguistique informatique, à l'informatique linguistique ou aux applications de ces deux disciplines, nous utiliserons le terme générique de *traitement automatique de la langue naturelle* (*TALN*). Le présent travail est donc le fruit d'une recherche de type *linguistique informatique* basée sur des expérimentations en TALN.

Le développement du TALN exerce une pression très forte sur la linguistique en général (et sur la linguistique formelle en particulier) pour que cette science "produise quelque chose" qui alimente le travail fait dans ce champ d'activité. Nous admettrons le postulat suivant:

Postulat 1:

> Une des applications informatiques les plus intéressantes pour le linguiste est la *génération de texte* car elle lui permet notamment de traiter le problème du choix paraphrastique, qui se résout avant tout au moyen des connaissances linguistiques. La *compréhension de texte*, par contre, pose surtout le problème de la résolution d'ambiguïté, qui, lui, se résout au moyen d'un ensemble très hétérogène de connaissances.

Il existe aujourd'hui un nombre important d'ouvrages ou d'articles faisant état de recherches en génération de texte: cf. McDonald 1980, McKeown 1982, Appelt 1985, Danlos 1985, Jacobs 1987 et Patten 1988. La génération de texte est devenue, en très peu d'années, un champ d'étude à part entière avec ses publications et ses conférences spécialisées. Ce champ d'étude débouche sur plusieurs types d'applications, par exemple:

- la traduction automatique;
- les interfaces d'accès aux bases de données;
- les "systèmes experts explicateurs";

- la production automatique de rapports à partir de données formalisées.

Nous avançons un second postulat qui concerne la caractérisation des modèles linguistiques adaptés à de telles applications:

Postulat 2:

Un modèle linguistique, pour être utile en génération de texte, doit AU MOINS: (i) être conçu pour établir une correspondance entre ce qu'exprime un énoncé et cet énoncé lui-même; (ii) décomposer en sous-systèmes plus simples le système très complexe de la langue qui établit une telle correspondance; (iii) être formel, en ce sens qu'il doit être constitué de règles et descriptions formalisées permettant une manipulation automatique.

Dans le cadre de nos recherches antérieures en génération de texte, nous avons voulu mettre en application l'approche linguistique qui répondait à ces critères tout en nous semblant la plus "plaisante" à utiliser. Notre raisonnement était le suivant:

1) UNE THEORIE LINGUISTIQUE EST UN OUTIL PERMETTANT DE CONSTRUIRE DES MODELES LINGUISTIQUES et est donc un outil de description des langues naturelles.
2) Plus un outil est adapté à sa fonction plus son utilisation semble plaisante; cette constatation s'applique aux marteaux, aux stylos, etc., mais aussi aux théories linguistiques.
3) Considérant le caractère éminemment linguistique des problèmes posés en génération de texte, l'approche théorique qui nous semblait la plus adéquate dans un contexte d'étude strictement linguistique devrait être tout aussi valide dans un contexte de génération de texte.

En conséquence, nous avons voulu démontrer qu'il était pertinent d'utiliser, dans un contexte d'application en génération de texte, une modélisation linguistique basée sur la *Théorie Sens-Texte* (*TST*). Une telle démonstration prend beaucoup de temps et ne peut se faire par la seule programmation d'un "système" qui s'insérerait dans le cadre d'un travail de doctorat.

Même si elle est basée sur un travail expérimental lié aux applications en génération de texte (cf. 1.3 ci-dessous, la présentation des systèmes GOSSIP et FoG), cette thèse n'a pas pour finalité de décrire un système de TALN et aucun système n'est le VERITABLE produit de cette recherche.

Soulignons que nous ne sommes pas les premiers à nous être intéressé à l'activation procédurale des descriptions Sens-Texte. Depuis maintenant de nombreuses années, Y. Apresjan et ses collaborateurs travaillent à la programmation de systèmes de traduction automatique fondés sur la TST (cf. Apresjan *et al.* 1984, 1985 et 1989). Cependant, c'est à M. Boyer (cf. Boyer 1984 et Boyer & Lapalme 1985) que l'on doit la première implémentation de la composante sémantique Sens-Texte, qui est l'objet de la présente étude (voir ci-dessous).

Un modèle d'une langue particulière construit à partir de la TST est appelé *Modèle Sens-Texte* (*MST*). Il est constitué (i) d'une grammaire, qui est un ensemble de règles déclaratives de correspondance entre des niveaux de représentation d'un énoncé, et (ii) d'un lexique formel, qui est au centre de la description linguistique. On peut donc dire que prouver la pertinence "computationnelle" des MST revient, en tout premier lieu, à définir un système procédural qui permet de mettre en jeu, d'activer, ces modèles essentiellement déclaratifs.

Il était hors de question de prétendre traiter ici tous les problèmes posés par l'activation des MST. Même si, dans le cadre de

nos expérimentations, nous avons abordé dans sa totalité la correspondance entre Sens et Texte, nous nous sommes ici focalisé sur une étape de cette correspondance que nous pouvons décrire ainsi:

> Nous étudions la mise en relation de la représentation du Sens d'un énoncé avec la représentation de sa structure syntaxique la plus proche de ce Sens — c'est-à-dire la plus "profonde". Formellement, le problème à résoudre est celui de la correspondance entre un réseau de Sens et une hiérarchie syntaxique de lexèmes, c'est-à-dire un arbre syntaxique.

Ce problème particulier a retenu notre attention pour deux raisons majeures:

1) La correspondance entre le niveau de la *Représentation Sémantique* (*RSém*) d'un énoncé et le niveau de la *Représentation Syntaxique Profonde* (*RSyntP*) est certainement la *composante des MST*[2] qui a été la moins étudiée même sur le plan strictement théorique.

Nous parlons ici de représentations des énoncés et non de représentations des lexèmes. Nous verrons en effet, dans le cours de l'exposé, que la correspondance entre les lexèmes de la langue et leur représentation sémantique est au contraire un des axes de recherche les plus développés en linguistique Sens-Texte.

2) La traduction d'un réseau en un arbre présuppose des opérations formelles très particulières.

---

[2] *Composante d'un MST*, de même que *RSém* et *RSyntP*, sont des termes techniques qui seront examinés en 2.

Il nous a semblé qu'il était important de rendre compte de ce type d'opérations pour ouvrir la voie à des implémentations "grandeur nature" de la TST, au moins comparables aux travaux entrepris sur l'implémentation des grammaires systémiques (voir 2.1. ci-dessous).

La recherche présentée ici est ambitieuse en ce sens qu'elle vise à établir des bases théoriques pour l'activation des MST; elle est très modeste en ce sens qu'elle isole, dissèque et analyse un problème dont la dimension paraît microscopique lorsqu'il est comparé à l'ensemble des problèmes que pose la génération de texte.

1.2 Génération de texte et synthèse linguistique

1.2.1 Les modèles linguistiques en TALN

Effectuons tout d'abord une mise au point terminologique. Traditionnellement, la *génération de texte* est considérée comme l'activité inverse de l'*analyse de texte* pour ce qui a trait à la communication homme-machine. Nous pensons qu'il existe cependant une disproportion entre ces deux termes. L'opération inverse de l'analyse est, quelque soit le domaine ou l'application considérés, la *synthèse*; or, la génération de texte présuppose une activité bien plus complexe que la *synthèse de texte*. Générer un texte, c'est produire un texte "à partir de rien", ou plutôt, à partir de rien qui soit textuel. L'opération inverse étant la *compréhension de texte* ou l'*interprétation de texte*; c'est-à-dire la production, à partir d'un texte, de quelque chose qui n'est pas une représentation de ce texte ou même de son contenu. Cela peut être, par exemple, un nouvel ensemble de connaissances ou la mise à jour d'une représentation de ce que l'on considère comme "l'état du Monde". La figure 1 récapitule la description terminologique des notions de génération, compréhension, synthèse et analyse (sous-entendu "de texte") selon les critères qui nous semblent pertinents et qui s'appuient sur la spécification des "entrées" et "sorties" de systèmes effectuant ces quatre opérations:

	entrées	sorties
GENERATION	• connaissances sur le Monde • bases de données	texte
COMPREHENSION (INTERPRETATION)	texte	• nouvelles connaissances sur le Monde • nouvelles bases de données
SYNTHESE	représentation du contenu d'un texte	texte
ANALYSE	texte	représentation du contenu d'un texte

Figure 1
Génération/compréhension *vs.* synthèse/analyse

Dans ce tableau, la spécification des entrées et sorties est volontairement floue pour permettre de tenir compte des variations dues aux différents contextes dans lesquels on trouve mises en jeu ces quatre opérations. (Les textes peuvent être sous une forme écrite ou sonore, les représentations du contenu peuvent être de type syntaxique, sémantique ou conceptuel, etc.) Dans le cas de la génération et de la compréhension, nous considérons des opérations qui sont impliquées soit dans l'utilisation "naturelle" de la langue, soit dans le fonctionnement de systèmes qui reproduisent cette utilisation. L'étude de la mise en jeu de ces opérations ne relève pas de la théorie linguistique mais plutôt de ce que l'on appelle les sciences cognitives ou de l'*Intelligence Artificielle* (*IA*) — application d'une programmation basée sur des concepts abstraits tentant, par son fonctionnement, de reproduire le raisonnement humain (cf. Schank 1987). La linguistique théorique et descriptive propose cependant des modèles de la langue qui peuvent être utilisés lorsque

l'on veut simuler l'activité linguistique. Plus précisément, les modèles linguistiques vont intervenir dans la structuration des opérations de synthèse et d'analyse telles que nous les avons définies ci-dessus. Il nous semble en effet que ces deux opérations correspondent à une activité tout à fait artificielle, liée à une certaine description formelle que l'on fait d'une langue (un modèle particulier). Rien ne prouve en effet qu'il est NATUREL d'analyser des phrases à l'aide d'un lexique et d'une grammaire d'une langue donnée. Ainsi, la compréhension de la langue est un phénomène extrêmement complexe sur la structuration duquel on ne possède pas pour l'instant beaucoup d'informations. On peut donc se demander quel serait la part d'un processus d'analyse dans un processus plus vaste de compréhension. Ces remarques valent aussi pour la synthèse et la génération. Nous pensons cependant que:

Il existe une plus grande corrélation entre génération et synthèse qu'entre compréhension et analyse.

Voici quelques remarques permettant d'appuyer cette affirmation:

1) "Parler grammaticalement" est une locution qui se comprend très bien comme signifiant que l'on parle en produisant des phrases qui respectent les règles de la grammaire. "Écouter (comprendre) grammaticalement" paraît une locution très étrange. On a d'abord un problème du fait qu'il n'existe pas dans la plupart des langues de verbe pour dénoter l'activité réciproque de 'parler', qui serait 'écouter la parole' (cf. Mel'čuk 1988c:19). Ensuite, on peut immédiatement constater que quelqu'un s'exprime ou ne s'exprime pas grammaticalement, mais il est beaucoup plus difficile de savoir si, en interprétant ce qu'on lui a dit, il a maîtrisé du même coup la composition grammaticale de nos phrases. Il n'y a donc pas de manifestation immédiate d'une compréhension grammaticale alors qu'il existe une manifestation immédiate d'une parole grammaticale.

2) Une conséquence de ce qui vient d'être dit est qu'il est LINGUISTIQUEMENT beaucoup plus difficile de faire preuve d'une expression correcte que d'une compréhension correcte puisque cette première activité est jugée aussi (et, dans certains cas, surtout) en fonction de la grammaticalité de ce que l'on produit. En toute logique, on devrait donc considérer qu'il est beaucoup plus important de réserver une place de choix à une composante linguistique dans un système effectuant de la génération que dans un système effectuant de l'interprétation: les performances du premier seront jugées par les utilisateurs autant en fonction de la qualité des informations transmises qu'en fonction de la qualité de la langue qui les véhicule. Même si ceci nous semble frappé au sceau de la logique la plus élémentaire, nous remarquons que la courte histoire du TALN montre une nette tendance des chercheurs à adopter des priorités inverses.
3) Il est possible, en parlant, de décomposer ce que l'on fait et de dissocier le moment où l'on "calcule" le contenu de notre phrase de celui où l'on calcule comment on va verbaliser cette phrase. Ceci est bien sûr particulièrement évident lorsque l'on parle une langue étrangère ou lorsque l'on s'exprime par écrit. Cette décomposition a ceci de particulier qu'elle semble isoler deux activités tout à fait SEQUENTIELLES. À l'inverse, la séparation des activités d'analyse linguistique et d'interprétation, bien que possible dans certains cas (voir, là encore, le cas de l'interprétation d'une phrase dans une langue que l'on maîtrise mal), pourra très difficilement mettre en évidence deux activités consécutives. On peut percevoir que l'on fait appel à nos connaissances grammaticales pour mener à bien l'interprétation d'une phrase, mais ce recours à la grammaire semble toujours fondu dans un processus global de compréhension guidé par des lois avant tout non linguistiques.

Ces remarques peuvent bien sûr être relativisées; elles ne font, de plus, que soulever des problèmes sans s'appuyer sur autre chose que

des perceptions plus ou moins intuitives de ce que sont les activités de génération et d'interprétation des textes. Nous pensons cependant qu'elles sont suffisantes pour mettre en évidence une disproportion entre le rôle des connaissances linguistiques dans la génération et leur rôle dans la compréhension — surtout lorsque l'on considère en plus la différence entre gestion de la paraphrase et résolution de l'ambiguïté (cf. le premier postulat énoncé en 1.1). Cette disproportion nous permet de considérer que les recherches sur la génération de texte sont un terrain privilégié pour effectuer un travail sur l'implémentation des modèles linguistiques. On peut en effet considérer qu'un système de génération efficace doit intégrer une composante de synthèse purement linguistique et que cette composante a tout avantage à être structurée en fonction d'un modèle linguistique donné.

Il est en revanche beaucoup plus difficile de "vendre" la nécessité du recours à un modèle linguistique dans le cadre d'une application de compréhension de la langue naturelle. Même si l'on peut considérer que dans l'absolu tout système devrait intégrer un modèle linguistique complet, il est douteux que les systèmes qui doivent comprendre la langue aient besoin d'effectuer une phase d'analyse linguistique PILOTEE par ce modèle.

Notons, pour conclure, que ces remarques tendent à définir le rôle des modèles linguistiques par rapport à des types donnés d'applications. D'un point de vue strictement théorique, cependant, l'approche Sens-Texte adoptée ici postule qu'un modèle linguistique doit être conçu de telle façon qu'il permet d'effectuer aussi bien la synthèse que l'analyse des phrases.

1.2.2 Le "véritable" titre de cette thèse

Si les travaux qui ont servi de champ d'expérimentation à la présente étude (voir 1.2.3) relèvent des applications en génération de texte, ce qui est présenté ici est directement connecté à un domaine d'étude linguistique et non à un domaine d'application. Pourquoi alors

parler de "génération de texte" dans le titre même de notre travail? Le terme exact à utiliser serait évidemment "synthèse de texte", mais un titre est destiné à ceux qui VONT lire ou NE VONT PAS lire un texte. Il ne concerne plus ceux qui lisent effectivement. Il s'agit donc (i) d'intéresser, et (ii) de sélectionner le public approprié. Il faut, dans un titre, tenir compte de ce que le lecteur DU TITRE ne sait pas. Ainsi, avant de lire la présente section, la plupart des lecteurs ne savent pas ce que nous entendons par génération *vs.* synthèse de texte. Il est presque sûr que l'expression "synthèse de texte", apparaissant dans le titre, serait lue comme "synthèse de texte" et finalement comprise comme "synthèse de la parole". S'il faut donc avoir lu un texte pour en comprendre le titre, nous ne voyons pas bien à quoi ce titre peut servir. Pour le même genre de raison, nous ne parlons pas de sémantique formelle dans notre titre même si la sémantique formelle est au centre de nos préoccupations dans la présente étude. Cependant, comme nous le montrerons dans la Partie **1**, le terme "sémantique" est ici un terme très technique, qui doit être clairement défini. Si nous avions annoncé dans le titre une recherche sur la sémantique formelle, quatre-vingt pour cent de nos lecteurs feraient, a priori, une mauvaise interprétation de ce terme, et seraient finalement déçus de ne pas trouver ici ce qu'ils cherchent (une étude des problèmes de représentation logique, par exemple). La formulation de notre titre n'est donc pas tout à fait exacte car elle est basée sur ce que nous savons être la terminologie généralement admise dans le domaine du TALN et non sur la terminologie qui est réellement la nôtre. Nous présenterons donc maintenant le "véritable" titre cette thèse, qui définit précisément les limites (étroites) de notre étude:

"Structuration et mise en jeu procédurale
de la composante sémantique d'un modèle Sens-Texte
dans un cadre de synthèse linguistique"

D'une certaine façon, il faudra attendre la Conclusion pour que toutes les notions permettant de comprendre ce titre aient été explicitées.

1.2.3 Les systèmes GOSSIP et FoG

Développé à Montréal dans le cadre des activités de recherches de ORA Inc., GOSSIP[3] (cf. Iordanskaja, Kittredge & Polguère 1988 et Iordanskaja & Polguère 1988) est un système prototype destiné à produire automatiquement des rapports en anglais sur l'activité d'un système d'exploitation[4]. On sait que la sécurité des ordinateurs est un des problèmes majeurs posés aujourd'hui par l'usage intensif de l'informatique dans tous les secteurs de l'activité humaine. Les risques associés à l'utilisation de l'informatique sont, entre autres:

1) les atteintes aux libertés individuelles dûes à une plus grande capacité à manipuler des informations très diverses sur les personnes;
2) les risques de pertes d'information (l'argent pouvant aussi être maintenant stockée sous forme de données informatiques) dues à des accidents ou à des erreurs de manipulation;
3) les risques de pertes d'information dues à des actes que l'on qualifie habituellement de "malveillants".

De nombreuses recherches sont actuellement en cours pour trouver des moyens de limiter les risques posés par l'utilisation de l'informatique. Si la première catégorie de risques mentionnée ci-dessus met en évidence un ensemble de problèmes fondamentaux, ces problèmes sont habituellement considérés par les gouvernements (et les organismes finançant la recherche) comme "éthiques" et ne suscitent donc que peu d'investigations. On comprendra aisément, en revanche, que les deux autres catégories de risques informatiques

---

[3] **G**enerating **O**perating **S**ystem **S**ummaries in **P**rolog.

[4] Un système d'exploitation correspond grosso modo à un ensemble de programmes qui sont installés sur un ordinateur pour en permettre le fonctionnement "logique" (par opposition à "physique").

aient permis l'émergence d'un nouveau secteur d'activité de recherche et de développement: la sécurité informatique.

Les recherches menées autour de GOSSIP visent la production de rapports concernant l'utilisation que les usagers font d'un système d'exploitation auquel ils ont accès. Il s'agit de faciliter la surveillance de l'activité des usagers pour déceler les anomalies pouvant être les indices d'une fraude, d'une tentative de pénétration, etc. De nombreuses recherches en surveillance des ordinateurs sont en cours depuis quelques années (voir, par exemple, Denning 1985). Un des outils de base permettant la surveillance est le programme d'audit. Ce programme enregistre des données de plus ou moins haut niveau sur l'activité du système d'exploitation; elles vont du simple enregistrement de certains événements élémentaires qui se produisent sur la machine ("tel usager a effectué telle opération à telle heure", etc.), jusqu'à une certaine analyse de ces événements afin d'enregistrer des informations plus complexes qui en ont été déduites. Ces audits sont bien souvent d'énormes bases de données dont l'examen détaillé est fastidieux et extrêmement couteux (compte tenu du fait qu'il doit être effectué presque en permanence). Dans ce contexte, la production automatique de rapports en langue naturelle qui feraient état d'une analyse critique des audits de sécurité peut être considérée comme un outil intéressant pour améliorer la surveillance d'un système d'exploitation. Dans le cadre de GOSSIP, la recherche a été limitée à l'étude de l'architecture d'un système de production de rapports pour ce qui est des composantes d'analyse des audits, de planification et de structuration du contenu des textes et, enfin, de production linguistique des textes proprement dite.

La composante de GOSSIP qui nous concerne ici est bien sûr celle prenant en charge la synthèse des phrases d'un rapport. Nous avons développé un mini-MST (anglais) pour expérimenter l'activation des règles déclaratives de la grammaire et l'exploitation de l'information contenue dans un lexique de type Sens-Texte. La définition initiale du projet nous a forcé à explorer, dans le cadre de l'implémentation du système, les principaux niveaux de traitement

conditionnés par le MST (du Sens au Texte)[5] sans pouvoir isoler véritablement une étape à traiter en profondeur. Ceci nous a permis d'avoir la satisfaction de faire "tourner" un synthétiseur complet Sens-Texte, mais a aussi créé beaucoup de frustrations. Les enseignements de cette expérimentation ont été précieusement collectés et les frustrations savamment cultivées comme autant de motivations pour approfondir la recherche dans un cadre plus théorique (la présente thèse).

FoG[6] (Bourbeau *et al.*, à paraître) n'est pas un prototype mais une véritable application qui doit devenir opérationnelle dans le courant de l'année 1990. Elle permet la production automatique de bulletins de prévision météorologique en anglais et en français à partir de données formalisées — plus précisément, la version FoG-89 de FoG prend en charge les prévisions maritimes pour l'est du Canada (émises à Gander et Halifax) ainsi que pour la région des Grands Lacs (émises à Toronto). Ce système de génération bilingue est lui-même une composante d'un système appelé FPA[7], développé à Toronto par Environnement Canada (McLeod, à paraître).

Comme GOSSIP, FoG prend en charge une phase de planification du contenu du texte, une phase de segmentation en phrases et une phase de synthèse linguistique. Son module de synthèse est lui aussi de type Sens-Texte et a été développé en important de façon très directe les méthodes de programmation mises au point pour GOSSIP. Ce module linguistique possède cependant les trois caractéristiques suivantes:

1) Il ne fait pas appel à un véritable niveau de représentation sémantique. Pour le traitement de chaque phrase, il reçoit en

---

[5] GOSSIP et FoG, qui sera présenté plus bas, n'utilisent qu'un seul niveau de représentation morphologique et ne prennent pas en compte les niveaux phonologiques (production de textes écrits).

[6] **Fo**recast **G**enerator.

[7] **F**orecast **P**roduction **A**ssistant.

effet en entrée une représentation déjà syntaxiquement hiérarchisée et qui est très proche d'une RSyntP de type Sens-Texte.

2) La structure reçue en entrée, par le module de synthèse linguistique, est une sorte d'interlingua puisqu'elle sera traduite en deux arbres syntaxiques profonds (un pour la phrase anglaise et un pour la phrase française). Cette opération de traduction est basée essentiellement sur l'introduction des lexèmes de la langue considérée et sur certaines opérations de restructuration de sous-arbres syntaxiques.
3) La couverture linguistique des deux grammaires et des deux lexiques construits est complète pour les sous-langages considérés (sur la notion de sous-langage, voir Kittredge & Lehrberger 1982).

L'utilisation d'une interlingua "syntaxique" est rendue possible dans FoG car la structuration des rapports en phrases et la hiérarchisation syntaxique globale de chacune de ces phrases sont identiques pour les deux sous-langages considérés (Polguère 1988b). Les textes (1) et (2), produits par FoG et extraits de Bourbeau *et al.* 1989, illustrent ces similitudes[8] :

(1) "Winds northerly 20 to 30 knots becoming northwest Wednesday morning. Snow and rain tapering to flurries Wednesday morning. Visibility zero to 3 in precipitation. Temperatures near zero." (p.9)

(2) "Vents du nord de 20 à 30 nœuds devenant du nord-ouest mercredi matin. Neige et pluie se changeant progressivement en averses de neige mercredi matin. Visibilité zéro à 3 sous les précipitations. Températures près de zéro." (p.10)

La similitude des structures syntaxiques profondes est en fait une motivation pour utiliser une interlingua syntaxique. Il existe une autre raison pour justifier le fait que cette interlingua soit syntaxique

---

[8] Nous avons légèrement modifié ici l'impression effectuée par FoG.

et non sémantique. En effet, comme nous le verrons en détail dans la première partie de cette thèse, le niveau de représentation sémantique Sens-Texte sert à rendre compte des phénomènes de paraphrase linguistique. Or, les sous-langages en général, et ceux que FoG traite en particulier, ne "connaissent" pas la paraphrase linguistique. Une fois isolé un ensemble d'informations à communiquer dans une phrase donnée, il n'existe pratiquement qu'une seule bonne façon de l'exprimer linguistiquement. Par exemple, la phrase de (2):

"Visibilité zéro à 3 sous les précipitations."

ne pourrait pas être paraphrasée par quelque chose comme:

"Précipitations donnant une visibilité de zéro à 3".

On peut dire que, pour ces sous-langages français et anglais, un certain contenu à exprimer dans une phrase IMPLIQUE un choix lexical et une hiérarchie syntaxique profonde donnés.

Pour conclure ces deux brèves présentations, nous voudrions insister sur un point particulier. Si un système comme FoG ne manipule pas de représentation sémantique de type Sens-Texte, ce n'est pas parce que le niveau de traitement sémantique linguistique pose trop de problèmes dans le développement de réelles applications. Cela tient uniquement au fait que le sous-langage traité n'admet pratiquement aucune paraphrase. Nous allons énoncer maintenant un troisième postulat. Selon nous:

Postulat 3:

Pour synthétiser un langage moins télégraphique et plus souple que le sous-langage traité par FoG, il est préférable d'avoir recours à un niveau de représentation sémantique afin de gérer la paraphrase.

Encore faut-il, cependant, disposer d'un "mode d'emploi" des composantes sémantiques des MST. Notre recherche représente, nous l'espérons, une étape vers la mise au point d'un tel mode d'emploi.

2 FONDEMENTS THEORIQUES

Nous ne proposons pas ici une véritable introduction à l'approche Sens-Texte qui a été décrite et commentée dans un nombre impressionnant de textes et publications, en plusieurs langues (anglais, français, allemand, russe,...) et pour divers publics (anthropologues, linguistes, informaticiens,...). On trouvera notamment des introductions à la TST dans les publications suivantes: Mel'čuk 1974, 1978, 1981a, 1981b, 1988a, 1988c, Mel'čuk & Pertsov 1987 et Mel'čuk & Polguère 1987.

La composante sémantique des MST, qui seule nous intéresse ici, sera par contre décrite de façon extrêmement détaillée. En effet, la Partie **1** de ce travail EST, dans sa totalité, une présentation et une étude de cette composante. Nous ne ferons donc ici que rappeler de façon très générale ce qu'est l'approche Sens-Texte et établir quelques mises au point terminologiques préalables. La TST va être présentée de façon comparative. Nous la situerons par rapport à une autre approche linguistique connue sous le nom de *linguistique systémique* (cf. Halliday 1961 et Halliday & Martin 1981). Nos motivations pour procéder de la sorte sont les suivantes:

1) Il nous semble que le travail le plus considérable qui ait été entrepris sur la synthèse linguistique, dans un cadre autre que la TST, est celui mené depuis maintenant de nombreuses années par le Penman Natural Language Generation Group dans le cadre du développement de la grammaire systémique NIGEL (cf. Mann & Matthiessen 1983, Penman 1988).

2) L'approche Sens-Texte et l'approche systémique se définissent toutes deux comme étant, notamment, "stratificationnelles" et "fonctionnelles". Il peut être intéressant de voir si cette similitude de termes correspond à une parenté de l'approche théorique.

Avant d'aborder la comparaison linguistique Sens-Texte *vs.* linguistique systémique, nous voudrions souligner que notre connaissance des grammaires systémiques est, pour l'instant, purement livresque. Il ne nous est donc pas possible de mener une comparaison très poussée et nous nous excusons par avance de certaines inexactitudes qui pourraient apparaître dans la présentation qui suit.

2.1 Comparaison de deux approches: linguistique Sens-Texte *vs.* linguistique systémique

L'approche Sens-Texte et l'approche systémique se définissent toutes deux comme étant, notamment, fonctionnelles et stratificationnelles. Nous allons montrer que cette apparente similitude des qualificatifs masque en fait des différences fondamentales.

Commençons tout d'abord par analyser le terme *fonctionnel* à travers l'examen d'un article de dictionnaire:

> "**FONCTION** (...)
> **I.Ⓑ.2** *Ling.* Ensemble des propriétés d'une unité par rapport au processus d'ensemble (communication); rôle dans l'énoncé (...)
> **II.1** *Math.* Relation qui existe entre deux quantités, telle que toute variation de la première entraîne une variation correspondante de la seconde (...)" (*Petit Robert* 1981:801)

Si de ces deux définitions nous déduisons deux lexèmes dérivés FONCTIONNEL**I** (dérivé de FONCTION**I.Ⓑ.2**) et FONCTIONNEL**II** (dérivé de FONCTION**II.1**), nous formulerons de la façon suivante une première distinction entre les deux approches considérées:

> Un modèle Sens-Texte d'une langue est un modèle fonctionnel**II** de cette langue alors qu'un modèle systémique en est un modèle fonctionnel**I**.

En effet, un MST est fonctionnel en ce sens qu'il est basé, notamment, sur plusieurs ensembles de règles mettant en correspondance des niveaux intermédiaires de représentation d'un énoncé. Chaque ensemble de règles décrit une fonction qui a chaque représentation d'un niveau N associe une représentation d'un niveau N_{+1}. Un modèle systémique est fonctionnel en ce sens qu'il vise à décrire la langue en tant qu'outil permettant d'assurer certaines fonctions dans un processus d'échange d'informations; il décrit aussi chaque énoncé ou chaque partie d'énoncé par ses fonctions à l'intérieur du système général de la langue. Un modèle systémique vise l'explication du fonctionnement communicatif de la langue plus que la description des structures valides de la langue (cf. Halliday 1985 pour une présentation des caractéristiques fonctionnelles des grammaires systémiques).

Ceci nous conduit à examiner le second terme: *stratificationnel* (cf. Lamb 1966). Un MST vise à décrire la correspondance entre un Sens et les multiples Textes qui en sont l'expression linguistique. Il fait appel à plusieurs niveaux intermédiaires de représentation d'un énoncé qui sont autant d'étapes logiques dans la description de la correspondance entre Sens et Texte. Les niveaux servent donc ici à décomposer la "chaîne de production" qu'est le modèle fonctionnel**II**.

"$\Leftrightarrow$" étant le symbole indiquant une correspondance entre deux niveaux de représentation, les SIX correspondances considérées par la TST sont les suivantes:

Représentation Sémantique (RSém)
⇔ *Représentation Syntaxique Profonde* (RSyntP)
⇔ *Représentation Syntaxique de Surface* (RSyntS)
⇔ *Représentation Morphologique Profonde* (RMorphP)
⇔ *Représentation Morphologique de Surface* (RMorphS)
⇔ *Représentation Phonologique Profonde* (RPhonP)
⇔ *Représentation Phonologique de Surface* (RPhonS)

Nous ne nous arrêterons pas ici sur la justification théorique de ces niveaux qui ont été maintes fois présentés dans la littérature (voir, par exemple, Mel'čuk 1981a). Nous voulons simplement mettre en évidence le fait que l'approche stratificationnelle Sens-Texte peut à la fois être caractérisée (i) par l'utilisation (fonctionnelle**II**) qui est faite des niveaux, et (ii) par le nombre et le type de niveaux envisagés.

Si l'on compare cette approche avec celle — stratificationnelle — de la grammaire systémique, on constate des différences importantes. M. Halliday considère le langage comme:

> "...a basically tristratal system: semantics, grammar, phonology." (Halliday 1978:39)

le niveau "grammatical" recouvrant ici grosso modo ce qui est inclus dans la syntaxe, la morphologie et le lexique Sens-Texte. Une grammaire systémique décrit, pour un niveau donné, l'ensemble des choix offerts au locuteur dans la production d'un énoncé. La terminologie utilisée est la suivante:

trait = C'est la "valeur" d'un choix. On peut donner comme exemple de traits: *indicatif*, *impératif*, *déclaratif*, *interrogatif*, *futur*, *présent*, *passé*, *avec douceur*, *de façon brutale*, etc.

système = C'est un regroupement de traits, s'excluant mutuellement, qui décrit un choix grammatical. Un *système* se représente graphiquement de la façon suivante:

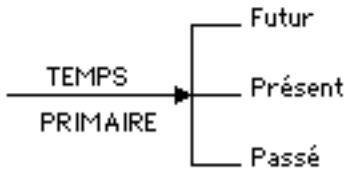


Figure 2
Exemple de *système* (tiré de Penman 1988:15)

réseau de systèmes =

C'est en quelque sorte un système de plus haut niveau ou un "système de systèmes". Il fait apparaître l'interdépendance entre les choix; un exemple typique de réseau de systèmes (pour l'anglais ou le français) serait:

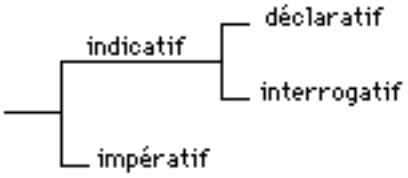


Figure 3
Exemple de *réseau de systèmes* (tiré de Patten 1988:25)

Les systèmes décrivent donc des "catégories grammaticales". Il faut cependant prendre ce terme dans une acception très large puisque les catégories envisagées peuvent être de toutes sortes — par exemple, rhétoriques: énoncé conciliant/dominateur/agressif, etc.

Une grammaire systémique fonctionne avec deux types de règles: (i) des règles permettant d'effectuer le passage à travers les réseaux de systèmes — et permettant donc de sélectionner un ensemble de traits à réaliser, (ii) des règles permettant la réalisation linguistique proprement dite — c'est-à-dire la mise en relation d'un ensemble de traits avec une structure linguistique. Ce sont ces dernières règles qui assurent la correspondance entre les différents niveaux de description mentionnés ci-dessus. Chaque composante d'un énoncé possède une description fonctionnelle selon plusieurs plans (héritée des choix systémiques correspondants): la composante

correspond à un "Acteur", un "Processus", etc., c'est un thème ou un rhème, etc.

Sans aller plus avant dans l'étude comparative de la structure d'un modèle Sens-Texte et celle d'un modèle systémique, on voit que les deux approches diffèrent radicalement dans le statut donné à la notion de niveau. Les niveaux des modèles Sens-Texte sont des NIVEAUX DE REPRESENTATION qui conditionnent des règles de correspondance (ou de traduction, cf. Mel'čuk 1978) et permettent de rendre compte du fonctionnement de la langue. Les niveaux des modèles systémiques sont des NIVEAUX DE TRAITEMENT qui conditionnent des règles d'interprétation de l'information qu'ils contiennent (encodée sous forme de *systèmes*) et servent, finalement, à rendre compte de la fonction de chaque élément d'un énoncé grâce aux traits particuliers qui lui sont associés.

L'approche systémique est le résultat d'un mûrissement progressif de réflexions sur le langage énoncées dans un cadre d'études anthropologiques — travaux de B. Malinowski; ces études ont ensuite été exploitées dans des recherches plus spécifiquement linguistiques (cf. Firth 1957 et Halliday 1961). On trouvera dans Patten 1988:3.1 un bref historique décrivant ce continuum qui unit les travaux en anthropologie de B. Malinowski à ceux, linguistiques, de J.R. Firth et M. Halliday[9]. L'approche systémique a en quelque sorte hérité génétiquement de ces origines anthropologiques deux postulats voulant que la description linguistique (i) tienne compte du contexte socio-culturel d'utilisation de la langue, et (ii) soit fondée sur les différentes fonctions remplies par la langue en tant que système sémiotique. Ces postulats ne sont en aucun cas repris par la TST. Rappelons que l'approche Sens-Texte se base, elle, sur trois postulats très différents que nous reformulerons ainsi (pour une présentation plus détaillée, voir Mel'čuk 1988c:14-18):

---

9 On peut aussi consulter Winograd 1983 pour une présentation succincte des grammaires systémiques.

1) La langue est considérée comme un système établissant une application représentée par la correspondance

$$\{RSém_i\} \Leftarrow \text{langue} \Rightarrow \{RPhon_j\} \quad , \quad 0 < i , j < \infty$$

2) L'application ci-dessus doit être décrite, pour chaque langue, par un modèle fonctionnel**II**.
3) Pour établir une modélisation adéquate de l'application ci-dessus, il faut avoir recours à des niveaux de représentation intermédiaires et considérer l'application initiale comme étant la composée[10] de six applications distinctes:
   - $\{RSém_i\} \Leftarrow$ sémantique $\Rightarrow \{RSyntP_k\}$
   - $\{RSyntP_k\} \Leftarrow$ syntaxe profonde $\Rightarrow \{RSyntS_l\}$
   - $\{RSyntS_l\} \Leftarrow$ syntaxe de surface $\Rightarrow \{RMorphP_m\}$
   - $\{RMorphP_m\} \Leftarrow$ morphologie profonde $\Rightarrow \{RMorphS_n\}$
   - $\{RMorphS_n\} \Leftarrow$ morphologie de surface $\Rightarrow \{RPhonP_o\}$
   - $\{RPhonP_o\} \Leftarrow$ phonologie $\Rightarrow \{RPhonS_j\}$

Pour conclure cette comparaison des deux approches, on peut dire que la TST privilégie la description des structures que l'on trouve dans une langue alors que l'approche systémique privilégie la description des fonctions**I.Ⓑ.2** des éléments de la langue car elle s'intéresse à la façon dont la langue est UTILISEE. Une description Sens-Texte ne procure pas une compréhension immédiate de ce qu'est le phénomène langagier dans son ensemble. Elle est néanmoins utilisable par tous ceux qui veulent "faire quelque chose avec la langue". On trouve normal que la chimie moderne nous donne les moyens de modéliser la structure moléculaire de l'eau même si cela n'explique pas directement le fait qu'un verre d'eau est plus agréable à boire qu'un verre de boue, ou que l'eau peut intervenir dans l'évocation de la nostalgie amoureuse:

"Sous le pont Mirabeau coule la Seine" (Apollinaire)

---

10 Au sens mathématique du terme.

Pourtant, il est certainement possible de trouver un lien entre ces deux observations à propos de l'eau et la composition chimique d'une molécule d'eau (qui conditionne, notamment, son comportement physique). La position, sans doute sectaire, de la TST sera donc de dire que: de même qu'il n'appartient pas à la chimie de décrire ce que l'on fait quotidiennement avec les corps qu'elle étudie, de même, il n'appartient pas à la linguistique formelle SENS-TEXTE de décrire ce que l'on peut faire avec la langue naturelle.

2.2 Mises au point terminologiques

Nous avons déjà expliqué plus haut ce que nous entendons par théorie (Sens-Texte) *vs.* modèle (Sens-Texte). Nous avons aussi présenté les sept niveaux de représentation mis en jeu dans un MST et les six correspondances que ces niveaux impliquent.

Nous allons maintenant effectuer quelques mises au point terminologiques à propos des quatre notions de: *composante d'un MST*, *signe linguistique*, *union linguistique* ⊕ et *transition*.

composante d'un MST

On appelle *composante d'un MST* un sous-ensemble donné de la grammaire de ce modèle qui établit la correspondance entre deux niveaux de représentation adjacents. Chaque composante porte le nom du niveau de représentation le plus profond mis en jeu. Il existe donc six composantes dans un MST — correspondant aux six correspondances définies en 2.1: la *composante sémantique*, la *composante syntaxique profonde*, la *composante syntaxique de surface*, la *composante morphologique profonde*, la *composante morphologique de surface* et la *composante phonologique profonde*.

signe linguistique

La notion Sens-Texte de *signe linguistique* est en quelque sorte une extension de la notion définie originellement par F. de Saussure[11]. Un signe linguistique (Sens-Texte) S_i est un triplet constitué d'un *signifié* $S_é$, d'un *signifiant* S_a et d'un *syntactique* S_y pour la paire $<S_é,S_a>$.

Une présentation détaillée de la notion de signe linguistique et de ses composantes n'est pas nécessaire ici. Rappelons cependant que, dans le cadre de la TST:

- Le $S_é$ de S_i est une information linguistique communicable qui est associée à S_i.
- Le S_a de S_i est le moyen d'expression linguistique du $S_é$ de S_i.
- Le S_y de S_i est l'ensemble de toutes les informations décrivant la combinatoire linguistique de la paire $<S_é,S_a>$ associée à S_i.

Il existe une très grande variété de signes linguistiques; en voici deux exemples pour français (choisis parmi des millions d'autres!):

(3) CORSAGE $S_é$ = 'vêtement féminin ...'
S_a = /kɔʀsaʒ/
S_y = racine nominale, masculin, morphologie régulière, ...

(4) -ER $S_é$ = "infinitif"
S_a = /-e/
S_y = suffixe, peut apparaître dans le S_y des verbes du premier groupe, ...

union linguistique ⊕

⊕ est appelée "méta-opération d'*union linguistique*". X ⊕ Y est l'union linguistique des deux signes ou objets linguistiques X et Y.

---

11 Cette extension n'est absolument pas triviale. Cf. Mel'čuk 1982 et Mel'čuk (à paraître, Première Partie, II.1) pour des présentations et justifications approfondies de la notion Sens-Texte de signe linguistique.

On peut ainsi appliquer cette opération à des signifiés — 'à nouveau' ⊕ 'partir' = 'repartir', à des signifiants — /ʀ(ə)/ ⊕ /paʀtiʀ/ = /ʀ(ə)paʀtiʀ/, etc.

transition

Cette notion n'est pas une notion Sens-Texte "standard" mais elle est définie pour les besoins de la présente étude:

> Une *transition* entre deux niveaux de représentation N_x et N_y est la mise en relation de ces deux niveaux selon une orientation synthétique, symbolisée par $N_x \Rightarrow N_y$, ou analytique, symbolisée par $N_x \Leftarrow N_y$.

Nous pouvons déduire de cette définition la remarque suivante:

> Dans ce travail, la seule transition considérée sera la *transition (synthétique) RSém ⇒ RSyntP*.

Cette transition sera aussi appelée *transition sémantique*.

3 STRUCTURE DE LA THESE

Cette thèse se divise en deux parties. La Partie **1** (Chapitres I et II) est consacrée à la présentation détaillée de la correspondance RSém ⇔ RSyntP. Elle couvre, en quelque sorte, les aspects purement déclaratifs du problème à traiter. La Partie **2** (Chapitres III, IV, V et VI) est consacrée à l'étude de la transition RSém ⇒ RSyntP. Elle couvre, en quelque sorte, les aspects plus procéduraux du problème à traiter.

On voit que la Partie **1** traite avant tout de la structure du modèle linguistique lui-même à travers la description des données grammaticales et lexicales mises en jeu par la composante sémantique. La Partie **2** traite du modèle de synthèse linguistique

qui, lui, englobe ce premier modèle "purement" déclaratif. Les deux parties sont donc extrêmement liées, en voici la composition détaillée.

3.1 Partie 1: "La Composante Sémantique d'un MST"

La présentation des règles sémantiques d'un MST ne se ramène pas à une simple formalité. En effet, on peut raisonnablement dire qu'il n'existe aucun travail traitant explicitement de la correspondance sémantique Sens-Texte, même dans le cadre de la modélisation d'une langue particulière. Les efforts dans ce domaine ont été concentrés sur le travail de description lexicographique. Même si, comme nous le verrons, les entrées de dictionnaire sont impliquées dans la composante sémantique, elles ne servent à traiter qu'une partie du problème; de plus, elles ne sont pas nécessairement opérationnelles d'un point de vue procédural dans leur formulation actuelle. Le but de la Partie **1** est ainsi de décrire formellement les différentes règles déclaratives de la composante sémantique, ce qui présuppose les tâches suivantes:

1) préciser les formalismes déjà existants pour les rendre exploitables par des procédures dans un cadre de transition RSém ⇒ RSyntP;
2) proposer des formalismes pour les règles qui n'avaient jamais été étudiées auparavant;
3) préciser le rôle de chaque type de règle dans la composante sémantique en général, mais aussi dans un contexte de transition sémantique.

Il est clair ici que, si nous avons pu faire l'économie d'une présentation détaillée de l'approche Sens-Texte dans cette Introduction (en renvoyant pour l'essentiel à l'abondante littérature déjà existante), nous sommes obligé de disséquer la composante particulière sur laquelle nous travaillons avant d'aborder le problème de la transition sémantique proprement dit.

Cela explique qu'un certain nombre de notions nouvelles devra être introduit dans cette partie; ces notions seront ensuite précisément définies ET JUSTIFIEES dans la Partie **2**.

Nous pensons pouvoir montrer qu'une nouvelle analyse des représentations et des règles mises en jeu dans la composante sémantique permet de mieux les situer dans le cadre de la transition RSém ⇒ RSyntP. Elle permet notamment de mieux cerner le lien théorique existant entre les deux niveaux de représentation considérés et de mieux décrire à quelles conditions, selon quelles méthodes, une RSyntP peut être associée à une RSém. La Partie **2** ira quant à elle plus loin, puisqu'elle décrira les conditions et méthodes présidant à la CONSTRUCTION d'une RSyntP à partir d'une RSém. L'utilité principale de la Partie **1** sera donc de décrire minutieusement la composante sémantique afin de préparer la démonstration de la thèse suivante:

Thèse:

A partir de l'examen des caractéristiques formelles des RSém et des RSyntP, ainsi que des différentes règles sémantiques d'un MST, on peut élaborer une méthode rigoureuse et générale d'organisation de la transition RSém ⇒ RSyntP.

C'est dans la Partie **2** que nous décrirons la méthode d'organisation de la transition sémantique que nous sommes parvenu à élaborer. Nous verrons que celle-ci est intéressante, pour une théorisation de la transition sémantique, de deux points de vue:

1) Elle modélise la façon dont un utilisateur du MST "lit" une RSém pour interpréter son contenu sous la forme d'une phrase de la langue.

2) Elle permet de décrire le rôle de la structure communicative d'une RSém dans le processus de construction d'un arbre syntaxique.

Ce dernier point est à notre avis très important puisqu'il n'existait pas, jusqu'à présent, de méthode formelle d'utilisation de la structure communicative, dans un cadre de synthèse, afin de contrôler les choix lexicaux et syntaxiques. (Dans la plupart des recherches antérieures, seul l'ordre linéaire des éléments de la phrase avait été étudié comme manifestation de la structure communicative.)

La Partie **1** est divisée en deux chapitres qui peuvent être résumés de la façon suivante:

1) Le **Chapitre I** présente le formalisme des représentations mises en jeu par la composante sémantique d'un modèle Sens-Texte. Nous décrirons donc les RSém et les RSyntP en tant qu'outils formels de représentation et élaborerons quelques justifications des formalismes ainsi que des niveaux de représentation eux-mêmes. Notons cependant que notre but ne sera pas de mener ici un débat théorique sur la validité de l'approche Sens-Texte et la justification des outils qu'elle propose d'utiliser dans la description linguistique. Nous éviterons donc au maximum les problèmes théoriques qui risqueraient de nous lancer dans une polémique puisque tel n'est pas le propos de cette thèse et que rien n'est plus inutile qu'une polémique menée à moitié.
2) Le **Chapitre II** décrit les règles déclaratives qui selon nous constituent la composante sémantique d'un MST. Il s'agit (i) des *règles de bonne formation*, (ii) des *règles de paraphrase sémantique*, (iii) des *règles grammatico-sémantiques*, et (iv) des *règles de correspondance sémantique proprement dites*[12]. Nous n'avons pas ici la prétention de traiter complètement le problème d'une telle

---

[12] La présentation des règles de correspondance sémantique proprement dites comprend aussi une Annexe, située à la fin du Chapitre II.

description mais nous introduirons les données minimales nécessaires à notre présentation des procédures d'activation. Nous définirons notamment, lors de la description de ces règles, quatre opérations de transition RSém ⇒ RSyntP, qui seront étudiées dans la Partie **2** : (i) la *hiérarchisation*, (ii) la *grammaticalisation*, (iii) la *lexicalisation au sens propre*, et (iv) l'*arborisation*.

3.2 Partie 2: "La Transition Sémantique"

La Partie **2** est dédiée à la description de la transition sémantique (ou transition RSém ⇒ RSyntP). Son but est de montrer comment il est possible de mettre en jeu les règles sémantiques déclaratives d'un MST dans un cadre de synthèse linguistique en implémentant les quatre opérations de base mentionnées ci-dessus.

Nous mettrons tout particulièrement l'accent sur la description de l'opération de *hiérarchisation* qui est la clé du système procédural que nous proposons. En ce sens, les trois autres opérations seront abordées essentiellement au niveau de leur interaction avec les notions définies pour décrire la hiérarchisation. Le système proposé est loin d'être complet mais il est, à notre avis, suffisamment structuré pour représenter un net progrès dans la compréhension de l'activation de la composante sémantique des MST.

Bien que nous essayions toujours de limiter au maximum la construction de nouveaux outils théoriques, il nous faudra, dans cette partie, définir certaines notions nouvelles car nous abordons des problèmes qui n'avaient pas été traités systématiquement auparavant. Nous serons aussi amené à proposer de nouvelles descriptions de notions telles que l'opposition thème/rhème ou la présupposition, descriptions qui rendent compte plus adéquatement du rôle spécifique de ces notions dans le cadre d'une approche Sens-Texte de la modélisation linguistique. EN CE SENS, CETTE SECONDE PARTIE N'A PAS SIMPLEMENT POUR BUT DE DECRIRE UN ENSEMBLE DE PROCEDURES; ELLE VISE

AUSSI à PRECISER LES OUTILS THEORIQUES NECESSAIRES A LA MODELISATION DE LA TRANSITION SEMANTIQUE.

Les procédures d'application des différentes règles qui constituent la composante sémantique d'un MST permettent d'effectuer une analyse de la représentation de départ en même temps qu'elles nous disent quand appliquer les règles déclaratives du modèle. Nous appellerons *parcours sémantique* une opération consistant à examiner systématiquement et complètement une représentation. Cette notion sera définie avec précision et nous verrons que deux parcours sémantiques distincts sont nécessaires en vue d'assurer, d'une part, la *hiérarchisation*, et d'autre part, l'*arborisation*.

Pour déterminer ces parcours, deux méthodes sont possibles:

1) une *méthode heuristique*, qui consiste à examiner le réseau à l'aide d'une collection de règles dont l'ordre d'application n'est pas contraint a priori;

2) une *méthode procédurale* (ou *algorithmique*), qui consiste à déterminer un ordre précis de parcours des nœuds et d'application des différentes règles.

Du point de vue d'une implémentation informatique, il est évident qu'un traitement n'est jamais purement heuristique, mais que l'on peut utiliser des programmations (toujours algorithmiques par nature) qui modélisent une approche heuristique.

Nous énonçons deux nouveaux postulats qui définissent notre approche de l'activation de la composante sémantique Sens-Texte:

Postulat 4:

L'utilisation d'une simple approche heuristique basée sur la mise en jeu non-ordonnée des règles du modèle est condamnée par l'interaction de toutes les contraintes potentielles sur l'application de ces règles.

Postulat 5:

L'examen d'une RSém de type Sens-Texte ne peut s'effectuer de façon sûre et systématique qu'à partir d'une approche qui soit la plus procédurale possible.

Il est bien sûr possible de modéliser le niveau d'application des règles sémantiques selon une approche heuristique, mais nous sommes convaincu qu'une telle approche limite par avance la complexité des réseaux qu'elle peut traiter de façon cohérente. Si nous ne sommes pas en mesure de prouver les deux credo énoncés ci-dessus, nous pensons en revanche qu'il est possible, dans le cadre de ce travail, de démontrer l'existence d'un ensemble de procédures de parcours des RSém relativement bien structuré et permettant la transition vers le niveau syntaxique profond.

La Partie **2** se divise en quatre chapitres, qui peuvent être résumés de la façon suivante:

1) Le **Chapitre III** précisera ce que nous entendons par transition sémantique et récapitulera les opérations impliquées dans cette transition. Certaines nouvelles notions de base, notamment les notions de *parcours sémantique* et de *réseau sémantique réduit*, seront aussi définies.
2) Le **Chapitre IV** prendra en charge la première étape de l'opération de hiérarchisation: la détermination du *nœud dominant*. Après avoir défini plus précisément cette notion — et

les liens qu'elle entretient avec la structure communicative d'une RSém, nous proposerons un ensemble de règles de calcul du nœud dominant. Nous examinerons le rapport existant entre de telles règles et les règles de bonne formation des RSém et RSyntP, décrites dans la Partie **1**.

3) Le **Chapitre V** complétera la description de la hiérarchisation en mettant à jour les procédures permettant d'effectuer le parcours sémantique des RSém. Des observations importantes seront aussi faites sur la *topographie des réseaux sémantiques* et, notamment, sur la pertinence de la prise en compte des composantes présuppositionnelles.

5) Finalement, le **Chapitre VI** décrira l'arborisation, essentiellement à partir d'un exemple précis de RSém relativement complexe. Nous montrerons comment un parcours sémantique, équivalent à celui décrit au Chapitre V, peut être utilisé pour activer les règles de correspondance sémantique dans le cadre de l'arborisation.

Arrivé au terme de cette dernière partie, toutes les notions qui, selon nous, sont nécessaires à l'organisation de la transition sémantique auront été présentées; de plus, les différents ajustements de la composante sémantique des MST, proposés dans la Partie **1**, auront trouvé leur justification.

Tous les résultats qui sont exposés dans cette thèse sont fondés sur une étude à la fois théorique et pratique de la transition sémantique. En ce qui concerne le problème de la *lexicalisation* — choix des lexèmes apparaissant dans la RSyntP, et celui de la *grammaticalisation* — choix des composantes sémantiques qui seront réalisées grammaticalement, il faut admettre que nous n'avons pu développer de façon satisfaisante les aspects appliqués de notre travail qui auraient pu nous permettre de tester certaines hypothèses toutes théoriques présentées ici. En effet, un travail d'expérimentation des procédures de lexicalisation, notamment, présuppose les deux activités suivantes:

1) la construction d'un lexique important comportant des définitions complètes qui permettent d'établir des liens sémantiques entre les différents lexèmes décrits;
2) la production d'un grand nombre de réseaux sémantiques présentant une grande variété de configurations tant au niveau de leur structure sémantique que de leur structure communicative;
3) le traitement de ces RSém à l'aide des règles de paraphrase sémantique (les définitions) afin de faire varier ce que nous appellerons leur *topographie* ainsi que leur niveau de décomposition sémantique.

Un tel travail représente une recherche en soi, recherche qui n'a pu être entreprise dans le cadre du projet GOSSIP, et encore moins dans le cadre du projet FoG. En conséquence, aucun chapitre de cette thèse n'est dédié aux opérations de lexicalisation et grammaticalisation. Toutefois, certaines données réunies à ce propos, qui pourraient être utiles lors de futures recherches, apparaissent dans le cours de l'exposé.

Partie 1

La Composante Sémantique d'un Modèle Sens-Texte

CHAPITRE I

Caractérisation formelle des RSém et des RSyntP

La composante sémantique d'un MST décrit la correspondance entre le niveau sémantique et le niveau syntaxique profond de représentation d'un énoncé. Elle est composée de trois ensembles de règles:

- les règles de bonne formation de la représentation sémantique;
- les règles de bonne formation de la représentation syntaxique profonde (qui appartiennent aussi à la composante syntaxique profonde d'un MST);
- les règles statiques de correspondance entre les deux niveaux.

Dans un cadre opérationnel (synthèse linguistique, analyse linguistique, etc.), il faut bien sûr ajouter les règles procédurales de mise en jeu des trois premiers ensembles de règles.

La description des correspondances proprement dites présuppose que l'on définisse d'abord les formalismes des niveaux de représentation mis en jeu par la composante sémantique. C'est ce que nous allons faire maintenant en cherchant avant tout à compléter les descriptions qui en sont habituellement proposées (voir, par exemple, Mel'čuk 1981a et Mel'čuk & Polguère 1987). Nous procéderons en cinq étapes: (i) rappel de qu'il faut entendre par "Sens" dans l'approche Sens-Texte, (ii) étude de quelques aspects formels

des RSém, (iii) survol d'autres approches utilisant les réseaux sémantiques, (iv) description des RSyntP, et (v) justification théorique de ces deux formalismes et niveaux de représentation.

Remarque: Le lecteur qui voudrait consulter, à fin d'illustration, des exemples de RSém et de RSyntP relativement complexes pourra se reporter à la RSém de la figure IV-5 (Chapitre IV, Partie **2**) et à la RSyntP correspondante de la figure VI-11 (Chapitre VI, Partie **2**).

1 RAPPEL DES PRINCIPES DE BASE DE LA SEMANTIQUE SELON LA TST

1.1 Sens et paraphrase

Mel'čuk 1988b:15 définit ainsi le Sens d'un énoncé:

> "Le sens est l'invariant des paraphrases langagières, c'est-à-dire la seule propriété commune à tous les énoncés ayant le même sens."

Cette définition, qui semble à première vue circulaire, est explicitée par I. Mel'čuk de la façon suivante. Il faut comprendre ici "même sens" comme un terme technique renvoyant à une notion plus primitive que le Sens lui-même. Le "même sens", c'est-à-dire la synonymie de deux énoncés, est perçu INTUITIVEMENT par tout locuteur sans que ce phénomène puisse être décrit linguistiquement plus en profondeur et, surtout, sans qu'il soit fait appel, pour cela, au Sens de ces énoncés. Ainsi, tout locuteur du français va immédiatement percevoir que *"les gens qui habitent la maison d'en-face"* et *"les habitants de la maison d'en-face"* sont des paraphrases, sans être nécessairement en mesure de donner une quelconque description de leur Sens qui rendrait compte de ce lien paraphrastique. Un peu comme la perception de la grammaticalité, la perception du lien de paraphrase entre deux énoncés est une chose immédiate qui vient avant la modélisation

linguistique. Du point de vue méthodologique, pour explorer le Sens des énoncés, il faut donc se baser sur la notion de "même sens".

Nous ne nous aventurerons pas plus dans ces considérations pour nous limiter à un bref approfondissement de la notion de paraphrase langagière.

Quand peut-on dire que deux phrases sont des paraphrases? Doit-on se référer à des notions comme la dénotation ou la valeur de vérité pour définir la paraphrase, ou bien la perception intuitive de la synonymie est-elle le seul critère acceptable?

Le terme de "paraphrase", employé seul, est ambigu et peut recouvrir des phénomènes assez différents. Soit les trois phrases suivantes:

(1a) *On ne parvient pas à trouver les plans de la fusée lunaire.*

(1b) *Les plans de la fusée pour la Lune sont introuvables.*

(1c) *On ne parvient pas à trouver les plans de la fusée du professeur Tournesol.*

Nous pouvons affirmer que, pour tout locuteur du français possédant certaines connaissances littéraires[1], ces phrases EMISES DANS UN CONTEXTE DONNE pourraient communiquer la même information. D'un point de vue logique, on dirait alors qu'elles dénotent la même situation, ou qu'elles ont la même valeur de vérité. Nous dirons quant à nous qu'elles sont en relation de *paraphrase cognitive*. Il est certain, cependant, que (1a) et (1b) entretiennent une relation encore plus étroite. Non seulement elles peuvent dénoter la même situation, mais, de plus, tout locuteur du français pourrait dériver l'une de l'autre simplement en utilisant sa connaissance de la langue française. Si un soi-disant locuteur du français n'était pas à même d'effectuer une telle dérivation, cela serait à coup sûr un indice d'une certaine méconnaissance de la langue. Dans le cas de (1c), en revanche, d'une part la paraphrase (cognitive) n'est pas nécessairement évidente pour tout le monde, d'autre part la dérivation à partir de (1a) s'effectue en utilisant des connaissances essentiellement non-linguistiques. On dira donc que (1a) et (1b) sont,

---

[1] Voir Hergé 1953.

en plus d'être en relation de paraphrase cognitive, en relation de *paraphrase langagière*.

La paraphrase langagière implique la paraphrase cognitive mais la RSém d'une phrase est son invariant de paraphrase langagière uniquement. Nous récapitulerons donc ainsi nos termes définissant le niveau de représentation sémantique selon la TST:

RSém,	représentation de l'invariant de paraphrase langagière;
paraphrase,	relation d'équivalence informative entre deux phrases (Textes);
langagier,	qui est particulier à chaque langue.

En conclusion, la RSém d'une phrase sera à la fois un reflet direct de son contenu informatif et un reflet de la structuration du lexique d'une langue donnée. Cette conception très limitée (et technique) du Sens ne doit pas être confondue avec celle en vigueur dans les approches plus cognitives de la description du phénomène langagier où *représentation sémantique* signifie bien souvent *interlingua*. En règle générale, les travaux non linguistiques traitant de la notion de sens ont souvent tendance à imbriquer celle-ci dans la notion plus vaste de *concept*, sans donner explicitement ses caractéristiques propres. Un bon exemple, bien qu'un peu dépassé maintenant, est Quillian 1968, qui tend à utiliser la plupart du temps le terme *concept* sauf lorsqu'il parle de certains concepts en tant qu'objets de la pensée exprimables par des mots de la langue. Dans ce dernier cas, on voit apparaître les termes *sens* et *sémantique*.

La sémantique couverte par la TST ne doit donc aucunement être considérée comme LA (seule et la vraie) sémantique[2]. *Sens* sera toujours écrit ici avec une majuscule initiale pour particulariser cette acception technique du terme.

---

[2] Mel'čuk (à paraître:Introduction Ch.II) oppose d'ailleurs le Sens dans la TST au "vrai sens".

1.2 Théorie sans texte

Nous n'avons pu nous empêcher de faire ce jeu de mots douteux car il attire l'attention sur un problème véritable de limitation de la TST. Nous venons de voir que, dans le cadre de cette théorie, le terme *Sens* devait être compris dans une acception bien précise, justifiée à notre avis, qui limite le champ d'application de la théorie au domaine strictement linguistique. *Texte* pose un peu le même type de problème. Pour les personnes non familiarisées avec l'approche Sens-Texte, ce terme laisse croire que la théorie prend en compte les problèmes dits de linguistique textuelle alors que le texte, en tant que COMBINAISON DE PHRASES COHERENTE AU NIVEAU LOGIQUE, INFORMATIF, ETC., est justement écarté de la description linguistique Sens-Texte. Le niveau d'énoncé considéré ici est exclusivement celui de la phrase et de ses composants. *Texte* signifie donc la représentation ultime d'une phrase ou de n'importe quelle partie de phrase. Le Texte est ainsi, en termes plus techniques, la *Représentation Phonologique de Surface* de ce qui est émis par le locuteur et perçu par le destinataire d'un énoncé phrastique. Nous pensons que l'ambiguïté est un peu malheureuse mais nous devons reconnaître que nous n'avons pas trouvé de meilleur terme. Nous nous contenterons d'écrire toujours avec une majuscule initiale *Texte* pour signifier le niveau le plus superficiel de représentation linguistique (qui est, dans l'approche Sens-Texte, le niveau de *RPhonS*) et simplement *texte* pour un ensemble cohérent et autonome de plusieurs phrases adjointes.

Une description linguistique qui se limite au niveau de la phrase est nécessairement insuffisante. En effet, les principes de cohérence textuelle qui pourraient être formulés en des termes très généraux s'appliquent différemment selon les langues, en utilisant des outils propres à chacune d'entre elles. De plus, la paraphrase, qui joue au niveau de la phrase, existe aussi au niveau du texte. Comparons les textes (2) et (3) ci-dessous. Le texte (2) est la reproduction exacte

d'un entrefilet du journal *La Presse*[3]; le texte (3) en est une paraphrase (un plagiat) que nous avons nous-mêmes commise:

> (2)
> "LA LOI DU DÉSIR: Pablo Quintero, cinéaste réputé, est un homosexuel qui a de nombreuses aventures tout en étant follement amoureux de Juan, jeune employé d'un bar d'Andalousie. Il est aussi attaché à sa sœur Tina, une actrice transexuelle qui fut jadis son frère, ainsi qu'à la fillette dont elle est la mère adoptive. Un nouvel ami de Pablo, Antonio, se prend d'une jalousie féroce envers Juan et en vient à la (sic)[4] tuer. L'enquête policière qui s'ensuit cause divers ennuis à Pablo et à ses proches."

> (3)
> LA LOI DU DÉSIR: Le fameux cinéaste Pablo Quintero est très attaché à son frère. Ce dernier, qui se fait appeler Tina, est devenu une actrice transexuelle et a adopté une fillette qui compte aussi beaucoup dans la vie mouvementée de Pablo (marquée par de nombreuses aventures homosexuelles). Une enquête, déclenchée par la police à la suite d'un acte de jalousie féroce, va venir perturber l'existence de ces personnes unies par l'affection. En effet, Juan, jeune homosexuel travaillant dans un bar andalous et dont Pablo était follement amoureux, vient d'être tué par Antonio, un ami de fraîche date du cinéaste[5].

DEUX TEXTES PEUVENT DONC SE PARAPHRASER SANS QU'AUCUNE DE LEURS PHRASES NE SOIENT DES PARAPHRASES. (2) et (3) utilisent sensiblement le même lexique, la même richesse (ou pauvreté) dans

---

[3] Chronique cinématographique de *La Presse* de Montréal du samedi 30 juillet 1988.

[4] La confusion qui règne dans cette histoire entre le sexe masculin et féminin a semble-t-il fait perdre au journaliste l'usage de sa grammaire quant à ce qui a trait au genre grammatical.

[5] Le lecteur attentif aura peut-être remarqué que notre variante du texte implique une lecture où Tina est la sœur de Pablo. Il serait toujours possible, dans le texte initial, de comprendre qu'il/elle est en fait la sœur de Juan. Cette dernière interprétation, défendue par I. Mel'čuk nous semblait cependant très douteuse pour des raisons de cohérence textuelle; notre conscience professionnelle nous a poussé à aller voir ce film pour tirer cette affaire au clair: inutile de dire que c'est nous qui avions raison!

les structures syntaxiques, mais un MST, tel que conçu jusqu'à présent, ne peut pas rendre compte du lien paraphrastique qui les unit si l'on n'intègre pas ce modèle dans un appareillage descriptif plus vaste traitant aussi le niveau du texte. Nous pensons que ces deux textes sont en (presque) parfaite relation de paraphrase LANGAGIERE. Ce serait donc du ressort d'une théorie linguistique de rendre compte de cette paraphrase entre deux textes dans le cadre d'une mise en relation entre Sens et Texte.

Nous ajouterons que le problème de la paraphrase textuelle dépasse le seul niveau sémantique: certaines décisions syntaxiques, comme la pronominalisation ou le choix de déterminants, se prennent en fonction de considérations textuelles. Il faut donc disposer des outils nécessaires pour calculer et transmettre à un MST des informations sur la structure textuelle qui lui permettront d'effectuer toutes ces opérations s'appliquant à la phrase. Dans ce travail, nous n'avons cependant pas cherché à élargir la couverture linguistique des MST malgré notre conviction qu'elle est insuffisante. Ils restent des modèles de la phrase qui doivent être mis en contact avec une modélisation du texte[6]. Une approche possible serait de considérer que le Sens d'un texte peut être représenté à l'aide d'un vaste réseau sémantique dans lequel il conviendrait de "couper" pour obtenir les représentations sémantiques des phrases successives. Nous pensons qu'il est vain d'essayer de traiter ce problème de la structuration d'un texte à travers une extension des MST. Le Sens d'un texte, et non plus celui d'une phrase, doit être calculé à partir d'un niveau de connaissance et de représentation beaucoup plus proche du "vrai" sens auquel nous avons fait allusion ci-dessus. Le formalisme des réseaux sémantiques ne serait vraisemblablement d'aucune utilité pour effectuer une tâche de structuration textuelle. Il est cependant

---

[6] Nous prenons donc pour acquise, dans ce travail, la pertinence d'une description linguistique Sens-Texte limitée au champ de la phrase. Notons, à titre indicatif, que certains arguments, basés sur une approche psychologique de la communication, ont été avancés pour considérer la phrase comme l'unité de planification lors de la production de la parole. On pourra consulter, à ce propos, Garrett 1975. De tels arguments ont bien sûr trouvé des détracteurs (cf. McNeill 1979:I.II); il ne nous appartient pas d'alimenter ici un tel débat.

possible d'envisager la construction de modèles s'appliquant à la dimension du texte qui respecteraient la philosophie générale de la TST: niveaux multiples de représentation, orientation des concepts vers le Texte, etc. (voir Iordanskaja 1989)[7].

2 QUELQUES PARTICULARITES FORMELLES DES RESEAUX SEMANTIQUES

2.1 Généralités

Une représentation sémantique Sens-Texte est constituée de trois structures:

- la structure sémantique, qui contient des Sens (de lexèmes, de phrasèmes, de grammèmes, etc.) connectés par des relations actancielles (notées **1**, **2**, **3**, **4**, **5**, ou **6**);
- la structure communicative, qui spécifie les oppositions telles que thème/rhème, posé/présupposé, information en premier plan/information en arrière plan, etc.;
- la structure rhétorique, qui explicite les intentions stylistiques du locuteur (style humoristique, poétique, etc.).

Seules les deux premières structures seront pertinentes pour ce travail. Pour ce qui est de la deuxième structure considérée, nous tenons à souligner le fait suivant, qui prendra toute son importance dans la Partie **2**:

---

[7] Dans cette discussion, nous avons présupposé une nette distinction entre phrase et texte. La frontière entre ces deux notions n'est bien sûr pas toujours très claire. Il existe des cas intermédiaires comme, par exemple, les phrases à structures communicatives multiples correspondant à des conjonctions de phrases. De tels cas doivent pouvoir être traités dans le cadre d'un MST. Nous reviendrons sur ce problème dans la Partie **2**.

Dans ce travail, nous considérons la *structure communicative* presque exclusivement sous l'angle de l'opposition thème/rhème — c'est-à-dire, la structure thématique. Par défaut, lorsque nous parlerons de structure communicative, nous référerons donc en réalité à une structure thématique.

AINSI, DANS LE CADRE DE CE TRAVAIL, "COMMUNICATIF" SIGNIFIERA (PAR DEFAUT) "QUI RELEVE DE LA STRUCTURE THEMATIQUE". Nous reviendrons en détail dans la partie **2** (IV:2 et IV:4) sur la caractérisation de l'opposition thème/rhème au niveau sémantique. Il sera aussi question, dans cette même partie (V:4), de l'opposition POSE/PRÉSUPPOSE mais sans que cette opposition soit assimilée à la composante thématique de la structure communicative.

Avertissement:

L'étude que nous avons menée sur les notions de *thème*, *rhème* et *présupposition* a été motivée par la mise à jour du rôle central joué par la structure communicative dans le processus d'activation des MST. Nous avons voulu refléter ce fait en reportant la présentation de ces notions à la Partie **2** (consacrée à la transition RSém ⇒ RSyntP).

Pour l'instant, nous ignorerons la notion de présupposition et nous nous contenterons de dire que le *thème* de la RSém d'un Texte est le "bloc de Sens" présenté comme étant ce dont parle le Texte; le *rhème* de la RSém est un "bloc de Sens" présenté par le Texte comme ce qui est dit à propos du thème. Ces définitions sont bien sûr très approximatives mais on peut proposer un autre outil dont il est fréquemment fait usage (cf. Iordanskaja 1989) pour cerner les notions de thème et rhème: il s'agit de la technique du questionnement. Cette technique consiste à considérer le Texte comme une réponse APPROPRIEE à une question donnée. Dans ce cas, l'élément de réponse

comme tel, c'est-à-dire l'information apportée sur le point précis interrogé, pourra être appelé rhème. Si nous reprenons l'exemple (1b), nous voyons que plusieurs questions hypothétiques peuvent lui être associées SELON LE CONTEXTE D'ENONCIATION CONSIDERE. Comme nous le verrons en IV, cela tient au fait que le français ne reflète pas nécessairement à la surface — de façon non ambiguë, la structure communicative d'une RSém:

(1b) *Les plans de la fusée pour la Lune sont introuvables.*
← 'Qu'est-ce qui est arrivé aux plans de la fusée pour la Lune?'
← 'Qu'est-ce qui est introuvable?'
etc.

<u>Remarque</u>: Si nécessaire, les phrases utilisées comme exemples dans ce travail seront accompagnées d'une question hypothétique donnant une indication sur la structure communicative de leur RSém. Le lien entre cette question hypothétique et son énoncé-réponse est indiqué par le symbole "←".

Au niveau de leur structure sémantique proprement dite, les RSém sont des *graphes orientés connectés*, c'est-à-dire qu'ils se définissent formellement au moyen de deux ensembles:

- un ensemble non vide de nœuds;
- un ensemble d'arcs définis par une paire ordonnée de nœuds (l'ordre correspondant à la direction indiquée par les flèches sur les arcs), tel qu'il existe une connexion directe ou indirecte entre tous les nœuds.

Dans ce travail, nous référons cependant aux RSém en tant que *réseaux* dans la mesure où nous considérons la plupart du temps leur représentation graphique, qui est un réseau. De plus, la notion de graphe est trop générale et s'applique aussi aux RSyntP, qui, en tant qu'arbres, sont un cas particulier de graphe.

Une caractéristique essentielle de notre utilisation du formalisme de réseau pour les RSém est que nous faisons un usage systématique de la notion de *nœud dominant* (cf. Mel'čuk 1988c:21). Nous définirons précisément, dans le cours de l'exposé, cette notion de nœud dominant sur laquelle s'appuie toute notre conception de la transition sémantique. Suivant la convention adoptée dans certains travaux récents sur les MST (cf. Mel'čuk 1988c et Mel'čuk & Polguère, à paraître), nous soulignerons le nœud dominant sémantiquement un réseau. Cette usage va même être rendu systématique grâce à la règle suivante:

> Tout sous-réseau (**thème**, **rhème**, etc.) d'un réseau sémantique, QUI EST UNE RSEM, doit contenir la spécification d'un nœud dominant (souligné). Il en va de même pour tout réseau qui n'est pas lui-même subdivisé en sous-réseaux.

Nous allons maintenant faire deux observations sur certaines caractéristiques formelles des RSém, vues en tant que réseaux.

2.2 Représentation linéarisée des réseaux sémantiques

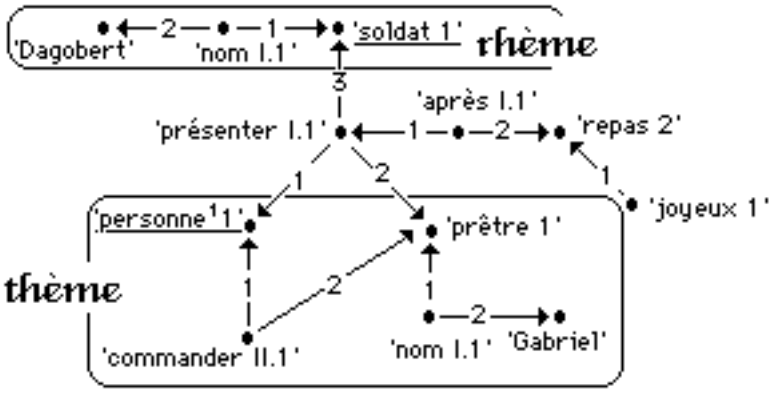
Comment représenter les réseaux sémantiques sous une forme linéaire? La question est importante car, pour effectuer des calculs sur les réseaux à l'aide d'un programme informatique, il est nécessaire de pouvoir leur associer une représentation linéaire. Par conséquent, dans toutes les recherches impliquant une implémentation de modèles (linguistiques ou non) fondés sur les réseaux sémantiques, il convient de définir une stratégie de représentation linéaire de ces figures "multidimensionnelles". Bien que nous ne nous intéressions pas ici directement à décrire comment se programme un MST, nous voudrions apporter quelques précisions sur ce point particulier. Les remarques qui suivent découlent bien sûr de notre propre expérience de la manipulation informatique des RSém. Nous ne parlerons cependant ici que de ce qui, dans le choix du type

de représentation linéaire, est directement dépendant de notre approche globale (si l'on veut, "linguistique procédurale" mais non nécessairement "informatique") de la transition RSém ⇒ RSyntP.

On trouve dans Sowa 1984 un équivalent linéaire pour la représentation des *graphes conceptuels* (sur lesquels nous reviendrons en 3). En l'adaptant légèrement, nous pouvons dire que la technique de linéarisation proposée par J. Sowa nous permettrait de donner la forme linéaire (5b) au réseau (5a) qui est la RSém de la phrase (4).

(4) *La personne qui commande au prêtre Gabriel a présenté ce dernier au soldat Dagobert après un joyeux repas.*

(5a)



<u>Remarque</u>: Afin de respecter la contrainte formulée précédemment, les nœuds dominant les sous-réseaux **thème** et **rhème** ont été soulignés dans (5a).

(5b) 'présenter'^{1} -
(1) → 'personne'^{10} -
(1) ← 'commander'^{9} -
(2) → 'prêtre'^{4} -
(1) ← 'nom'^{6} -
(2) → 'Gabriel'^{11}
(2) → 'prêtre'^{4}
(3) → 'soldat'^{2} -

(1) ← 'nom'[7] -
(2) → 'Dagobert'[3]
(1) ← 'après'[8] -
(2) → 'repas'[5] -
(1) ← 'joyeux'[12]

En (5b), les relations étiquetant les arcs sont écrites entre parenthèses et l'orientation des arcs est indiquée par une flèche, "→" ou "←", située à droite du nom de la relation. Afin d'assurer une certaine équivalence entre les deux formalismes, il est nécessaire d'attribuer un indice aux éléments sémantiques de la représentation linéaire (pour faciliter la lecture, nous avons supprimé dans (5b) les numéros spécifiant les Sens de lexème). Cela revient en quelque sorte à attribuer un numéro de référence[8] à chacun des nœuds du réseau; ces numéros sont ici attribués de façon arbitraire et n'ont d'autre valeur sémantique que la spécification univoque de l'occurrence d'un Sens. On peut noter que J. Sowa propose d'utiliser des variables pour spécifier de façon univoque un nœud du graphe: 'nom':x, 'nom':y, etc.; cela lui permet notamment d'utiliser, dans les graphes conceptuels, les quantificateurs logiques pour lier ces variables. Dans le cas de nos réseaux sémantiques, nous préférons considérer que ce sont les noms des Sens eux-mêmes qui sont des sortes de variables, variables qui doivent être "liées" au moyen d'une indexation[9].

Le principal problème que pose une représentation linéaire comme (5b) est qu'elle présuppose une sorte de hiérarchie entre les nœuds, qui n'existe pas dans la représentation graphique. Sowa souligne lui-même que ces formules sont construites en grappes à

---

[8] Le terme "référence" est choisi ici volontairement pour son ambiguïté. On verra plus loin, dans le chapitre VI de la Partie **2**, qu'il est préservé lors de la transition au niveau syntaxique profond et sert justement à spécifier la co-référence entre nœuds de la RSyntP.

[9] Nous ne considérons pas dans ce travail le problème de l'expression linguistique, au niveau sémantique, des phénomènes de portée des quantificateurs logiques. Il est certain qu'une étude poussée du phénomène de la quantification, dans un cadre Sens-Texte, demanderait le développement d'autres outils formels pour compléter les RSém et les règles les manipulant.

partir d'un nœud qui en est la "tête". Il faut considérer que le choix de cette tête et la hiérarchie qui en découle n'ont pas de signification dans la représentation linéaire. Il est bien sûr possible d'adopter une telle convention et de n'accorder aucune pertinence à la notion de hiérarchie d'une représentation linéaire; il se trouve cependant que celle-ci pourrait interférer dangereusement, dans le cas de nos RSém, avec certains aspects cruciaux de la transition sémantique telle que nous allons proposer de la modéliser dans la suite de ce travail. Nous pensons donc qu'un formalisme beaucoup simple et "plat" doit être adopté lorsque l'on manipule des versions non graphiques des RSém: c'est-à-dire, UN SIMPLE **ENSEMBLE** DE NOEUDS SEMANTIQUES.

Dans les réseaux sémantiques de la TST, les relations entre nœuds correspondent simplement à des marques de relations prédicat-argument. Ainsi, LE SYSTEME DE RELATIONS du réseau (5c) est formellement équivalent à l'ensemble des relations prédicat-argument de (5a).

(5c) 'présenter'^{1} ('personne'^{10} , 'prêtre'^{4} , 'soldat'^{2})
'nom'^{6} ('prêtre'^{4} , 'Gabriel'^{11})
'nom'^{7} ('soldat'^{2} , 'Dagobert'^{3})
'après'^{8} ('présenter'^{1} , 'repas'^{5})
'commander'^{9} ('personne'^{10} , 'prêtre'^{4})
'joyeux'^{12} ('repas'^{5})

Si (5c) rend bien compte des RELATIONS existant entre les nœuds de (5a), cette collection de relations prédicat-argument ne constitue cependant pas une REPRESENTATION formellement équivalente au réseau. En effet, elle ne décrit que les foncteurs puisque ce sont les seuls éléments du réseau à posséder des dépendants. Les noms d'objets n'apparaissent qu'indirectement comme des arguments d'un ou plusieurs foncteurs. De plus, il est difficile de coder clairement la structure communicative du réseau dans une telle transcription. Nous préférons donc considérer la représentation (5d) ci-dessous comme équivalente au réseau (5a). Elle décrit les nœuds du réseau plutôt que de décrire les relations. Ainsi, les noms d'objets correspondent en quelque sorte à un cas particulier de foncteur d'arité zéro.

(5d) **thème:**
'personne'^{10}
'prêtre'^{4}
'nom'^{6} ('prêtre'^{4} , 'Gabriel'^{11})
'commander'^{9} ('personne'^{10} , 'prêtre'^{4})
'Gabriel'^{11}

rhème:
'soldat'^{2}
'Dagobert'^{3}
'nom'^{7} ('soldat'^{2} , 'Dagobert'^{3})

reste du réseau:
'présenter'^{1} ('personne'^{10} , 'prêtre'^{4} , 'soldat'^{2})
'repas'^{5}
'après'^{8} ('présenter'^{1} , 'repas'^{5})
'joyeux'^{12} ('repas'^{5})

On notera que nous avons dans (5d) autant de déclarations (à savoir, douze) qu'il y a de nœuds dans le réseau sémantique (5a). Il est aussi possible de décider, par convention, que le nœud dominant des sous-réseaux **thème** et **rhème** est ici mentionné en premier.

Remarque: (5d) permet d'éviter d'introduire une hiérarchie entre les nœuds sémantiques. Cette hiérarchie, bien que moins gênante que celle de (5b), est tout de même implicite dans (5c). En effet, dans ce dernier cas, la représentation linéaire n'introduit les objets sémantiques qu'indirectement, en introduisant avant tout les foncteurs. Une telle représentation est un ensemble de foncteurs et non un ensemble de nœuds sémantiques.

2.3 Acyclicité

Il faut relever une autre caractéristique importante des réseaux sémantiques: ceux-ci ne peuvent être cycliques. Par exemple, une chaîne de relations entre nœuds du type "W est argument de X qui est argument de Y qui est argument de Z qui est argument de W" est impossible. En fait, il est toujours possible de construire un réseau

sémantique avec cycle, mais la réalisation textuelle d'une telle RSém sera alors impossible. Considérons la figure I-1 ci-dessous:

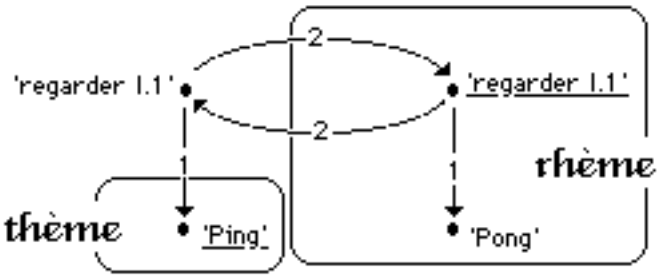


Figure I-1
RSém avec cycle

Si l'on voulait associer une phrase française à ce réseau, il faudrait dire quelque chose comme:

(6) *Ping regarde Pong regarder Ping regarder Pong regarder Ping ...*
← 'Qu'est-ce que regarde Ping?'

qui est une phrase sans fin. Il est impossible de stopper à un moment la récursion. Ainsi, une phrase comme:

(7) *Ping regarde Pong regarder Ping.*
← 'Qu'est-ce que regarde Ping?'

n'a pas pour RSém la figure I-1 mais plutôt:

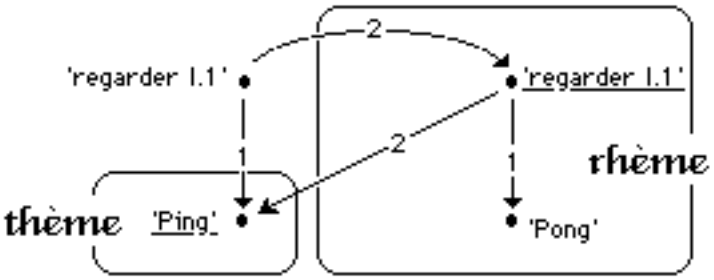


Figure I-2
RSém de *"Ping regarde Pong regarder Ping."*

On voit donc que les RSém sont très contraintes au niveau de leur syntaxe (au sens de syntaxe d'un langage formel); elles peuvent cependant intégrer les configurations sémantiques les plus complexes. Il n'y a pas de limite a priori au nombre de nœuds qui peuvent être les ancêtres d'un nœud sémantique donné.

Cette double caractérisation des réseaux sémantiques de la TST, simplicité de la syntaxe et très grande complexité potentielle des configurations à traiter, est d'ailleurs une des raisons qui nous a poussé à développer une méthode de parcours des RSém qui soit la plus algorithmique possible (voir Introduction).

3 AUTRES APPROCHES UTILISANT LES RESEAUX SEMANTIQUES

Pour terminer cette description des RSém, nous allons brièvement passer en revue certains travaux liés au traitement de la langue naturelle qui utilisent des représentations pouvant être caractérisées comme des réseaux sémantiques. On peut remarquer que les principales théories linguistiques formelles n'ont pas recours aux réseaux comme formalisme de représentation. Cela tient surtout au fait que ces théories reposent pour la plupart sur une approche syntaxique plutôt que sémantique (les descriptions syntaxiques étant généralement fondées sur la notion d'arbre). Dans le cas d'approches plus sémantiques (grammaires de Montague et modèles dérivés notamment), celles-ci ont recours à la logique formelle et utilisent des représentations linéaires avec quantificateurs. On trouve néanmoins de nombreux travaux en linguistique informatique, ou plus généralement en IA ou en sciences cognitives, qui mettent en jeu les réseaux comme structure de représentation des connaissances[10], structure sur laquelle diverses opérations d'inférence ou de transformation peuvent être effectuées; en voici quatre exemples:

---

[10] "Knowledge representation"; ces connaissances peuvent être très générales ou porter sur des domaines techniques spécifiques et peuvent même être des connaissances linguistiques.

1) McNeill 1979 dont l'ambition est de trouver comment caractériser une correspondance du type Pensée ⇔ Parole en s'appuyant sur des représentations du contenu des énoncés sous forme de graphes conceptuels ("conceptual graphs");
2) Sowa 1984 qui utilise les graphes conceptuels comme outils de représentation des connaissances et de modélisation du raisonnement;
3) Schank 1975:Chapitre 3, qui propose un formalisme de dépendance conceptuelle ("conceptual dependency") pour représenter le contenu de phrases analysées à l'aide d'un nombre restreint de primitives conceptuelles et un nombre potentiellement illimité d'objets dénotants ("picture-producers") liés entre eux par des relations de dépendance;
4) Frederiksen 1975 qui, dans une approche de psychologie cognitive, utilise un formalisme de réseau pour modéliser la compréhension du discours au niveau du texte; ces réseaux possèdent une double structure: la structure "sémantique" qui encode les connexions entre des concepts et la structure "logique" qui encode les connexions entre propositions.

Tous ces travaux mériteraient, à divers titres, une présentation plus poussée. Peut-être même une étude systématique de la notion de réseau et de son utilisation comme formalisme de représentation dans le traitement de la langue naturelle serait-elle une recherche intéressante en soi. Comme nous ne pouvons pas nous engager ici sur ce terrain, nous nous contenterons de présenter plus en détail le formalisme des graphes conceptuels de Sowa 1984. Ce choix peut s'expliquer de deux façons:

1) Le formalisme des graphes conceptuels développé par Sowa est actuellement très en vogue dans les différentes recherches en IA et en sciences cognitives; cela, à cause de sa généralité et de la relative rigueur avec laquelle il a été défini.

2) Nous expliciterons, en V:5 (Partie **2**), certaines notions proposées par Sowa qui présentent des ressemblances frappantes avec des notions que nous avons nous-mêmes dû définir pour organiser procéduralement la transition RSém › RSyntP. Il sera donc utile, à cette étape de notre exposé, de pouvoir référer directement à des graphes conceptuels formalisés.

On trouvera dans le troisième chapitre de Sowa 1984 une description formelle détaillée des graphes conceptuels. Nous ne faisons ici que récapituler les caractéristiques essentielles de ce formalisme de représentation des connaissances.

Les graphes conceptuels sont, pour Sowa, un moyen de représenter la façon dont un cerveau humain (ou une machine qui simule son fonctionnement) interprète le Monde[11] à partir (i) de sa perception de celui-ci, (ii) d'un stock déjà existant de perceptions antérieures, et (iii) de sa capacité à effectuer des raisonnements. Un graphe conceptuel est donc le résultat d'une certaine interprétation du Monde. Les éléments du Monde perçus, que Sowa appelle "percepts", sont interprétés en tant que concepts et mis en relation les uns avec les autres. En conséquence, un graphe conceptuel comprend trois types distincts de composantes:

1) des concepts, représentés par des nœuds rectangulaires étiquetés;
2) des relations entre les concepts, représentées par des nœuds circulaires[12] étiquetés;
3) des arcs orientés, qui unissent chacun un concept (un nœud rectangulaire) et une relation (un nœud circulaire).

---

[11] Nous prenons comme convention d'écrire *Monde* avec une majuscule initiale pour référer à ce plan de la réalité sur lequel s'effectuent les opérations d'interprétation dans les divers processus cognitifs.

[12] Nous ne pensons pas dénaturer Sowa 1984 outre mesure en utilisant ici l'ovale plutôt que le cercle!

Comme illustration du formalisme, on peut examiner la figure I-3 qui est le graphe conceptuel que Sowa 1984:77 propose d'associer au syntagme *"un chat assis sur un tapis"*:

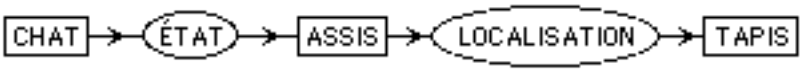
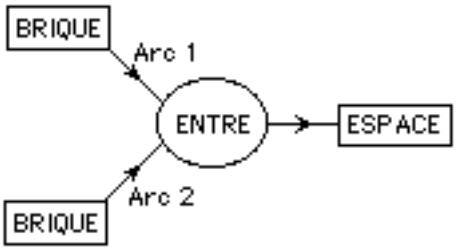


Figure I-3
Exemple de graphe conceptuel selon Sowa 1984

Nous donnons cet exemple tel quel sans vraiment chercher à "justifier" cette représentation. Pourquoi tel élément de la représentation est-il une relation conceptuelle et tel autre un concept proprement dit? Pourquoi les arcs sont-ils orientés de cette façon? Ces questions et bien d'autres méritent une réponse qui peut être trouvée, en partie seulement à notre avis, dans Sowa 1984. Les critères qui permettent à Sowa de structurer de cette façon sa représentation de la figure I-3 ne nous semblent pas vraiment très clairs. Il soutient notamment (page 74) que le lien entre la structure d'un graphe conceptuel et le Texte auquel il peut être associé n'est pas direct, mais le reste de l'exposé pose souvent problème sur ce point particulier.

A la différence des réseaux sémantiques Sens-Texte, les graphes conceptuels mettent en jeu des relations qui sont conçues comme des nœuds à part entière. Il y a donc plusieurs sortes de relations possibles selon le nombre d'arcs qui s'y rattachent. On peut ainsi trouver des relations "triadiques" comme dans:



que Sowa propose comme une représentation de *"un espace entre une brique et une brique"*. Il faut noter que, dans de tels cas, le formalisme des graphes conceptuels se trouve alourdi d'une donnée

supplémentaire puisque qu'il faut (un peu comme pour les RSém) particulariser les arcs par un étiquetage ("Arc 1" et "Arc 2").

Pour Sowa, les graphes conceptuels ne sont pas des réseaux sémantiques. Il réserve ce dernier terme pour caractériser le "système de relations" qui unit un concept avec le reste de l'appareillage sensoriel et cognitif considéré: c'est-à-dire les autres concepts, les percepts et tous les participants des divers processus cognitifs. En conséquence, le "sens d'un concept" est tout entier contenu dans ce système de relations représenté par un réseau sémantique. On voit ici le gouffre qui sépare l'usage fait par Sowa et par la TST des termes, apparemment identiques, de "sens" et "réseau sémantique". Il est donc fondamental de toujours garder à l'esprit le caractère technique de cette terminologie pour éviter toute confusion. DANS CE CAS PRÉCIS ET CONTRAIREMENT AUX APPARENCES, NOUS NE PARLONS PAS DES MEMES CHOSES ET NOUS LE FAISONS AVEC DES TERMES TECHNIQUES DIFFERENTS.

Le formalisme des graphes conceptuels doit posséder au moins la puissance de représentation du formalisme de la logique du premier ordre. En tant que représentation des connaissances, il doit aussi être équipé pour permettre d'effectuer toutes sortes de connexions avec les données du Monde (interprétation dénotationnelle) et les systèmes de raisonnement. Le formalisme de base doit donc être sophistiqué pour pouvoir s'intégrer à des systèmes complexes fonctionnant sur plusieurs plans simultanément. Mentionnons ici la fameuse relation IS-A (EST-UN). Elle spécifie des relations hiérarchiques entre concepts qui se superposent aux autres relations introduites dans les graphes conceptuels (chat EST-UN animal EST-UN être vivant, etc.). Nous laisserons de côté le problème du lien entre graphes conceptuels et dénotation. Notre propos était simplement d'introduire les principales caractéristiques du formalisme utilisé et de souligner en quoi les préoccupations de Sowa 1984 peuvent être différentes des nôtres.

Nous remarquerons en conclusion que, par rapport aux types de réseaux qui ont été évoqués ci-dessus, les RSém de la TST sont extrêmement pauvres. Notamment, il est important de considérer que

chaque nœud d'une RSém est un prédicat dont l'arité est fixe. Un Sens comme 'présenter**I.1**' (utilisé dans (5a) ci-dessus) est un prédicat à trois actants sémantiques, qui n'apparaîtra jamais dans un réseau sous une autre forme. De plus, le faible nombre de relations étiquetant les arcs (six au maximum) et leur informativité très réduite (spécification d'une relation actancielle **1**, **2**, **3**, **4**, **5** ou **6**) fait finalement de la RSém un formalisme très "dépouillé", ce qui a deux conséquences immédiates:

1) Les RSém de la TST contiennent toute l'information nécessaire pour faire fonctionner de façon optimale les règles sémantiques données par les MST;

2) Les RSém de la TST ne contiennent aucune information destinée à faciliter l'exécution d'une tâche autre que le paraphrasage linguistique sémantique; elles ne permettent notamment pas de raisonner sur le contenu d'un Texte afin d'inférer quelque chose sur l'état du Monde avec lequel il est en rapport (cf. Iordanskaja & Polguère 1988).

4 LES ARBRES SYNTAXIQUES PROFONDS

La composante syntaxique (profonde et de surface) d'un MST est fondée sur une *syntaxe de dépendance*. Avant de discuter des RSyntP, nous allons faire quelques remarques à propos de ce type de syntaxe.

4.1 Remarques sur le recours à une syntaxe de dépendance

Tesnière 1959 définit la syntaxe comme l'étude des mots et des connexions entre eux à l'intérieur d'une phrase, ces connexions étant des dépendances entre mots "régissant" et mots "régis". En syntaxe, un mot peut régir plusieurs subordonnés mais n'est régi que par au

plus un dominant; de plus, toute phrase possède un et un seul nœud non régi. On voit donc que la syntaxe selon Tesnière implique qu'une représentation syntaxique d'une phrase sera un arbre exprimant les dépendances syntaxiques entre les mots de cette phrase. D'une certaine façon, il existe un consensus chez la plupart des grammairiens pour reconnaître l'adéquation du formalisme d'arbre comme outil de représentation de la structure syntaxique d'une phrase. Cependant, les arbres syntaxiques sont-ils toujours des arbres de dépendance? On peut répondre par l'affirmative à cette question uniquement si l'on prend le terme de *dépendance* dans son sens le plus classique: quelque chose comme "lien entre des entités introduisant une hiérarchie entre ces entités". Cependant, deux grandes tendances existent à l'heure actuelle chez les grammairiens qui répartissent ceux-ci en deux groupes[13]:

1) les partisans d'une syntaxe de DÉPENDANCE ENTRE MOTS, par simplification "syntaxe de dépendance" — les dépendances étant en général explicitement nommées par une étiquette sur les arcs des arbres syntaxiques;
2) les partisans d'une syntaxe de HIERARCHIE DE CONSTITUANTS, ou "syntaxe de constituants" - syntaxe qui implique la définition et la manipulation de nœuds syntaxiques "abstraits", correspondant à des syntagmes ou parties de syntagmes, et où seules les feuilles des arbres syntaxiques sont étiquetées par les mots de la phrase.

Dans le second cas (syntaxe de constituants), la notion de dépendance est effacée au maximum pour laisser la place à celle d'INCLUSION de constituants. Ce qui compte, dans une telle approche, c'est de rendre compte de l'enchâssement mutuel des constituants,

---

[13] Nous simplifions ici à l'extrême le tableau. Des approches "mixtes" ont aussi été proposées; cf. Gladkij 1985 qui propose de recourir à des descriptions formelles utilisant à la fois la notion de syntaxe de dépendance et celle de groupe syntaxique.

enchâssement qui peut être constaté directement en surface à partir de l'ordre linéaire d'apparition des éléments de la phrase[14].

On trouve dans Mel'čuk 1988a:Introduction une argumentation en faveur des grammaires de dépendance. Elle met en évidence le fait que la suprématie actuelle des grammaires de constituants est essentiellement due à quatre raisons plus "politiques" que scientifiques:

1) emphase sur l'étude de l'anglais, langue qui utilise les constituants (les syntagmes) pour exprimer les dépendances syntaxiques sans faire appel, pour cela, à la morphologie (pas de marques casuelles, etc.);
2) fascination pour une hyper-formalisation aux dépens du recours aux critères introspectifs d'exploration linguistique;
3) description centrée sur l'utilisation d'un outil formel mathématique: le format de règle Backus-Naur (abréviation anglaise, *BNF*);
4) emphase mise sur l'analyse plutôt que sur la synthèse, avec un relatif désintéressement (au moins initialement) pour la sémantique.

Nous ajouterons que les travaux en *informatique linguistique* ont servi, en quelque sorte, de "caisse de résonance" à ces tendances jouant en faveur d'une éclipse des grammaires de dépendance. Les linguistes-informaticiens ont avant tout essayé de mettre en pratique, dans des applications orientées vers l'analyse, les descriptions grammaticales des grammaires génératives. La principale conséquence de cela est que la seule linguistique qui a été rendue accessible aux non-linguistes concernés, essentiellement les informaticiens, a été celle fondée sur une syntaxe de constituants. Il faut reconnaître que l'informatique exerce, dans les milieux de recherche, une certaine hégémonie sur les autres disciplines (au niveau des moyens de

---

[14] Il convient de noter que les grammaires de dépendance sont généralement développées par les structuralistes des écoles européennes alors que les grammaires de constituants sont maintenant typiques des approches générativistes.

recherche, de publication, etc.). On comprend alors fort bien le "tort" que l'informatique linguistique a pu faire à la recherche sur les grammaires de dépendance.

Pour ce qui est des outils informatiques comme tels, l'apparition de langages de haut niveau comme Prolog, dont les bases du formalisme permettent très facilement d'encoder les BNF, a pu faire croire pendant un certain temps que les grammaires de constituants étaient la solution de l'avenir en termes d'implémentation. W. Clocksin et C. Mellish montrent ainsi dans leur livre d'introduction à Prolog, Clocksin & Mellish 1981, à quel point il est facile d'implémenter dans ce langage une micro grammaire de constituants. L'argument de la simplicité ne tient bien sûr pas. Il suffit, pour s'en convaincre, de considérer autre chose que de simples "grammaires-jouets" et d'examiner une version Prolog d'une grammaire un peu plus puissante comme Chat-80 de Pereira 1983. Cette dernière, outre le fait qu'elle ne fonctionne que pour l'analyse, commence à ressembler beaucoup plus à du code informatique qu'à une description linguistique.

Pour conclure ces quelques considérations sur la notion de syntaxe de dépendance, notons que la dominance des approches fondées sur la syntaxe de constituants tend à devenir chose du passé, même dans le domaine de l'informatique linguistique où l'on trouve de plus en plus de recherches en rupture avec le courant générativiste (cf., par exemple, les travaux présentés dans Schubert 1987).

4.2 Formalisme des arbres syntaxiques profonds

Dans l'approche Sens-Texte, les RSyntP des phrases sont habituellement décrites comme des arbres de dépendance DONT LE SOMMET EST VERBAL et dont les nœuds sont étiquetés par des unités lexicales profondes de la langue et les arcs par une des neuf relations suivantes: **I**, **II**, **III**, **IV**, **V**, **VI**, **APPEND** (appenditive), **ATTR** (attributive) et **COORD** (coordinative).

Une unité lexicale est dite "profonde" si elle N'APPARTIENT PAS à l'une des classes suivantes:

- lexème vide — ETRE**V** (auxiliaire), À**I** (cf. *"penser à qqch."*), etc.;
- lexème pronominal — pronom, adjectif possessif, etc.;
- lexème qui résulte de l'application d'une *fonction lexicale* — DONNER en tant que $Oper_1$ de RÉPONSE**1**, etc.

La dernière classe de lexèmes mentionnée ci-dessus étant absente des RSyntP, ceci implique que certains nœuds de la RSyntP, au lieu d'être étiquetés par des lexèmes, pourront être étiquetés par une formule composée d'une fonction lexicale appliquée à un argument donné — par exemple, "$Oper_1$(RÉPONSE**1**)".

Les phrasèmes (⌜PASSER À TABAC⌝, etc.) peuvent, comme les lexèmes profonds, étiqueter les nœuds des RSyntP. Cependant, afin de simplifier notre terminologie, nous appellerons les unités lexicales profondes *lexèmes profonds* ou même, par défaut, *lexèmes*.

Les RSyntP comprennent, outre la structure syntaxique proprement dite, une structure communicative, une structure anaphorique et une structure prosodique. Cette dernière structure ne sera pas prise en compte dans ce travail. Nous avons concentré notre attention sur les trois autres structures.

Il nous semble important de mettre en évidence le fait que la structure anaphorique empêche d'attribuer à la RSyntP un caractère purement arborescent. En effet, la puissance formelle de la RSyntP vient du fait qu'elle est une représentation hybride, à la fois arbre de dépendance (structure syntaxique) et réseau (structure anaphorique). Les liens de co-référence (doubles flèches pointillées dans les figures) servent en fait à préserver, à travers la hiérarchie syntaxique, une information de type sémantique.

Une fois observée cette relative proximité formelle entre RSém et RSyntP, est-il possible d'effectuer d'autres rapprochements entre

les deux formalismes? Pour répondre à cette question, il convient de comparer les deux constituants essentiels de chacune des représentations: les nœuds et les relations entre nœuds (arcs).

a) Nœuds sémantiques et nœuds syntaxiques profonds:

Selon la TST, les nœuds sémantiques sont étiquetés par des Sens alors que les nœuds syntaxiques profonds le sont par les lexèmes profonds. En réalité, pour bien comprendre cette distinction, il faut revenir à la définition du terme *Sens* dans la TST (cf. 1.1 ci-dessus). Le Sens est l'invariant des paraphrases langagières et, donc, le Sens d'un lexème profond est en fait représenté par un nœud portant une étiquette sémantique qui renvoie directement à ce lexème AINSI QU'A TOUTES LES DECOMPOSITIONS SEMANTIQUES PLUS OU MOINS PROFONDES DE CE SENS. On voit donc que le nœud syntaxique profond, par opposition, représente avant tout un choix qui a été effectué à partir du niveau sémantique; il existe cependant un lien univoque entre le Sens d'un lexème (qui étiquette un nœud sémantique) et ce lexème lui-même (qui étiquette un nœud syntaxique profond).

b) Relations sémantiques et relations syntaxiques profondes:

Observons tout d'abord que les relations sémantiques, de même que les relations syntaxiques profondes, sont universelles en ce sens qu'elles ne varient pas d'une langue à l'autre. De plus, aux six relations sémantiques **1**, **2**, **3**, **4**, **5**, et **6**, qui spécifient des actants sémantiques, correspondent six relations syntaxiques profondes "de base" **I**, **II**, **III**, **IV**, **V** et **VI**, qui spécifient des actants syntaxiques profonds. Le principal écart est constitué par l'introduction des relations **APPEND**, **ATTR** et **COORD** qui ne semblent pas être en corrélation directe avec des relations du niveau sémantique. Nous démontrerons plus loin que ces trois dernières relations ont un statut très particulier puisqu'elles correspondent à un ajout d'information par rapport aux relations actancielles. Cette information est purement

syntaxique mais elle est le reflet direct d'une certaine lecture qui est faite du réseau sémantique. En ce sens, elle est elle aussi la marque d'un choix.

Nous voudrions étendre cette caractérisation de la distinction RSém/RSyntP, fondée sur la notion de choix effectué, à l'ensemble des composantes de la structure syntaxique profonde. Ainsi, la construction d'un arbre syntaxique profond passe par le choix de lexèmes donnés, le choix d'un sommet (racine) donné et le choix de relations appropriées pour assurer une hiérarchisation syntaxique correcte en fonction des caractéristiques syntaxiques des différents lexèmes sélectionnés (essentiellement leur classe syntaxique). Tous ces choix sont opérés à partir de l'information présente dans la RSém et dans le lexique. La principale caractéristique d'une RSyntP par rapport à la RSém est donc qu'elle représente une certaine lecture de cette dernière; MODELISER LA TRANSITION ENTRE LES NIVEAUX SEMANTIQUE ET SYNTAXIQUE PROFOND REVIENT EN FAIT A MODELISER CETTE LECTURE DU RESEAU SEMANTIQUE.

Un dernier mot pour souligner que, si les relations syntaxiques profondes sont universelles, les règles qui vont les introduire à partir de la RSém ne le sont pas puisqu'elles se basent sur les caractéristiques syntaxiques des lexèmes introduits. Ces caractéristiques sont, elles, entièrement liées au lexique et à la grammaire d'une langue donnée. (Le comportement des classes syntaxiques est en effet propre à chaque langue de même que les écarts entre l'orientation des dépendances sémantiques et syntaxiques; cf. Mel'čuk 1981b.)

5 JUSTIFICATION THEORIQUE DES RSEM ET DES RSYNTP

La justification des niveaux de représentation dans un modèle stratificationnel du type Sens-Texte passe par une description des opérations spécifiques effectuées à chacun de ces niveaux. On sait,

en effet, que plus un système fonctionnel possède de niveaux, plus il est complexe dans la mesure où il manipule des formalismes de règles et de représentations différents; cependant, chacun de ces formalismes est plus simple à manipuler puisqu'il ne s'applique qu'à un petit fragment du problème à résoudre.

La TST prend en compte deux types de paraphrasage au niveau sémantique. Le premier est essentiellement basé sur la possibilité de substituer, DANS UNE RSEM, un Sens de lexème donné à un sous-réseau sémantique fait de plusieurs Sens de lexèmes interconnectés qui en constituent la définition (l'opération inverse de décomposition sémantique relevant elle aussi bien sûr du paraphrasage). Le second paraphrasage est peut-être plus riche puisqu'il englobe la possibilité d'associer plusieurs configurations syntaxiques profondes, c'est-à-dire plusieurs hiérarchies syntaxiques de lexèmes, à un même réseau sémantique, et, aussi, la possibilité d'utiliser les *fonctions lexicales* pour modifier ces hiérarchies à partir d'équations sur le Sens APPLIQUEES AU NIVEAU SYNTAXIQUE PROFOND[15]. Cette stratification est le reflet d'une certaine taxinomie des phénomènes linguistiques liés à la capacité de paraphrasage des langues naturelles. Nous ne pensons pas que cette taxinomie soit trop riche ni, d'ailleurs, qu'elle soit la seule envisageable. Cependant, notre expérience de l'implémentation de la transition sémantique dans le cadre d'un MST nous a montré que son principal mérite est de distinguer les phénomènes qui présentent une certaine autonomie. La dichotomie RSém/RSyntP ne se justifie donc pas par un souci de mieux modéliser ce qui effectivement se produit dans la tête du locuteur, mais elle va dans le sens de la modularisation des connaissances dans un système de SIMULATION d'une activité humaine: ici, la parole. En ce sens, un MST est parfaitement adapté aux approches d'IA qui veulent simuler au mieux un comportement en opposition avec des approches plus

---

[15] Cette ébauche de classification des différentes paraphrases sémantiques ne recoupe pas exactement celle de Mel'čuk 1988d. Cela tient essentiellement à la distinction nette que nous proposons d'effectuer entre règles de paraphrase sémantique et règles de correspondance sémantique. Un des objectifs de cette Partie **1** est justement de clarifier cette distinction.

ambitieuses ("perverses") qui cherchent à le reproduire tel qu'il existe naturellement[16].

Dans le cas d'une tentative de simulation de l'activité de parole, nous considérons donc normal de choisir de toujours séparer dans des niveaux de traitement différents les phénomènes de nature différente. Les résultats présentés dans ce travail ainsi que dans les autres recherches fondées sur la TST confirment à notre avis l'autonomie relative des niveaux postulés et justifient la dichotomie RSém/RSyntP[17].

Le recours à la distinction entre niveau de représentation sémantique et niveau syntaxique profond semble encore plus justifié si l'on considère l'exploitation du modèle linguistique dans un contexte d'application en génération de texte. En effet, dans un tel cadre, on construit des systèmes qui doivent faire cohabiter et interagir des niveaux de connaissance et de représentation linguistiques et non-linguistiques ou *conceptuels*.

Si nous postulons un niveau de représentation non-linguistique des phrases sous forme de réseaux conceptuels, il est possible d'affirmer que, dans de nombreuses approches en génération de texte, la dichotomie linguistique/non-linguistique est réalisée de la façon suivante:

---

[16] De telles approches n'ont rien de condamnable mais, quel que soit le domaine qu'elles abordent, elles relèvent avant tout d'une problématique psychologique.

[17] Nous sommes en revanche beaucoup moins convaincu que le niveau sémantique puisse être valablement traité sans aborder certains phénomènes habituellement considérés par la TST comme non-linguistiques (c.-à-d. conceptuels ou logiques). Nous reviendrons sur cette insuffisance de l'information contenue dans les réseaux sémantiques, mais sans réellement proposer de solution à ce problème qui dépasse les ambitions de cette thèse.

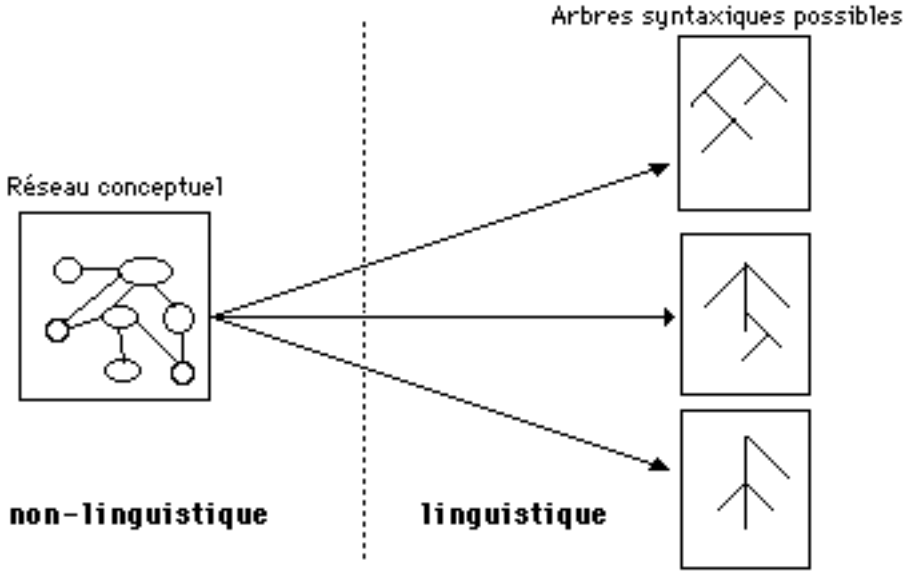


Figure I-4
Dichotomie linguistique/non-linguistique
dans les systèmes typiques de génération

Le type de transition qu'illustre la figure I-4 est trop direct et masque un niveau de traitement purement linguistique où sont effectués certains choix grammaticaux (choix entre les différents types de hiérarchies syntaxiques possibles) et lexicaux (choix des lexèmes profonds qui seront introduits dans la phrase).

Le recours à un modèle de type Sens-Texte en génération de texte permet d'obtenir les transitions suivantes:

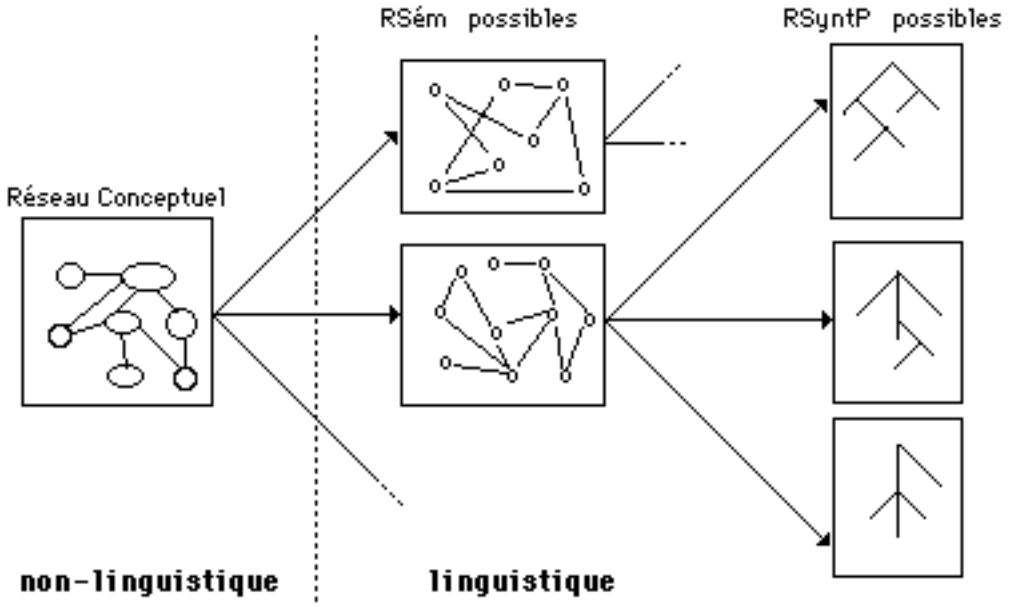


Figure I-5
Dichotomie linguistique/non-linguistique
permise par le recours à un MST

Remarque: Les différentes RSém possibles apparaissant dans la figure I-5 ne renvoient pas nécessairement à des paraphrases mais correspondent au choix qui peut être effectué pour exprimer selon différentes segmentations en phrases le contenu informatif d'un réseau conceptuel. Il n'est pas du tout certain qu'il faille à tout prix disposer d'une représentation conceptuelle pour chaque phrase dans un processus de génération. La question de savoir où et quand doit s'effectuer la segmentation reste à notre avis ouverte. (Pour plus d'informations sur la façon dont peut être planifié un texte en vue de la construction de représentations sémantiques, voir Carcagno & Iordanskaja 1989 ainsi que Iordanskaja 1989.)

La structuration des transitions dans la figure I-5 ne complexifie pas le processus de génération. Les systèmes de génération utilisent de multiples niveaux de représentation afin de mettre en jeu des règles traitant explicitement certains aspects particuliers du

processus en cause. L'approche Sens-Texte permet d'avoir des règles de transition relativement simples et spécialisées. La figure I-5, par exemple, montre qu'un modèle de génération utilisant les MST établit une distinction très forte entre la transition conceptuelle (du réseau conceptuel vers les réseaux sémantiques) et les transitions linguistiques (du réseau sémantique vers les arbres syntaxiques). Dans l'approche symbolisée par la figure I-4, au contraire, les deux problèmes sont traités simultanément, ce qui implique des types très complexes de connaissances et de règles. Au niveau terminologique, on notera que pour les chercheurs travaillant selon cette dernière approche (qui s'oppose à celle que nous préconisons), les termes *sémantique* et *conceptuel* peuvent la plupart du temps être considérés comme des synonymes; cf. par exemple, Herman *et al.* 1988 et Jacobs 1987.

On pourrait penser que le point de vue que nous défendons ici est biaisé par le fait que nous avons tendance à trop considérer le problème de la génération de texte selon une perspective linguistique. Peut-être voulons-nous donner trop d'importance à notre propre discipline de recherche pour mieux justifier son intervention dans les recherches plus vastes comme celles menées en IA. Nous pensons cependant que la stratification que nous proposons, qui fait cohabiter un niveau de représentation sémantique linguistique et un niveau de représentation sémantique non linguistique (ou conceptuel), s'imposera d'elle-même avec le temps. On peut noter que les chercheurs, non linguistes, travaillant en TALN se dirigent de plus en plus vers le type de stratification que nous proposons. Ainsi, un récent article de Sowa (à paraître) propose justement une réflexion sur la dichotomie entre structuration sémantique lexicale et structuration d'un domaine conceptuel. Sans que la compatibilité entre les présentes propositions de Sowa et les nôtres soit complète, certaines parentés sont tout de même intéressantes à observer. À titre d'illustration et parce que ce texte est encore difficilement accessible, nous reproduisons tel quel le schéma proposé dans Sowa (à paraître:9) pour structurer le processus de traduction d'un texte d'une langue A vers une langue B.

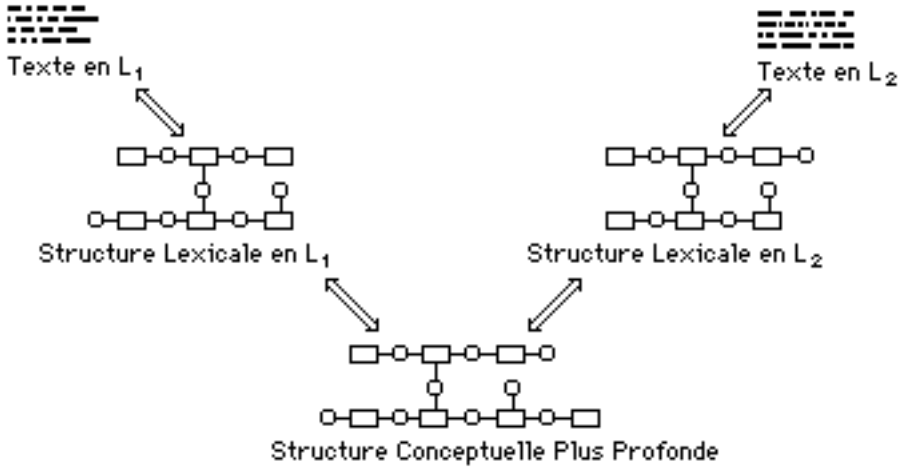


Figure I-6
Décomposition du processus de traduction selon Sowa (à paraître)

Les similarités existant entre cette figure et la figure I-5 ci-dessus (dont une première version avait été proposée dans Iordanskaja & Polguère 1988) apparaissent ici d'elles-mêmes. On notera toutefois que la figure I-6 ne mentionne pas le recours à des niveaux intermédiaires de représentation qui soient syntaxiques (hiérarchisés en arbre).

CHAPITRE II

Règles établissant les correspondances RSém ⇔ RSyntP

Nous venons d'examiner les représentations mises en jeu par la composante sémantique d'un MST, mais un tel modèle est constitué d'un ensemble de règles et non d'un ensemble de représentations! Il nous reste donc à accomplir la tâche la plus importante dans la description de la composante sémantique: la description des règles sémantiques. Nous répartissons celles-ci en quatre classes majeures qui seront étudiées successivement:

1) les *règles de bonne formation*;
2) les *règles de paraphrase sémantique*;
3) les *règles grammatico-sémantiques*;
4) les *règles de correspondance sémantique proprement dites*.

Afin de préparer le terrain à la présentation des aspects procéduraux traités dans la Partie **2**, nous introduirons aussi, sans toutefois décrire leur fonctionnement, les opérations majeures mettant en jeu ces règles dans le cadre de la transition sémantique; c'est-à-dire: la *lexicalisation*, la *grammaticalisation*, l'*arborisation* et la *hiérarchisation*.

1 LES REGLES DE BONNE FORMATION

La tâche de la composante sémantique est de prendre en charge la correspondance entre le niveau de représentation sémantique et le niveau de représentation syntaxique profond. Pour établir de façon convenable une telle correspondance, il est avant tout nécessaire de posséder un ensemble de règles définissant rigoureusement la syntaxe des RSém et des RSyntP. Ces règles, souvent implicites dans le modèle, contrôlent en fait la construction de toute représentation de même que l'écriture de toute autre règle sémantique. Il est intéressant de noter que dans le cas d'une application en génération de texte, là où un MST doit être alimenté en RSém par une composante extra-linguistique, cette composante a besoin d'un ensemble de règles de bonne formation des réseaux sémantiques spécifiant entièrement leur syntaxe. Les connaissances suivantes, notamment, doivent être entrées dans un système de construction automatique des RSém:

1) Ensemble de tous les nœuds reconnus par le modèle linguistique: c'est-à-dire l'ensemble de tous les Sens de la langue considérée, exprimables soit lexicalement soit grammaticalement;
2) Relations possibles entre les nœuds des RSém: elles se décrivent au moyen de règles de combinaison des Sens afin de former des significations plus complexes (par exemple, comment se construisent sémantiquement les différentes combinaisons de conjonctions, comment s'effectue la "prise d'argument" par un prédicat, etc.);
3) Contraintes sur la spécification des structures communicative et rhétorique: c'est-à-dire, de façon générale, comment un réseau se structure en sous-réseaux.

Toutes ces connaissances portent sur ce que nous appellerons, dans la Partie **2**, la *topographie des réseaux sémantiques*.

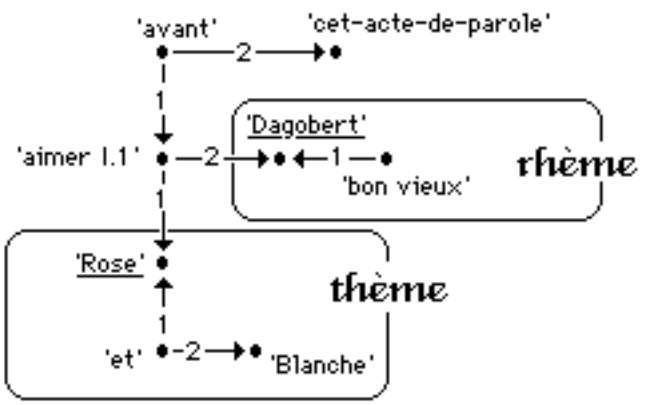
Il est justifié de considérer que les RSém sont écrites dans une sorte de langage formel graphique — les réseaux sémantiques. Comme tout langage formel, celui des réseaux sémantiques doit être clairement et exhaustivement défini au niveau de sa syntaxe et de sa sémantique. Nous ne prendrons ici qu'un exemple pour montrer qu'il n'existe pas d'écriture et d'interprétation "évidente" du langage formel des réseaux sémantiques.

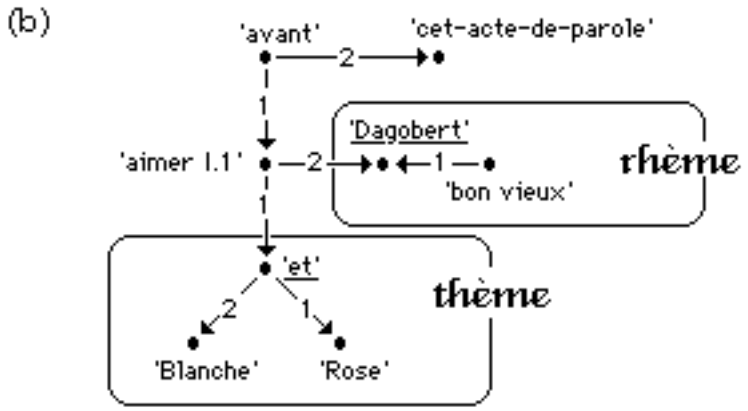
Soit (1), une phrase présentant un cas simple de conjonction:

(1) *Rose et Blanche aimaient le bon vieux Dagobert.*
← 'Rose et Blanche aimaient qui?'

et soit (a) et (b) deux tentatives de représentation sous forme de réseau du Sens de (1); nous faisons précéder (a) d'un astérisque car nous montrerons que seule (b) est une représentation correcte:

(a) *





Avant toute chose, remarquons que, dans ces RSém, le nœud 'et' est en fait une abréviation pour un prédicat sémantique très spécial que nous pourrions appeler 'ensemble^{et}'. Un tel prédicat admet potentiellement un nombre illimité d'arguments et correspond à une conjonction d'éléments (à l'opposé de 'ensemble^{ou}', qui correspond à une disjonction d'éléments). Il est utile de simplifier la représentation car nous nous intéressons ici uniquement au problème du rattachement au réseau global du sous-réseau correspondant à un ensemble de nœuds conjoints; notre propos n'est pas d'explorer le traitement de la conjonction comme tel[1].

La structure sémantique d'une RSém n'est pas un bloc homogène. On a vu que la structure communicative impose à la RSém une partition en sous-réseaux. On peut aussi considérer que, de façon générale, une RSém est en fait une mosaïque de micro-réseaux. Ainsi, en (a) et (b), le sous-réseau comprenant les nœuds 'et', 'Rose' et 'Blanche' forme, indépendamment de la partition imposée par la structure communicative, un "bloc de Sens". Tout le problème de la construction de la RSém de (1) réside dans le fait de savoir comment

---

[1] Notons que, dans GOSSIP, nous avions adopté une autre approche pour encoder sémantiquement la conjonction et la disjonction, en faisant appel à un nœud sémantique 'ensemble' possédant un argument et à une conjonction récursive. Ainsi, ce syntagme où le "champ" des différentes conjonctions est indiqué par les crochets en gras:

un chien, un chat et **[***une alouette, un cheval et trois ratons laveurs* **]**

se verrait associer, selon la méthode adoptée dans GOSSIP, une RSém comme:

'ensemble d'un chien et d'un chat et d'un ensemble d'une alouette et d'un cheval et de trois ratons laveurs'.

connecter ce bloc de Sens au reste du réseau. (a) implique que 'Rose' est le véritable premier actant sémantique de 'aimer**I.1**', premier actant qui est augmenté de 'et' ⊕ 'Blanche'[2]. (b) implique que le premier actant de 'aimer**I.1**' est un Sens plus abstrait, proche de la notion d'ensemble: le prédicat à deux arguments 'et' (mis pour 'ensemble^{et}').

Considérant les trois faits suivants:

1) 'et' ⊕ 'Blanche' ⊕ 'Rose' forme un bloc sémantique B,
2) 'et' (et non 'Rose') est le nœud dominant de B,
3) B est la caractérisation sémantique du premier actant de 'aimer**I.1**',

les règles de bonne formation des RSém vont imposer (b), et non (a), comme représentation valide du Sens de (1). La règle très générale utilisée ici est du type:

> Si γ est le nœud dominant d'un sous-réseau Γ et si ce sous-réseau doit être connecté à un autre réseau par une relation de n-ième actant sémantique d'un prédicat P, alors γ sera le n-ième actant de P.

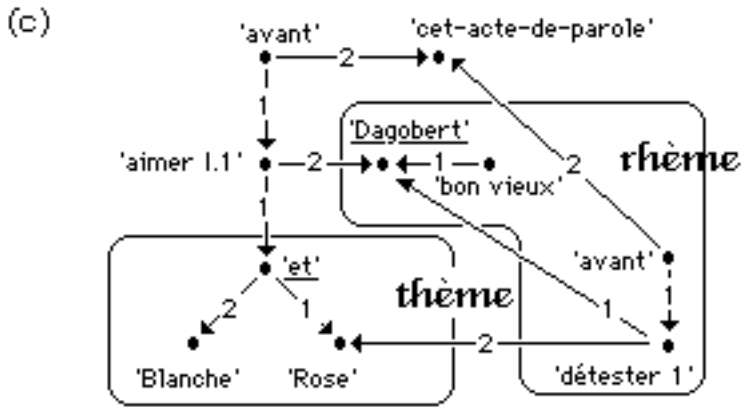
On peut aisément se convaincre de la pertinence du choix de (b) en essayant de construire, à partir de (a) et à partir de (b), une RSém d'une phrase comme (2):

(2) *Rose et Blanche aimaient le bon vieux Dagobert qui détestait Rose.*
← 'Rose et Blanche aimaient qui?'

Seule la stratégie utilisée en (b) va permettre de produire une figure correcte car elle offre la possibilité de distinguer formellement un argument référant à UN objet simple d'un argument correspondant à un ensemble d'objets, comme l'illustre (c):

---

[2] Rappelons que la méta-opération d'union linguistique ⊕ a été décrite dans l'Introduction.



Nous ne pousserons pas plus loin l'examen des règles de bonne formation car elles seront considérées à nouveau dans la Partie **2** (cf. IV:4) à la lumière des idées nouvelles que nous introduisons pour traiter de l'activation procédurale des MST. Dans le cas des RSém, nous avons montré que ces règles s'appuient sur d'autres composantes que les seules structures sémantique, communicative et rhétorique. La notion de bloc de Sens que nous avons utilisée demanderait une étude en soi que nous ne sommes pas à même de prendre en charge ici[3].

2 LES REGLES DE PARAPHRASE SEMANTIQUE

2.1 Formalisation des règles de paraphrase sémantique

On trouve dans la littérature sur la TST (cf., par exemple, Mel'čuk 1988b) les entrées lexicales présentées ainsi:

Définition du lexème ⇔ Lexème | Conditions

Les trois composantes d'une telle entrée peuvent être décrites de la façon suivante:

---

[3] On trouvera dans St-Germain 1988 des propositions originales pour ce qui a trait à une structuration "causale" des RSém; par exemple, J. St-Germain fait figurer les chaînes "causationnelles" (cause/cause-effet/effet) dans ses RSém au moyen de la spécification de sous-réseaux.

1) La partie définition a la forme d'un (sous-)réseau sémantique dans lequel un nœud est signalé comme dominant (souligné).
2) La partie lexème contient un (sous-)arbre syntaxique profond représentant le lexème et ses actants syntaxiques profonds abstraits (son régime).
3) Les conditions décrivent la combinatoire lexicale et syntaxique du lexème, de même que sa combinatoire morphologique.

Nous allons ici reformuler la présentation d'une entrée lexicale type afin, notamment, de préciser la notion de *règle de paraphrase sémantique*.

Une entrée lexicale (cf., Mel'čuk *et al.* 1984, 1988 et Mel'čuk & Polguère 1987) comprend une zone sémantique, une zone de combinatoire syntaxique et une zone de combinatoire lexicale. Chacune de ces zones est constituée d'une ou plusieurs règles dont la forme est conditionnée, bien sûr, par le type de données dont elles doivent rendre compte, mais aussi par leur contexte d'utilisation dans le cadre d'un MST[4]. Rappelons que, dans un MST, les entrées lexicales sont stockées dans un dictionnaire (lexique) formel appelé *Dictionnaire Explicatif et Combinatoire* (*DEC*).

La zone sémantique introduit la définition sémantique du lexème, c'est-à-dire la description du Sens de ce lexème en termes des Sens d'autres lexèmes de la langue considérés comme plus simples[5]. La définition va essentiellement permettre d'effectuer la paraphrase lexicale en établissant une équivalence entre l'apparition d'un Sens de lexème ou l'apparition de sa décomposition sémantique

---

[4] La frontière entre les notions de "donnée" et de "règle" est assez floue et dépend de la façon dont on considère l'information stockée dans le lexique. Ainsi, Magn(PEUR**2**) = BLEU**2** peut être aussi bien compris comme une donnée sur la combinatoire lexicale du lexème PEUR**2** que comme une règle permettant, par exemple, d'insérer de façon appropriée le lexème BLEU**2** dans une représentation syntaxique.

[5] Dans la présentation qui suit, nous donnons notre propre point de vue sur ce que doit être formellement la zone sémantique; voir la mise au point à la fin de cette section.

dans un réseau sémantique plus vaste: ce que nous appelons, dans un cadre de transition RSém ⇒ RSyntP, la *lexicalisation*. Notons que nous ne nous intéressons ici qu'à la *lexicalisation au sens propre*. La lexicalisation prise dans un sens plus général comprend, outre cette opération de sélection des Sens au niveau de la RSém, une opération (triviale) d'introduction des lexèmes correspondants lors de la transition au niveau syntaxique profond: opération qui peut être appelée *lexémisation* (cf. Mel'čuk & Polguère à paraître).

> La définition prend donc la forme d'une équation qui met en relation deux entités de même nature (c.-à-d. deux entités sémantiques). Du côté gauche du signe d'égalité de l'équation nous trouverons un sous-réseau constitué uniquement d'un nœud sémantique étiqueté du Sens du lexème et accompagné de ses actants sémantiques, le *definiendum*, du côté droit, un autre sous-réseau correspondant à la décomposition sémantique du lexème, le *definiens*.

Pour en faciliter la lecture (et aussi pour des raisons typographiques), les définitions du DEC français ne font pas appel au formalisme des réseaux sémantiques. Le réseau est remplacé par une proposition qui se veut une réalisation linéaire et "francisée" de la représentation sémantique. Ainsi, l'entrée lexicale de ACHETER**1** dans Mel'čuk *et al.* 1988 donne comme définition:

> *X achète**1** Y à Z pour W* = X acquiert de Z les droits de possession complète et permanente de Y en échange d'une quantité d'argent W que Z demande pour Y [=Conv_{3214i}(*vendre**I.1***)].

<u>Avertissement</u>: Lorsque, dans ce travail, nous aurons besoin d'examiner des descriptions lexicales soignées, validées par une recherche poussée, nous nous référerons à l'un des deux volumes du DEC du français déjà publiés: c'est-à-dire Mel'čuk *et al.* 1984

ou Mel'čuk *et al.* 1988. Dans tous les autres cas, nous construirons nos propres exemples. Ceux-ci ne devront alors être considérés que comme des illustrations utiles à l'examen d'un problème purement formel; nous ne prétendons aucunement présenter ici de nouvelles entrées de dictionnaire.

Il n'est pas absolument évident de transcrire les règles de paraphrase sémantique telles que spécifiées ci-dessus dans un véritable formalisme de réseau sémantique. Pour effectuer une telle transformation, au moins trois opérations sont nécessaires:

1) Repérage des Sens de lexème introduits dans la proposition.

Chaque Sens de lexème va donner lieu à un nœud dans un réseau sémantique. Nous trouvons dans notre exemple les Sens 'acquérir', 'argent', 'complet', 'demander', 'droits', 'permanent', 'possession' et 'quantité' plus les "variables" sémantiques 'W', 'X', 'Y' et 'Z'. Normalement, la structure actancielle sémantique peut être trouvée, pour chaque lexème, dans son entrée lexicale. La plupart des lexèmes présents dans notre exemple n'ayant pas encore été décrits dans le cadre du DEC, il faut leur attribuer provisoirement une arité donnée. Ainsi, 'demander' peut être considéré ici comme un prédicat à trois arguments seulement ('X demande Y pour Z') puisqu'il s'agit en réalité de 'demander**I.1**' correspondant à des phrases comme: *"Il demande une somme astronomique pour sa vieille voiture."* Dorénavant, nous allons donc référer à ce Sens comme 'demander**I.1**'.

2) Repérage des "mots-outils" utilisés.

Ceux-ci permettent (i) de mettre en évidence la structure actancielle des lexèmes — cf. *(droits)* ***de*** *(possession)*, *(demande)* ***pour*** *(Y)*, etc., et (ii) de faciliter la lecture de la formule comme si c'était un Texte — cf. ***les*** *(droits)*, *(complète)* ***et*** *(permanente)*, etc. Ces termes ne trouveront pas nécessairement d'équivalent formel dans le réseau sémantique.

3) Analyse de la "syntaxe" de la proposition.

Dans le cas où l'on définit un verbe, la proposition (au sens de la logique) est en fait réalisée par une phrase dont le verbe principal est le *Sens générique* de la définition; ici, par exemple, 'acquérir' est le sens générique de 'acheter**1**'. Une telle analyse permet de mettre en évidence le nœud central "dominant" la définition; ce nœud devra être souligné dans le réseau sémantique. Elle permet aussi d'introduire d'autres nœuds sémantiques qui sont, dans la définition, réalisés par une structure syntaxique particulière. Par exemple, la structure appositive ***une quantité d'argent*** *W* va être rendue dans le réseau grâce à l'introduction d'un nœud sémantique supplémentaire 'être**1II**'.

La figure II-1 donne la règle de paraphrase sémantique (sous forme de réseau) qui peut être associée au (Sens du) lexème ACHETER**1** tel que défini ci-dessus[6].

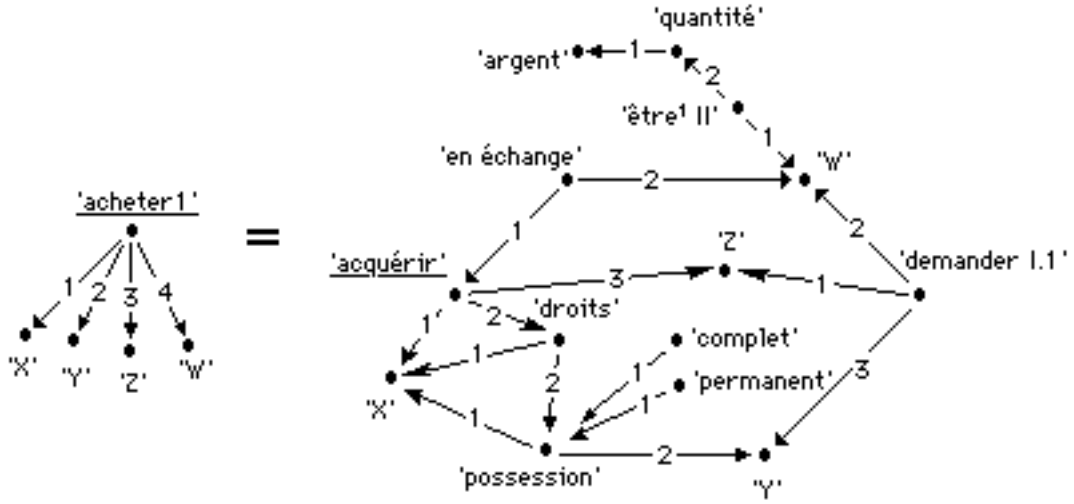


Figure II-1
Définition de ACHETER**1** sous forme de réseau

---

6 Un travail de spécification formelle complète de la pseudo-langue utilisée dans les définitions du DEC permettrait d'effectuer automatiquement le type de manipulation qui vient d'être présentée. C'est, dans l'état actuel du travail lexicographique, une lacune formelle du DEC de ne pas proposer cette formalisation complète.

La spécification de la structure actancielle du lexème défini, dans la partie gauche de la règle, permet de savoir comment s'effectuent les connexions en cas d'application de celle-ci. Cette information est nécessaire dans le cas d'une "réduction" aussi bien que dans celui d'une "décomposition" s'opérant sur un réseau plus vaste correspondant à la RSém d'un Texte. Les nœuds qui représentent les actants indiquent en fait les points d'ancrage du sous-réseau dans le reste du réseau dont il fait partie.

Bien sûr, dans le cas où le lexème défini n'est pas un prédicat mais un objet sémantique, la partie gauche de la règle comporte le seul nœud sémantique pour ce lexème, comme dans l'exemple suivant présentant sous forme de réseau la définition de ÉCOLE**I.1b** de Mel'čuk *et al.* 1988:

> *école**I.1b***. = Ensemble de toutes les personnes qui travaillent et qui étudient**I.3** dans une école**I.1a**.

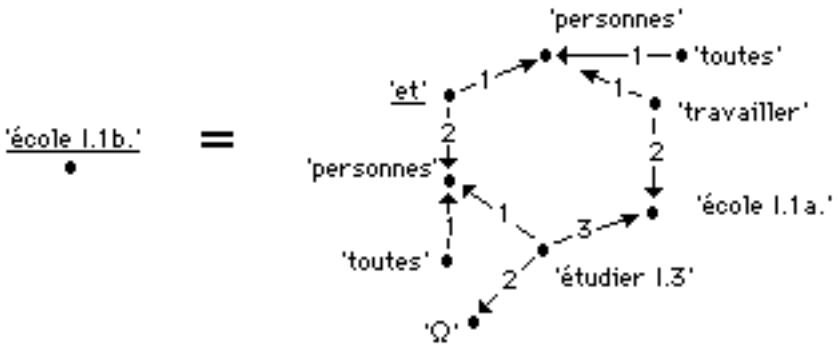


Figure II-2
Définition de ÉCOLE**I.1b** sous forme de réseau

On peut noter ici le caractère ambigu (pour qui n'a pas en main les conventions réglementant la formulation des définitions du DEC) du Sens 'personnes qui travaillent et qui étudient'. Notre représentation en forme de réseau limite le Sens à 'personnes qui travaillent et personnes qui étudient'[7].

---

[7] Noter aussi que, conformément à la convention adoptée en 2.1, nous utilisons ici un seul nœud sémantique 'et' pour représenter un ensemble d'éléments conjoints.

Les deux nœuds soulignés des parties gauche et droite de l'équation, appelés *nœuds dominants* (voir ci-dessus), entretiennent une relation privilégiée: le nœud dominant du definiens est le *Sens générique* du lexème défini. Dans la figure II-1, 'acquérir' est le nœud dominant de la définition de 'acheter**1**' car *acheter**1** quelque chose* signifie avant tout *acquérir quelque chose*. De même, une *école**I.1b*** est avant tout un *ensemble de personnes.*

Puisque le nœud dominant de la définition représente le Sens générique du lexème défini, les procédures qui prennent en charge la lexicalisation, dans un système de synthèse Sens-Texte, devront tenir compte de cette notion de nœud dominant pour ne pas faire jouer les règles de paraphrase sémantique de façon anarchique. En effet, les équations sémantiques ne peuvent être appliquées sur les réseaux (dès qu'un sous-réseau y a été identifié) sans déterminer quel est le nœud dominant de ce sous-réseau. L'exemple suivant, figure II-3, montre une RSém pour laquelle l'application de la règle de la figure II-1 est possible. Au contraire, dans la figure II-4, l'équation sémantique ne doit pas être appliquée sous peine de produire une phrase qui exprime un Sens différent de celui codé dans le réseau initial. Dans chacun de ces exemples, le nœud 'croire**I.1**' est souligné pour indiquer qu'il est considéré, en fonction de la structure communicative, comme le nœud dominant du réseau entier. Nous ne justifierons pas ce choix pour l'instant (voir le Chapitre IV dans la Partie **2**).

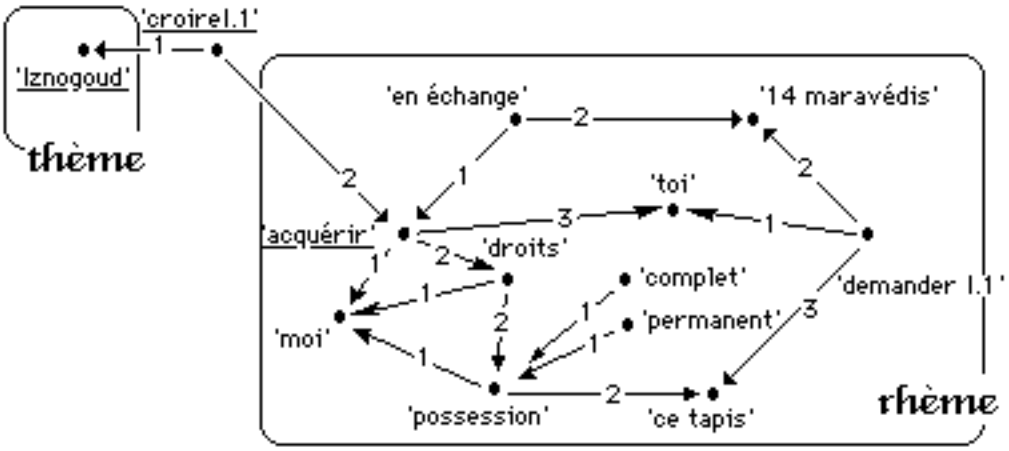


Figure II-3
RSém de (3')

(3') *Iznogoud croit que j'ai acquis auprès de toi les droits de possession complète et permanente sur ce tapis en échange des 14 maravédis que tu demandes pour ce tapis.*
← 'Qu'est-ce que croit Iznogoud?'

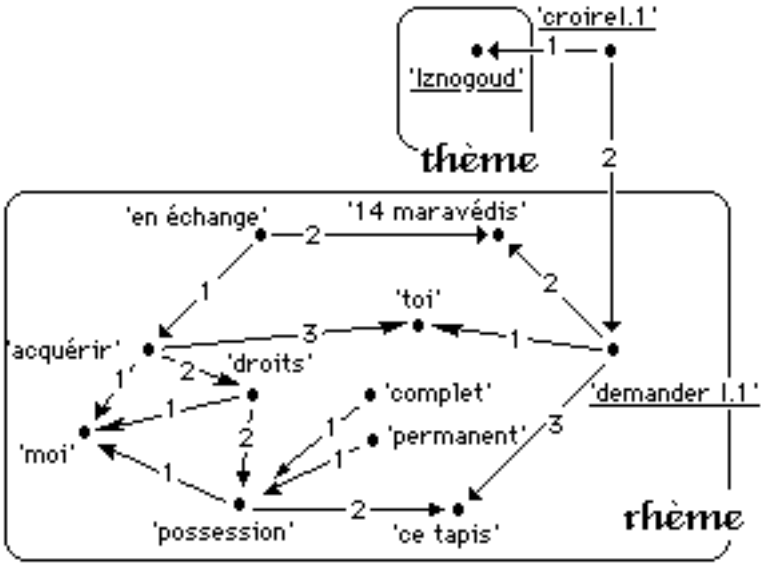


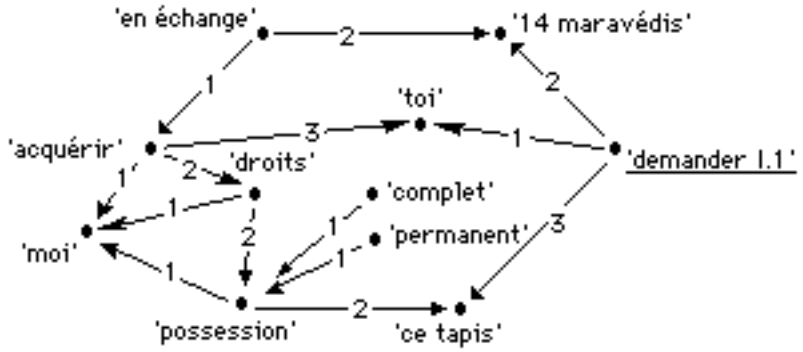
Figure II-4
RSém de (3'')

(3'') *Iznogoud croit que tu demandais pour ce tapis 14 maravédis en échange desquels j'ai obtenu de toi les droits de possession complète et permanente sur ce tapis.*
← 'Qu'est-ce que croit Iznogoud?'

Dans le cas de la RSém de la figure II-3, il est possible d'appliquer la règle de paraphrase sémantique II-1, ce qui permet de produire une phrase comme:

(3'') *Iznogoud croit que je t'ai acheté ce tapis pour 14 maravédis.*

En II-4, par contre, l'application directe de II-1 "effacerait" du réseau l'information qui est vraiment communicativement pertinente pour cette phrase puisqu'elle en domine le rhème. En effet, dans un tel cas, ce que croit Iznogoud c'est avant tout que 'quelqu'un demandait**I.1** quelque chose' et non que 'quelqu'un a acquis quelque chose'. Formellement, l'application de la règle II-1 au réseau de la figure II-4 est impossible car on doit considérer que le sous-réseau de cette figure qui correspondrait à la partie droite de l'équation sémantique (noter que c'est en même temps ici le rhème de la phrase) possède, du fait de la structure communicative du réseau total, un nœud dominant 'demander**I.1**' au lieu de 'acquérir'. Un tel sous-réseau:



ne peut donc "matcher" avec la partie droite de II-1.

Nous pouvons constater que la manipulation des réseaux sémantiques dans un contexte de lexicalisation est un outil de

vérification de la cohérence des descriptions faites dans le dictionnaire. En effet, la TST met l'emphase sur le test de substituabilité, en contexte, d'un lexème à sa définition comme critère d'adéquation des décompositions sémantiques. La lexicalisation, appliquée aux réseaux sémantiques, modélise ce test en permettant d'isoler les éléments superflus, manquants ou mal connectés dans la définition. Peut-être, d'ailleurs, est-il plus pertinent d'un point de vue méthodologique de construire d'abord une RSém où apparaît un nœud sémantique pour lequel on a construit une définition. On peut ensuite tenter de voir si la définition peut se substituer au nœud défini sans poser de problème formel (impossibilité d'assurer proprement les connexions, etc.) ou sémantique (apparition d'incohérences, de redondances, etc.). TOUTE DEFINITION VALIDE DEVRAIT PASSER SANS ENCOMBRE CE TEST DE LEXICALISATION "A L'ENVERS".

Si l'on considère la lexicalisation d'un point de vue procédural, on voit qu'une des difficultés de cette opération est la nécessité de déterminer un nœud dominant pour tous les sous-réseaux que l'on peut isoler dans un réseau global. Afin de faire percevoir intuitivement comment il est possible de mener à bien la lexicalisation en tenant compte de cette contrainte, nous pouvons anticiper sur le cours de notre exposé et énoncer la règle suivante:

> Pour déterminer quel est le nœud dominant d'un sous-réseau, on peut imaginer que le réseau global est "lu" à partir de son nœud dominant, qui est lui-même fonction de la structure communicative du réseau. Au cours de cette lecture, un nœud X va être considéré comme dominant à son tour un nœud Y si X est lu avant Y. Cela est une première description d'une notion que nous avons appelée *dépendance communicative*. Elle sera examinée plus en détail en 4.1.2 et dans la Partie **2**.

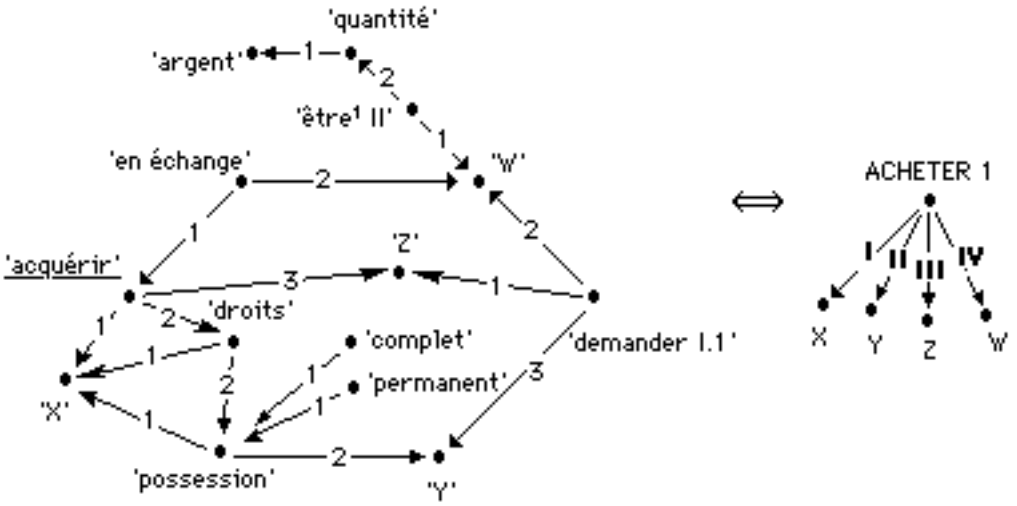
Cette règle, appliquée aux cas des figures II-3 et II-4, permet donc d'expliquer que la lexicalisation n'ait lieu que dans un cas seulement. Nous n'irons pas plus avant dans la description des équations sémantiques. Celles-ci seront considérées à nouveau dans la Partie **2** (même si, rappelons-le, nous ne proposons pas dans ce travail de véritables procédures de lexicalisation). De même, la notion de nœud dominant d'un réseau, associée ici au Sens générique d'une définition lexicale, sera étudiée dans une perspective plus large en 2.3 et IV:1.

Remarque: Toutes les composantes d'une définition n'ont pas le même statut en cas d'application d'une règle de paraphrase sémantique. Ainsi, la composante 'W être^{1}**II** quantité d'argent' de la figure II-1 peut fonctionner comme une contrainte sur la nature sémantique du nœud 'W'. '14 maravédis' peut donc correspondre à ce bloc sémantique si l'on considère qu'il inclut le Sens 'être^{1}**II** quantité d'argent'. Ce Sens est inséré dans la définition pour bien différencier ACHETER**1** d'autres lexèmes comme TROQUER ou ÉCHANGER, mais la façon dont il est manipulé en cas de lexicalisation doit être précisée. Tel qu'annoncé, nous reviendrons plus loin sur la manipulation des réseaux sémantiques lors de la lexicalisation mais sans proposer véritablement une méthode sûre permettant d'effectuer cette opération. Pour être générale, une telle méthode devra se baser sur une reformulation des définitions mettant en évidence la structure complexe du definiens: corps de la définition, présupposés, contraintes sur la nature des sous-réseaux pouvant se substituer aux variables sémantiques, etc. Une recherche approfondie, explorant notamment systématiquement les définitions déjà construites pour le DEC français, serait nécessaire pour mettre à jour le fonctionnement de la lexicalisation et compléter les

observations qui seront faites dans le cadre du présent travail.

Pour conclure sur la formalisation des règles de paraphrase sémantique, remarquons que les équations sémantiques qui viennent d'être présentées sont en général introduites de façon implicite dans la littérature sur la TST. Elles sont fondues dans les entrées lexicales et présentées dans des règles dites *règles lexico-sémantiques* qui sont un raccourci commode pour référer à une entrée lexicale au complet. En réalité, les règles lexico-sémantiques qui sont généralement citées (cf. Mel'čuk & Pertsov 1987:36) établissent une mise en relation entre un sous-réseau sémantique représentant la définition d'un lexème et ce lexème accompagné de sa structure actancielle de syntaxe profonde, spécifiant ainsi, du même coup, sa *diathèse*. De telles règles seraient en fait non-fonctionnelles pour au moins deux raisons:

1) Elles combinent deux opérations de nature complètement différente, voir ci-dessus: la *lexicalisation au sens propre* (qui relève d'une problématique de paraphrase sémantique) et la *lexémisation* (qui relève d'une problématique de correspondance/transition entre niveaux de représentation). Ainsi, la figure II-1 se réécrirait, dans cette optique, de la façon suivante:



2) Puisqu'elles présentent la lexicalisation comme impliquant nécessairement une correspondance avec le niveau de syntaxe profonde, elles empêchent toute lexicalisation récursive. Cependant, le principe du bloc maximal dans les définitions implique que les lexicalisations s'effectuent par étapes successives. Rappelons que ce principe veut que tous les éléments apparaissant dans une définition soient les moins simples possible (c.-à-d., qu'ils soient sémantiquement les plus riches possible). Plus précisément, aucune sous-partie de la définition ne devrait elle-même être la définition d'un lexème. Ce principe rend plus "lisibles" les définitions et s'inscrit dans la méthodologie générale de construction des définitions du DEC: du plus complexe au plus simple, sans dresser au préalable une liste de primitives sémantiques servant à construire les définitions[8]. Par exemple, si le sens 's_1' comporte le sens 's_2' qui lui-même comporte le sens 's_3', il est impossible de lexicaliser 's_3' en 's_1' sans passer par 's_2' puisque, en vertu du principe de bloc maximal, 's_3' n'apparaît pas dans la définition de 's_1'. Il est donc nécessaire, selon la logique interne de la TST, que les véritables règles prenant en charge la

[8] On peut consulter à ce propos Mel'čuk 1989, qui dresse un bilan de la notion de primitive sémantique vue à travers l'approche Sens-Texte tout en présentant d'autres notions importantes comme celle de bloc maximal.

paraphrase sémantique (par le biais des définitions) ne soient pas des règles de correspondance entre deux niveaux de représentation mais des EQUATIONS SEMANTIQUES.

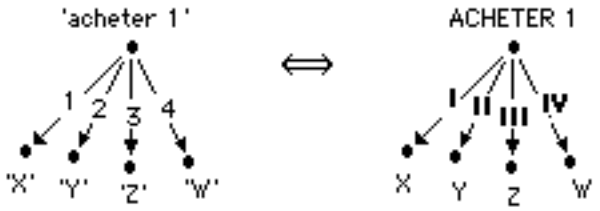
La tentation de présenter les règles lexico-sémantiques sous la forme de règles de correspondance s'explique par une volonté de définir quelque part dans le modèle la façon dont s'effectue cette correspondance RSém ⇔ RSyntP. Nous venons de prouver que ce formalisme de règle n'était pas le bon puisqu'il joue au détriment du processus de lexicalisation. De plus, le recours à ce type de règle n'apporte qu'une solution très approximative au problème de la correspondance RSém ⇔ RSyntP; il manque toujours, notamment, un appareillage décrivant comment sont introduites dans l'arbre syntaxique profond les relations **ATTR**, **APPEND** et **COORD**. Un des objectifs de ce travail est de proposer une formalisation des règles de correspondance RSém ⇔ RSyntP en considérant à la fois les règles statiques et les règles dynamiques impliquées[9].

Remarque: Notons que, dans ce travail, nous n'aborderons pas le problème de l'étude des significations prises en charge par les *fonctions lexicales* comme Magn, Incep, etc. Il faudrait vraisemblablement concevoir un type de règles sémantiques spécial permettant d'introduire, dans la RSyntP, les fonctions lexicales à partir de Sens présents dans la RSém. La puissance descriptive des fonctions lexicales, dans le cadre de la lexicographie, est un fait acquis. Pourtant, leur mode d'utilisation dans un contexte de synthèse, particulièrement au niveau de la transition sémantique, pose de très nombreux problèmes que nous nous permettons d'ignorer ici.

---

[9] En conséquence, nous introduisons en 4, ci-dessous, des règles de correspondance spécifiques à la grammaire.

Le fait de présenter la zone sémantique d'une entrée lexicale sous la forme d'une équation sémantique telle que spécifiée dans les figures II-1 et II-2, ne nous fait pas perdre le lien qui existait dans l'ancienne formulation entre la zone sémantique et le régime. On peut en effet considérer qu'il existe toujours une correspondance triviale entre la partie gauche de la définition et une structure arborescente (syntaxique profonde) minimale. Ainsi, on peut construire dans le cas du lexème ACHETER**1** la règle suivante:



Une telle règle établit simplement une correspondance directe entre actants sémantiques et actants syntaxiques profonds[10]. Cette correspondance, appelée *diathèse* dans le cadre de la TST, est spécifiée dans les entrées lexicales d'un DEC par les équations nommant chaque colonne du schéma de régime: "X = I", "Y = II",... (c.-à-d., l'actant X est le premier actant syntaxique profond, Y le deuxième,...).

2.2 Application facultative des règles de paraphrase sémantique

Nous venons de voir qu'un des principes fondamentaux de manipulation des RSém est la possibilité de substituer un nœud sémantique à un sous-réseau correspondant à la décomposition sémantique de ce nœud, et vice versa. Une telle substitution[11], que nous avons appelée *lexicalisation au sens propre* (par contraste avec la *lexémisation*), est rendue possible par l'emploi des règles de paraphrase sémantique qui viennent d'être décrites. L'application de ces règles n'est pas obligatoire. Il est en effet toujours possible de ne pas opérer

---

[10] Mel'čuk1978:284 mentionne explicitement ce type de règles qu'il appelle (p. 286) "règles sémantico-lexicales".

[11] C'est-à-dire le choix des Sens des lexèmes qui seront présents en syntaxe profonde.

de lexicalisation sans que cela entrave le fonctionnement normal du modèle. D'un point de vue strictement théorique, on peut donc considérer que les deux phrases ci-dessous sont l'une et l'autre acceptables:

(4) a. *Iznogoud croit que j'ai acquis auprès de toi les droits de possession complète et permanente sur ce tapis en échange de la quantité d'argent, que tu demandes pour ce tapis, de 14 maravédis.*
b. *Iznogoud croit que je t'ai acheté ce tapis pour 14 maravédis.*

Il est certain que (4b) est préférable à (4a), comme c'est d'ailleurs (presque?) toujours le cas lorsque l'on doit choisir entre deux paraphrases se distinguant par leur degré de lexicalisation. Il est donc possible d'énoncer le principe général suivant:

Même si, théoriquement, la lexicalisation au sens propre n'est pas une opération dont la réalisation est nécessaire, les phrases les plus lexicalisées seront, en cas de choix, considérées comme les plus convenables stylistiquement.

2.3 Le definiens et la notion de représentation sémantique réduite

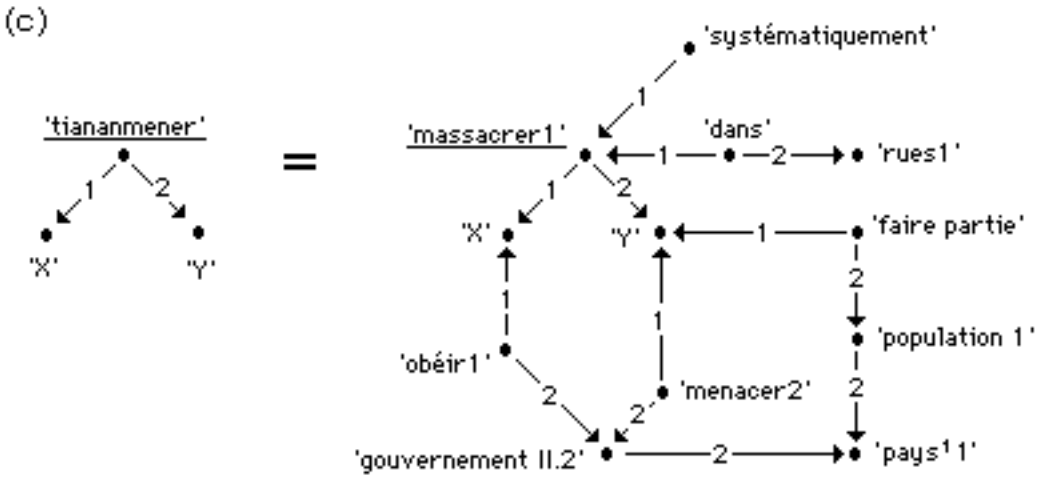
Pour terminer cette description des règles de paraphrase sémantique, nous allons tenter de donner une caractérisation formelle générale du definiens en tant que réseau sémantique.

Nous avons vu que le nœud dominant d'un definiens correspond en fait au Sens générique contenu dans le Sens complet du lexème défini; IL PEUT AUSSI PERMETTRE DE CONSTRUIRE LE SOMMET D'UN SOUS-ARBRE SYNTAXIQUE PROFOND. Un tel sous-arbre peut être associé à toute définition (sous-réseau sémantique définissant) en appliquant les règles de correspondance sémantique qui seront décrites en 4. Par

exemple, supposons que nous construisions un néologisme dont la définition serait[12]:

> *X tiananmène Y* = X, qui obéit**1** au gouvernement**II.2** d'un pays^{1}**1**, massacre**1** dans les rues**1** systématiquement Y qui fait partie de la population**1** de ce pays^{1}**1** et qui menace**2** ce gouvernement**II.2**.

ou, selon une représentation plus formelle:

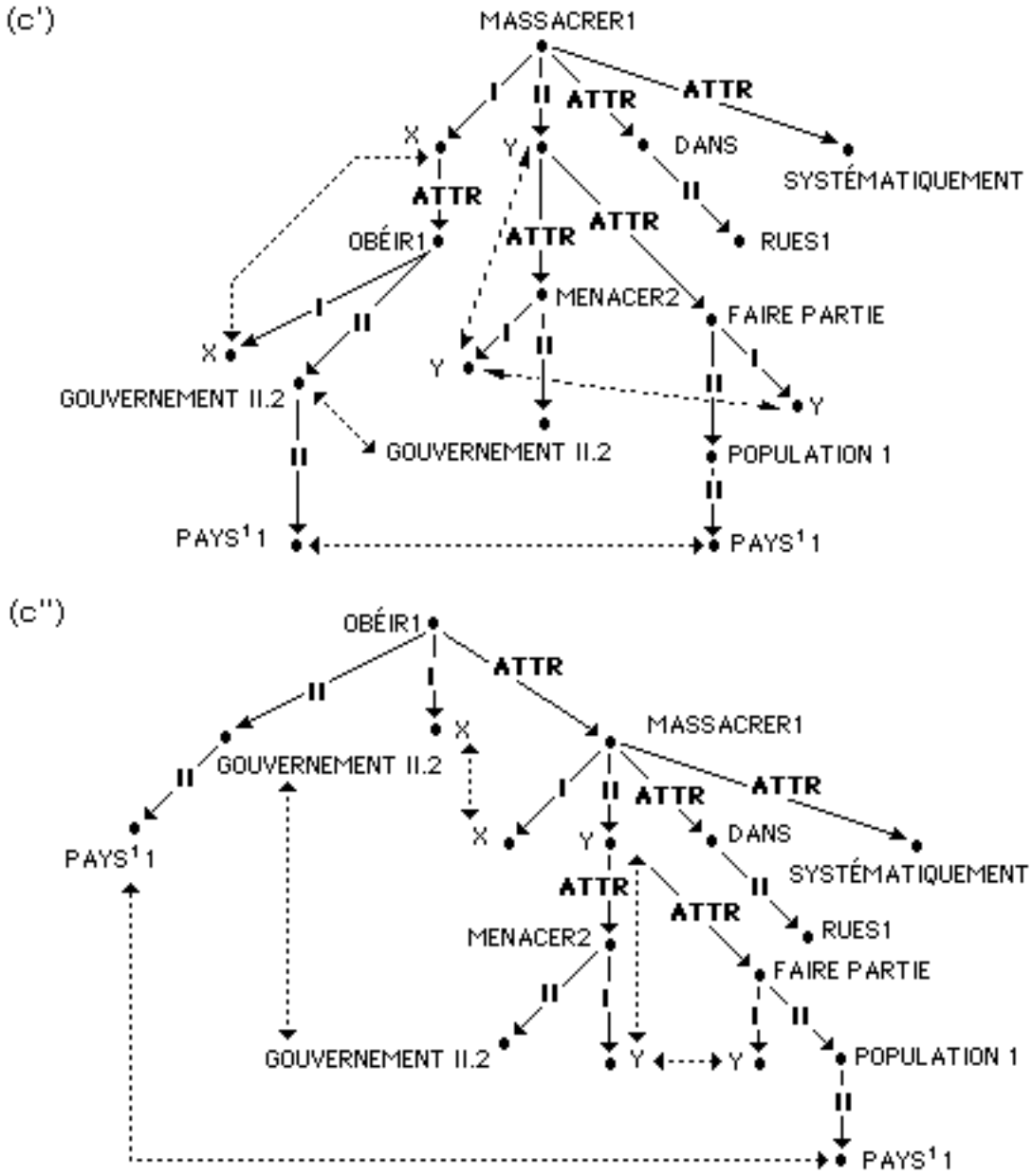


Exemple: *En 1871, les Versaillais ont-ils tiananmené les partisans de la Commune de Paris?*

Les règles de correspondance sémantique ainsi que les procédures de transition RSém ⇒ RSyntP (présentées dans la Partie **2**) vont nous permettre d'associer (c') et non pas (c'') à la partie droite de la règle COMPTE TENU DU NœUD SPECIFIE COMME DOMINANT:

---

[12] Ce néologisme est créé ici sur le modèle de PASTEURISER (de *Pasteur*), en utilisant une composante importante de la réalité extra-linguistique qui l'inspire.



Les réseaux sémantiques définissant un lexème ont donc ceci de particulier qu'ils peuvent directement être associés à un sous-arbre syntaxique profond donné. Cela tient à deux propriétés de ce type de réseau:

1) Les definiens ne sont pas de véritables REPRESENTATIONS sémantiques car ils sont caractérisés par un nœud dominant fixe qui impose une certaine hiérarchie figée entre les nœuds

qui en font partie; ils sont donc un cas particulier de réseau sémantique.

2) Les opérations de lexicalisation ne peuvent être effectuées dans les definiens dont le contenu lexical est contraint par la structuration lexicale de la langue et le principe du bloc maximal; un definiens est donc toujours, par nature, lexicalisé au maximum.

Nous dirons que les definiens sont une variété de *Représentations Sémantiques réduites*. Cette notion de RSém réduite est centrale dans la détermination des procédures d'activation de la composante sémantique. Elle sera étudiée dans la Partie **2**; pour l'instant, nous nous contenterons d'en donner la définition suivante, qui découle des deux propriétés relevées ci-dessus:

Une *RSém réduite* est un réseau sémantique tel qu'il entretient une relation univoque avec un arbre syntaxique profond donné.

Nous ne poursuivrons pas ici l'étude de cette notion qui sera largement discutée plus loin. Il suffit pour l'instant de noter qu'elle ne caractérise pas seulement les definiens, mais qu'elle peut être utilisée pour caractériser tout réseau sémantique dont on désire volontairement limiter la puissance représentationnelle. En effet, considérant notre ébauche de définition de la notion de RSém réduite, NOUS POUVONS DIRE QUE "RESEAU SEMANTIQUE REDUIT" IMPLIQUE "PARAPHRASE REDUITE".

3 LES REGLES GRAMMATICO-SEMANTIQUES

Certaines significations exprimées par les langues naturelles ont un statut bien particulier:

- leur inventaire peut varier radicalement selon la langue considérée;

- elles sont en nombre restreint et sont relativement déconnectées de la réalité extra-linguistique;
- elles forment un système logiquement structuré.

Les significations auxquelles il est fait référence ici sont directement liées au système grammatical de chaque langue et elles doivent intervenir dans les règles de la grammaire: elles sont donc appelées *significations grammaticales*, par opposition aux *significations lexicales*, prises en charge par les règles de paraphrase sémantique qui viennent d'être décrites[13]. Il découle de cela que certaines parties d'une RSém sont traitées par la composante sémantique d'un MST grâce à des règles sémantiques spéciales que nous appellerons *règles grammatico-sémantiques*. (Noter toutefois que ces règles ont été appelées "règles sémantico-morphologiques" dans Mel'čuk 1978:286.) Il existe bien sûr une grande variété de règles de ce type, chaque règle pouvant être classifiée en fonction de la signification qu'elle prend en charge et du moyen linguistique, grammatical ou lexical, qu'elle utilise pour exprimer cette signification. Chaque langue possède son propre jeu de règles grammatico-sémantiques permettant de prendre en compte son propre ensemble de significations grammaticales: ainsi, l'anglais et le français, langues pour lesquelles il existe une signification grammaticale de "nombre", expriment l'opposition singulier/pluriel sur les noms communs grâce à des variations morphologiques (*livre/livres*, *dog/dogs*, *man/men*, ...) alors que le mandarin n'a pas recours au nombre grammatical. (On peut bien sûr exprimer, dans cette dernière langue, le nombre non-grammatical en faisant appel à des moyens lexicaux.)

3.1 Examen d'un type de signification pris en compte

NOUS N'ALLONS PAS PASSER EN REVUE DE FAÇON EXHAUSTIVE LES DIFFERENTS TYPES DE REGLES GRAMMATICO-SEMANTIQUES. NOUS NOUS

---

[13] La terminologie et les moyens de description employés dans cette partie de notre exposé sont très directement tirés de Mel'čuk 1982 et Mel'čuk à paraître:1ère Partie, où l'on peut trouver une classification complète des différentes significations linguistiques.

CONCENTRERONS SUR UN CAS PARTICULIER, CELUI DES REGLES EXPRIMANT DES *GRAMMEMES*. Nous définirons tout d'abord ce qu'est un grammème et examinerons le cas particulier d'un type de grammème du français, puis nous verrons comment on peut formaliser et impliquer dans la transition sémantique les règles grammatico-sémantiques.

Une première caractéristique du grammème est qu'il correspond à une signification obligatoirement exprimée lorsque les lexèmes appartenant à une certaine classe syntaxique sont employés. Ainsi, en français, on dispose de grammèmes temporels car l'expression du temps grammatical est nécessaire lorsque l'on emploie un verbe fini.

Un grammème est donc toujours lié fonctionnellement à une certaine classe syntaxique. Par exemple, pour employer un verbe français, on doit répondre à plusieurs questions, dont "Quel temps grammatical?", pour employer un nom commun, on doit répondre à la question "Quel nombre grammatical?".

Une seconde caractéristique est que les grammèmes forment des *catégories* de significations. Mel'čuk (à paraître) insiste sur le caractère catégoriel de la structuration des grammèmes dans une langue en notant les deux points suivants:

1) Les grammèmes sont organisés en petits ensembles (au minimum des paires) de significations. Par exemple: nombre des noms français = {singulier, pluriel}.
2) Les grammèmes d'une même catégorie s'excluent mutuellement. Ainsi, en français, le singulier et le pluriel des noms ne cohabitent pas: un nom ne peut logiquement pas être employé à la fois au singulier et au pluriel, un verbe ne peut être à la fois au passé et au présent, etc.

Dans sa combinatoire avec les lexèmes, le grammème se distingue d'autres significations grammaticales, dites *significations dérivationnelles* (cf., par exemple, les significations réalisées par les suffixes nominaux en français *-age*, *-tion*, etc.), en ce sens qu'il ne

permet pas de créer de nouveaux lexèmes: on dit que c'est une *signification flexionnelle*. Ainsi, *patate* et *patates* représentent des variations flexionnelles du même lexème PATATE.

Finalement, on peut affirmer que la raison d'être du grammème est avant tout linguistique. L'expression du temps du verbe en français est une contrainte exercée sur le locuteur par la grammaire de la langue. Si l'on prend l'ensemble de TOUTES les significations exprimées par des adjectifs comme *petit*, *bleu*, *méchant*, etc., on voit que, non seulement elles ne forment pas une catégorie[14], mais que, de plus, rien ne force le locuteur lorsqu'il utilise un nom à le qualifier par un adjectif. La langue lui laisse le choix de dire *désir* sans dire *exacerbé* ou *coupable*, de dire *(un) français* sans dire *maudit* ou *petit*, etc.

Il n'est pas toujours évident de décider si telle ou telle signification, dans une langue donnée, est ou non un grammème. Par exemple, en mandarin, une phrase comme:

tā	zài	Zhõngguó
'3e pers.'	'se trouver'	'Chine'

signifie 'elle/il est en Chine', mais peut aussi bien signifier 'elle/il était en Chine' dans un contexte approprié (avec un adverbe de temps indiquant le passé, par exemple, ou bien dans un récit situé de façon évidente dans le passé)[15]. On pourrait donc considérer que cette langue ne possède pas de grammème de temps. Si le verbe ne subit pas, à proprement parler, de variation morphologique, il existe certaines particules dites "aspectuelles" et "modales" qui expriment par des moyens lexicaux des significations proches de celles véhiculées par le temps grammatical en français; mais la NECESSITE ou

---

[14] Il est cependant possible de distinguer, à l'intérieur de la classe des adjectifs, des catégories "générales" non grammaticales; par exemple, la catégorie des adjectifs de dimension: *petit*, *grand*, *moyen*, etc.

[15] En français, on peut très bien dire *"En 1715, le Révèrent Telesphore est en Chine"*, mais cela signifie simplement que le temps grammatical présent de l'indicatif peut servir dans certains contextes à exprimer le passé (présent dit "narratif"). Il n'en reste pas moins que dans de tels cas le verbe français, contrairement au verbe mandarin, possède la marque morphologique évidente d'un temps grammatical.

non d'exprimer ces significations avec le verbe mandarin reste à démontrer.

Pour approfondir ce dernier point et montrer dans quelle mesure la détermination des grammèmes d'une langue peut être problématique, nous allons maintenant examiner un peu plus en détail le cas des articles en français.

3.2 Le cas des articles en français

> "Cette partie du discours est peut-être la plus importante, eu égard à son usage fréquent et continuel, et sa qualité d'être particulière à certaines langues." Landais 1850:273.

Sans le vouloir, l'auteur[16] de cette citation met le doigt sur une caractéristique essentielle de l'article en français: le fait qu'il est "d'un usage fréquent et continuel" n'est que la conséquence du fait QU'IL EST L'EXPRESSION DE QUELQUE CHOSE D'OBLIGATOIREMENT EXPRIME EN FRANÇAIS. Ce caractère obligatoire, que nous tenterons de démontrer dans ce qui suit, est un premier indice du fait qu'il existe logiquement deux façons de rendre compte de l'article français dans le cadre d'un MST. Il peut être décrit soit comme un lexème profond, soit comme un mot grammatical. Dans ce dernier cas, il exprime un grammème et est introduit au moyen de règles grammatico-sémantiques. Pour savoir comment le traiter, il convient donc de déterminer (i) la nature de la signification qu'il exprime, et (ii) le type de moyen d'expression dont il relève.

---

[16] Nous recommandons fortement la lecture de ce livre édifiant dont le but est la présentation "de la solution analytique, raisonnée et logique de toutes les questions grammaticales anciennes et modernes". Comment ne pas y référer?

3.2.1 Signification de l'article

Voici les formes de surface du français qui sont habituellement considérées comme des manifestations des articles[17] et l'information grammatico-sémantique qu'elles véhiculent (voir Grevisse 1975: 278-308):

	± défini	**genre**	**nombre**
un	indéfini	masculin	singulier
une	indéfini	féminin	singulier
des	indéfini	masc/fém	pluriel
le	défini	masculin	singulier
la	défini	féminin	singulier
les	défini	masc/fém	pluriel

Figure II-5
Les articles français

Parmi ces options dans les formes de surface, celles liées à la dichotomie masculin/féminin relèvent sans conteste de la morphologie. Dans le cas d'une description des articles en tant que lexèmes profonds, *le* et *la* seront évidemment considérés comme des variantes morphologiques du même lexème, variantes qui sont contrôlées par un accord en genre avec le nom[18] que l'article détermine. Le cas du nombre, singulier/pluriel, est plus litigieux. Est-ce un Sens indépendant exprimé au moyen de l'article ou un Sens induit en syntaxe via l'accord en nombre de l'article avec le nom? Les logiciens qui se sont penchés sur la description des significations exprimées par les langues naturelles, en utilisant le langage de la logique, ont parfois considéré le nombre comme faisant partie du sens[19] de l'article. Ainsi Montague 1973, de même que Dowty *et al.* 1981, utilisent pour représenter le sens des articles anglais *a* et *the* des formalisations logiques qui définissent ces entités linguistiques

---

[17] Nous ne tenons pas compte ici des formes *au*, *aux*, *du* et *des*, résultant d'une contraction avec la préposition.

[18] Dans la suite de cette discussion, *nom* est utilisé pour référer à tous les éléments nominaux pouvant être déterminés par un article.

[19] Dans la suite de cette discussion, les mots *sens* et *signification* sont évidemment utilisés dans une acception plus large que *Sens* dans la TST.

avant tout comme des quantificateurs: *a* recevant une traduction logique basée sur le quantificateur existentiel (∃) et *the*, une traduction faisant appel au quantificateur existentiel et au quantificateur universel (∀)[20]. Une telle approche du sémantisme des articles ne nous sera guère utile ici puisqu'elle vise essentiellement à rendre compte de la façon dont l'article influence l'interprétation, en termes de dénotation, d'un syntagme nominal.

La grammaire de M. Grevisse *Le Bon Usage* propose, pour l'article, la définition suivante:

> p.278 "L'article est un mot que l'on place devant le nom pour marquer que ce nom est pris dans un sens complètement ou incomplètement déterminé; il sert aussi à indiquer le genre et le nombre du nom qu'il précède." (C'est nous qui soulignons.)

Quand l'auteur dit que l'article "sert aussi à indiquer le genre et le nombre", on peut s'étonner qu'il mette sur le même plan dans cette définition l'opposition défini/indéfini et ce que le reste de sa description dans les pages suivantes semble plutôt présenter comme un simple accord morphologique. On peut noter que, dans la définition de l'adjectif, M. Grevisse ne dit pas que ce mot sert à indiquer le genre et le nombre du nom alors qu'il subit en français les mêmes variations morphologiques que l'article. Ainsi, dans:

Oisive jeunesse
À tout asservie,
Par délicatesse
J'ai perdu ma vie. (Rimbaud)

---

[20] Les traductions exactes, dans le formalisme logique utilisant le calcul-lambda (cf. Dowty *et al.* 1981:195), sont:

(i) pour *a*, "$\lambda P[\lambda Q \exists x[P\{x\} \wedge Q\{x\}]]$";

(ii) pour *the*, "$\lambda P[\lambda Q \exists y[\forall x[P\{x\} \Leftrightarrow x=y] \wedge Q\{y\}]]$".

en l'absence d'article, n'est-ce pas l'adjectif *oisive* qui "indique" le genre de *jeunesse*? Un peu plus loin, dans le même texte, pourquoi l'article assumerait-il plus cette fonction que l'adjectif?

> *Et la soif malsaine*
> *Obscurcit mes veines.*

En fait, l'article français entretient avec la grammaire un lien beaucoup plus étroit que l'adjectif. C'est la raison pour laquelle Grevisse introduit dans sa définition cette composante qui, normalement, devrait apparaître plus loin dans l'approfondissement de la description (présentation des variations morphologiques, placements, etc.). Il est cependant clair que le nombre ne fait pas partie du sémantisme de l'article. Grevisse distingue ainsi, comme beaucoup d'autres grammairiens, deux "espèces" d'articles: l'article défini et l'article indéfini[21], isolant comme véritable Sens de l'article l'expression du caractère défini ou indéfini du nom déterminé.

Que recouvre cette opposition défini/indéfini? Pourquoi et quand en faire usage?

On peut tout d'abord noter que s'il existe un terme pour exprimer les oppositions grammaticales singulier/pluriel ("nombre") et présent/passé composé/imparfait... ("temps"), il n'en n'existe pas en français pour défini/indéfini[22]. Cela tient vraisemblablement au fait que, si défini/indéfini est motivé sémantiquement, son lien avec la réalité extra-linguistique est encore plus indirect que dans le cas des grammèmes de nombre et de temps. Dans certains cas, le Sens exprimé est si difficile à spécifier que l'on peut considérer que seule compte l'opposition défini *vs.* indéfini: le défini est ce qui n'est pas l'indéfini et vice versa. L'opposition défini/indéfini correspond donc parfaitement à ce qui a été appelé plus haut une catégorie: défini et indéfini s'excluent mutuellement (un nom ne pourra être déterminé à

---

[21] L'article partitif étant, selon lui, une variété d'article indéfini.

[22] Nous parlons ici de termes consacrés, d'un usage courant même chez les non-grammairiens.

la fois de façon définie et indéfinie dans une phrase française) et il n'existe pas d'alternative à cette opposition[23].

Ce dernier point pose cependant problème. On pourrait en effet argumenter sur le fait qu'il existe certains emplois des noms qui sont neutres quant à l'opposition défini/indéfini, lorsque, justement, aucun article n'apparaît devant le nom: dans ce cas, il existerait une troisième alternative et la catégorie complète à considérer serait {défini, indéfini, neutre}. Cependant, pour rendre compte de tels cas d'omission de l'article, certains (cf. Guillaume 1969) ont proposé d'admettre l'existence d'un article zéro. On pourrait en effet interpréter le fait que l'article français, tout en présentant un caractère obligatoire, puisse être absent dans certaines constructions comme un indice de l'existence d'un article sans manifestation de surface autre que sa propre absence. Le raisonnement serait le suivant: si l'article exprime une signification qui est obligatoire mais si sa présence en surface ne l'est pas, alors il convient de postuler un article "théorique" qui n'a pas de réalisation de surface mais exprime, par cette réalisation vide, cette signification obligatoire. Furukawa 1986 combat cette thèse au moyen d'une argumentation qui semble très solide, car elle se base notamment sur de nombreux exemples couvrant de façon relativement complète les différents cas possibles d'omission de l'article en français. Sa thèse est que l'omission de l'article n'a pas ici ce qu'il appelle une "fonction positive". Elle correspond à des cas aussi divers que des locutions verbales, des *translations* (voir ci-dessous) ou des quasi-translations, etc.; ces cas n'ayant rien à voir avec l'expression des significations habituellement associées à l'article. Bien sûr, il ne nous appartient pas ici de trancher un tel débat. Notons que l'existence d'un article zéro est peut-être plus facile à établir pour l'anglais, comme le montre la phrase anglaise suivante, que nous avions commise en rédigeant un rapport et dont les erreurs, typiques d'un francophone, ont été corrigées par R. Kittredge:

---

[23] Il existe bien sûr de nombreuses études tendant à attribuer à l'opposition défini/indéfini une valeur plus "concrète" et moins linguistique; cf., Krámský 1972 et, surtout, Hawkins 1978.

> "It would have been possible to ensure the same flexibility for the grammar rules and the lexical entries by making an extensive use of the conditions."

Ici, tous les articles soulignés correspondent à des fautes et devraient, en bon anglais, être supprimés. On voit notamment, dans cet exemple, que certaines constructions sont non-anaphoriques (*grammar rules*, *lexical entries*, *conditions*) du fait de l'absence d'article; pour des constructions équivalentes, le français utiliserait l'article défini. Dans de tels cas, l'absence d'article semble jouer en anglais un rôle positif, au sens de Furukawa 1986.

Il existe certains cas en français où la présence même de l'article manifeste autre chose que l'opposition défini/indéfini. Puisqu'il existe un lien privilégié entre article et nom — l'article fonctionne comme un attribut (au sens large) du nom, la présence d'un article peut être une des manifestations de la *translation* entre classes syntaxiques. Rappelons que la translation est une modification dans le syntactique d'un signe linguistique donné. Dans ce qui suit, nous nous basons sur la notion de translation AU SENS LARGE telle que définie par Tesnière 1959. Ainsi, pour L. Tesnière, *"le tapis de Telesphore"* est un syntagme présentant un cas de translation O > A (nom vers adjectif) puisque *Telesphore* fonctionne ici, par rapport à *tapis*, "comme un adjectif". Les exemples suivants sont des cas de translations basées sur la présence de l'article. Ce sont (a) des translations maintenant lexicalisées, (b) des translations régulières ou (c) des translations néologiques:

a- *le boire* *le manger*

b- *le rouge* *le petit* *le Marcel*

c- *le comprendre*

Dans de tels cas, l'alternance entre présence et absence d'article jouera la plupart du temps comme indice d'une translation et non de l'opposition défini/indéfini (**un boire*, **un Marcel*,...); voir, par exemple:

(5) *Le manteau que j'achète est rouge.*
Le manteau que j'achète est le rouge.

Il est intéressant de constater qu'il n'existe pas de possibilité d'obtenir en français une translation du type de (5) avec emploi indéfini. Ainsi, (6) ne paraît pas très correct; il faudra, pour exprimer ce Sens indéfini dans ce type de contexte syntaxique, utiliser un autre appareillage grammatical, très lourd, comme en (7).

(6) ?? *Le manteau que j'achète est un rouge.*
(7) *Le manteau que j'achète en est un qui est rouge.*

Il existe d'autres cas, présentant le phénomène inverse, où l'ABSENCE d'article en français accentue la translation nom > adjectif, opérée grâce à la construction prépositionnelle:

un chemin <u>de campagne</u> *une maison <u>de bambou</u>*

ou bien permet de réaliser une translation nom commun > (quasi) nom propre, comme dans la phrase:

(8) *Chien, cette nuit je fis un splendide cauchemar.* (F'murr)

Cette dernière construction est décrite par Grevisse 1975:305-306 comme un cas de nom mis en apostrophe. Cela qualifie l'effet de style mais non la portée grammaticale de ce type d'omission de l'article (translation) puisque l'apostrophe en français peut aussi être construite avec l'article[24]:

(9) *Amène-toi, le chien, et plus vite que ça!*

---

[24] On peut noter le caractère très marqué stylistiquement des constructions françaises comme (9). En espagnol castillan, par exemple, cette construction, décrite par Alonso 1984 comme "un 'vocatif' avec une fonction appellative ou expressive", est vraisemblablement moins marquée (même si nous donnons ici un exemple tiré de la langue poétique):

Ni4no, déjame; no pises
mi blancor almidonado. (Lorca)

Pour récapituler ce qui a été dit sur le Sens de l'article français, nous noterons les trois points suivants:

1) L'article français exprime la dichotomie défini *vs.* indéfini. Il n'exprime pas le nombre mais varie en genre et en nombre selon le nom qu'il détermine, en s'accordant avec celui-ci.
2) Les deux Sens possibles de l'article forment une catégorie de significations grammaticales, {défini, indéfini}, qui nécessairement accompagnent tout nom dans une phrase française. Ainsi, dès que l'on veut employer un nom en français, il faut décider si ce nom sera présenté de façon définie ou indéfinie. Les articles français expriment donc des Sens obligatoires.
3) La présence de l'article dans la phrase n'est pas obligatoire dans la mesure où l'expression obligatoire de son Sens peut entrer en conflit avec certaines autres lois de la grammaire qui devront avoir la priorité (cf. les cas de translation avec omission de l'article, ci-dessus).

Nous avons concentré ici notre attention sur l'article, mais il va de soi qu'une étude complète devrait tenir compte des lexèmes commutant avec l'article: adjectifs démonstratifs, adjectifs possessifs, etc. Ces lexèmes du français exprimant autre chose que des grammèmes, la commutation avec les articles rend litigieux et critiquable le traitement grammatical que nous proposons ci-dessous.

3.2.2 Traitement grammatical des articles du français

Les remarques qui viennent d'être faites montrent qu'il serait possible de considérer que l'article exprime en français un grammème appartenant à la catégorie {défini, indéfini}. Pour une catégorisation plus sérieuse de l'article français, on peut se reporter à Furukawa

1986 déjà cité[25]. S'il exprime un grammème, l'article est néanmoins un mot de la langue, qui doit être décrit dans le dictionnaire. Cependant, en tant que mot grammatical, il sera introduit par la composante sémantique d'un MST au moyen de règles grammatico-sémantiques. Nous ne pensons pas qu'il soit pertinent d'associer une définition à l'article, définition qui lui permettrait d'être introduit par des règles de paraphrase sémantique. Nous avons vu qu'il est intimement lié à la grammaire du français et, donc, sa description doit refléter cette particularité.

Nous donnerons, dans la section suivante, quelques exemples montrant que cette approche est dans la pratique très satisfaisante et qu'elle permet de contrôler de façon "élégante" l'apparition des articles dans la phrase française. Elle rend compte du fait que l'article (en tant qu'expression d'un grammème) est, au niveau de la communication, dépendant du nom. Il est évident que l'on cherche d'abord à exprimer le contenu sémantique du nom, puis que, en fonction des contraintes grammaticales propres au français, on se trouve obligé d'exprimer le caractère plus ou moins défini du Sens de ce nom. Communicativement parlant, l'expression des significations associées à un article est uniquement motivée par la présence du nom et l'emploi du nom est une condition nécessaire à l'emploi de l'article, comme l'emploi du verbe est une condition nécessaire à l'emploi du temps grammatical[26].

Cet examen (un peu long mais nécessaire) du cas de l'article français nous montre à quel point la détermination des grammèmes d'une langue est une tâche complexe. Nous allons maintenant laisser

---

[25] Rappelons que nous utilisons ici le cas de l'article comme une ILLUSTRATION et que nous ne prétendons nullement faire le tour de la question. Pour prouver le caractère grammatical du Sens de l'article, il convient d'examiner de façon exhaustive tous ses emplois possibles (anaphorique, cataphorique, avec ou sans un sens générique, etc.), tâche que nous ne pourrions entreprendre dans le cadre de ce travail.

[26] On trouve dans Apresjan *et al.* 1989:134 un exemple de règle traitant grammaticalement l'article anglais dans le cadre d'un système de traduction automatique de l'anglais au russe fondé sur l'approche Sens-Texte. La comparaison avec ce qui est proposé ici ne peut cependant pas aller très loin dans la mesure où le système décrit, ETAP-2, ne fait pas appel à de véritables niveaux de RSém et RSyntP mais à une représentation syntaxique dite "normalisée" comme niveau de représentation le plus profond.

de côté l'article pour revenir au problème plus général de la manipulation des grammèmes par des règles grammatico-sémantiques.

3.3 Formalisation des grammèmes et intervention dans la transition sémantique

Nous avons vu, en I:4.2, que la notion de choix était au centre du problème de l'implémentation des transitions Sens-Texte, et de la transition RSém ⇒ RSyntP en particulier. Plus loin, nous avons vu que, justement, les règles de paraphrase sémantique étaient un outil de base permettant d'effectuer certains choix lexicaux au niveau de la RSém d'un Texte. Contrairement à ces dernières règles, les règles grammatico-sémantiques offrent peu d'alternatives quant à leur application. Leur fonction est de *grammaticaliser* la RSém — c'est-à-dire, d'exprimer les différentes significations grammaticales présentes dans la RSém. Dans le cas des grammèmes, cette *grammaticalisation* est quasi-obligatoire si l'on considère que les transitions entre niveaux de représentation sont effectuées en série, c'est-à-dire en traitant complètement et définitivement un niveau de transition avant de passer au suivant. Il est certain que dans un texte comme:

Il pleuvait fort sur la grand-route,
Elle cheminait sans parapluie.
J'en avais un, volé sans doute
Le matin même à un ami. (Brassens)

le participe passé *volé*, employé comme adjectif, ne porte que le passé comme marque de temps grammatical. Cependant, la RSém de la phrase dans laquelle il apparaît doit contenir un grammème très complexe permettant de calculer un temps grammatical qui se trouve être exprimé de façon implicite même s'il n'est pas réalisé en surface. En effet, *J'en avais un, volé sans doute...* est en relation de paraphrase avec:

J'en avais un que j'avais volé sans doute
le matin même à un ami.

mais pas avec:

J'en avais un que j'allais voler sans doute...[27]

Le grammème de temps doit donc être exprimé, c'est-à-dire qu'il doit être identifié et traduit au niveau syntaxique profond par la composante sémantique au moyen de règles grammatico-sémantiques.

Les règles grammatico-sémantiques s'appliquent au niveau sémantique et mettent en correspondance une signification grammaticale et un "indicateur grammatical" de syntaxe profonde. Contrairement aux règles de paraphrase sémantique qui opèrent sur un seul niveau de représentation, elles sont une variété de règles de correspondance. On peut les schématiser de la façon suivante:

'Sens grammatical' ⇔
Marqueur grammatical (du niveau) syntaxique profond

L'écriture d'une règle de ce type nécessite une étude systématique du phénomène grammatical considéré. Comme dans le cas des règles de paraphrase sémantique, nous ne proposons ici aucune description nouvelle pour le français et nous nous limitons à quelques considérations purement formelles. Nous construirons donc provisoirement quelques règles grammatico-sémantiques à seule fin d'illustrer leur principe de fonctionnement.

Nous pouvons considérer que la phrase:

(10) *Le poète se souvient d'avoir volé un parapluie à un ami.*
← 'Le poète se souvient de quoi?'

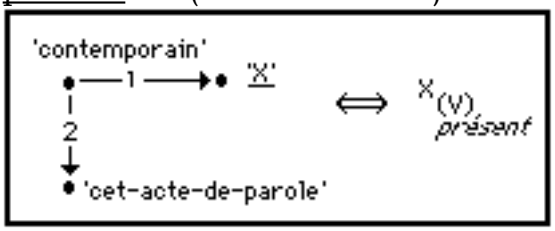
exprime, notamment, les cinq significations grammaticales suivantes pour lesquelles nous donnons des règles grammatico-sémantiques

---

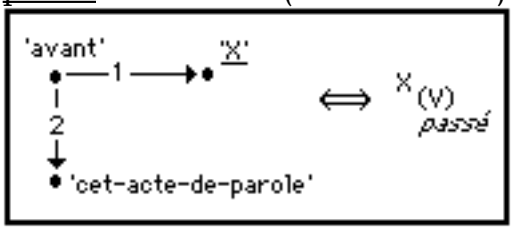
[27] Cette alternative est grammaticalement tout à fait acceptable; c'est seulement la situation qu'elle dénote qui est un peu plus complexe à imaginer.

prenant en charge, de façon très simpliste, la description sémantique des marqueurs grammaticaux du niveau syntaxique profond correspondants:

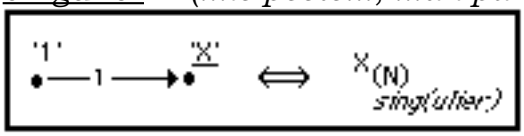
<u>présent</u> (*...se souvient...*)



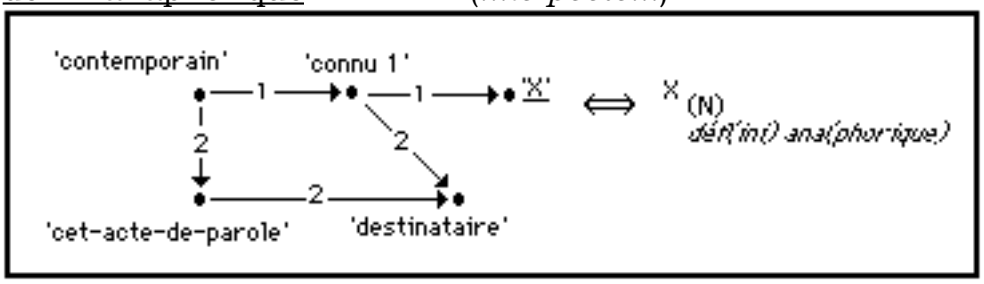
<u>passé</u> (*...avoir volé...*)



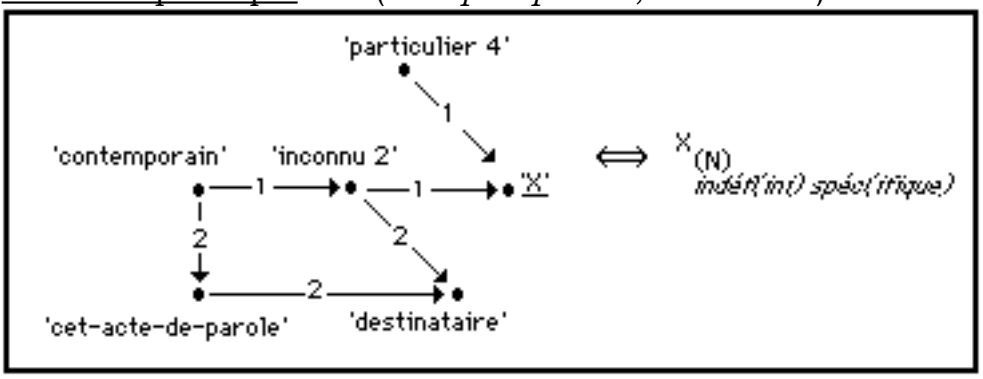
<u>singulier</u> (*...le poète...*, *...un parapluie...*, *...un ami...*)



<u>défini anaphorique</u> (*...le poète...*)



<u>indéfini spécifique</u> (*...un parapluie...*, *...un ami...*)



Pour ce qui est ici des significations exprimées par les articles, nous nous inspirons de la description de Furukawa 1986 qui distingue entre autres: (i) les emplois définis anaphoriques *vs.* définis cataphoriques (*Le livre est enfin écrit/ Il a enfin écrit le livre sur lequel il travaillait tant*); (ii) les emplois indéfinis spécifiques *vs.* non spécifiques (*Il cherche un coupable*[quelqu'un en particulier qui se trouve être coupable]/[quelqu'un qui soit coupable]). Ici, nous rendons compte de ces nuances sémantiques de façon très grossière en utilisant les Sens des lexèmes CONNU**1**, INCONNU**2** et PARTICULIER**4**.

Remarque: Dans nos deux derniers exemples de règles grammatico-sémantiques, les nœuds 'connu**1**' et 'inconnu**2**' sont en fait des abréviations pour, respectivement, 'connu**1** du destinataire de cet acte de parole au moment de cet acte de parole' et 'inconnu**2** du destinataire de cet acte de parole au moment de cet acte de parole'. Cette simplification est aussi utilisée dans la figure II-6 ci-dessous.

Les exemples de règles grammatico-sémantiques proposés ici sont, nous l'avons dit, très simplistes. Nous nous basons sur des notions grammaticales assez générales comme "passé", "présent", "indéfini spécifique", etc., en leur faisant correspondre des Sens qui ne sont que des indicateurs de ce que pourraient être les Sens véritables mis en jeu.

Soit la RSém de (10):

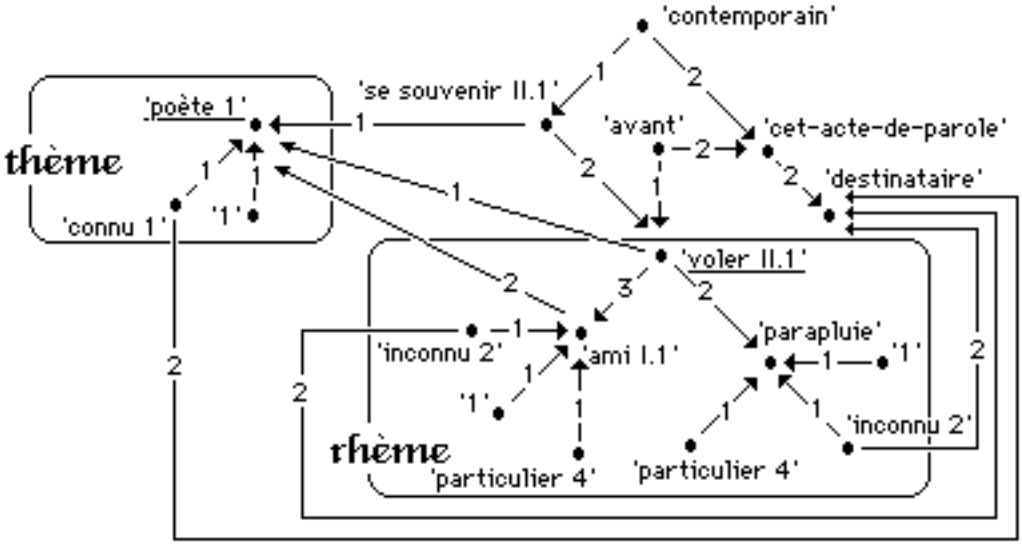


Figure II-6
RSém de (10)

Les règles grammatico-sémantiques énoncées plus haut nous permettent de construire l'arbre syntaxique profond de la figure II-7 à partir de la figure II-6. (La production d'un tel arbre présuppose aussi bien sûr d'autres opérations que la grammaticalisation, opérations dont il n'est pas question pour l'instant.)

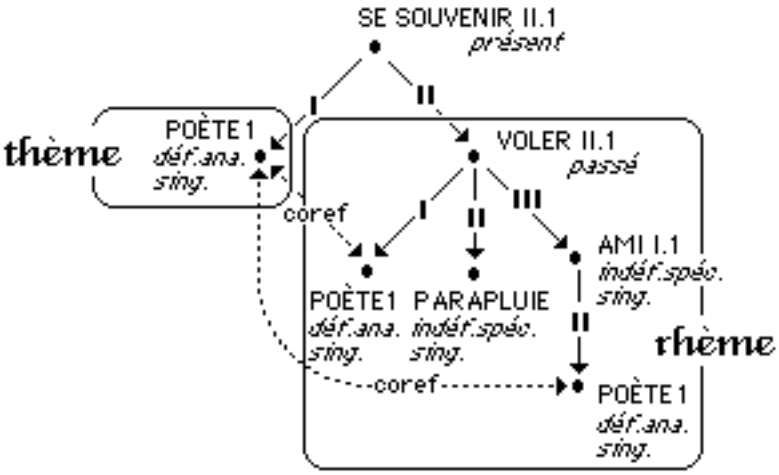


Figure II-7
RSyntP de (10)

Les marqueurs grammaticaux associés aux différents nœuds permettent de calculer le temps grammatical pour chaque nœud verbal et d'introduire le lexème de l'article comme un DEPENDANT SYNTAXIQUE DE SURFACE des différents noms communs. La figure II-8 donne la RSyntS résultant de ces calculs:

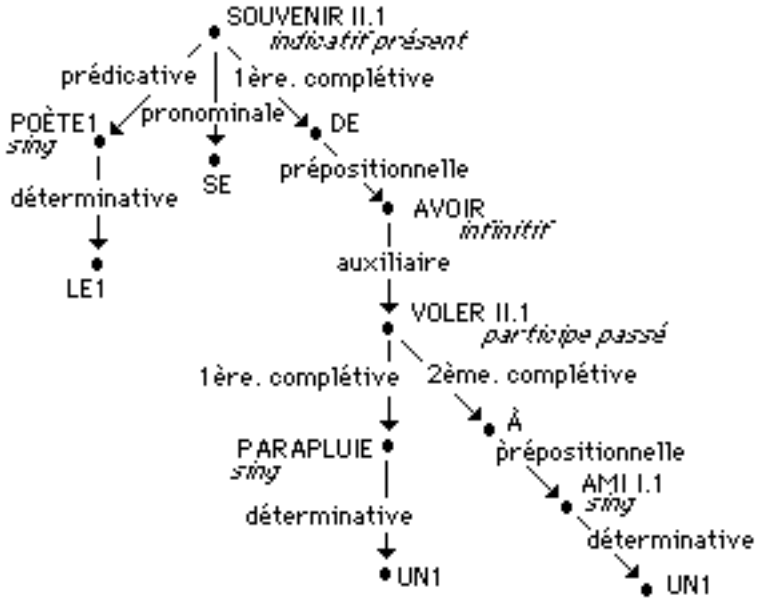


Figure II-8
RSyntS de (10)

Dans un contexte de transition RSém ⇒ RSyntP, les règles grammatico-sémantiques sont donc appliquées de façon quasi-obligatoire pour transmettre aux niveaux postérieurs de représentation une information permettant de contrôler des processus aussi divers que le calcul de certaines constructions grammaticales et indicateurs morphologiques (temps grammatical des verbes, nombre des noms, etc.) ou l'apparition de certains lexèmes grammaticaux (articles, verbes auxiliaires, prépositions régies, etc.). L'effet de ces règles sur les transitions suivantes n'est pas immédiat; par exemple, dans le cas des articles, ce n'est pas parce qu'un lexème nominal d'une RSyntP est marqué comme "défini anaphorique" qu'il possèdera, au niveau syntaxique de surface, un dépendant déterminatif LE**1**. Illustrons ceci par un exemple.

Supposons que nous disposions d'un programme informatique capable de déterminer la structure d'un texte et le contenu informatif que chaque phrase de ce texte doit exprimer. Supposons aussi que ce programme puisse construire une RSém pour chacune de ces phrases, RSém faisant apparaître une structure sémantique (comprenant tous les Sens lexicaux et grammaticaux) et une structure communicative. On peut imaginer que les deux grammèmes de "défini anaphorique" et "indéfini spécifique" seront calculés notamment en fonction de ce qui a été exprimé dans les phrases précédentes. Ainsi, si la phrase **n** exprime 'hier j'ai été très surpris de voir des extra-terrestres', la phrase **n+1** exprimera 'les extra-terrestres traversaient la route nationale' plutôt que 'des extra-terrestres traversaient la route nationale'[28]. Supposons cependant que la phrase **n+1** ait un rhème différent indiquant, par exemple, une donnée nouvelle qui est le nombre d'extra-terrestres aperçus ce jour-là. Dans ce cas, même si le Sens exprimé sera quelque chose comme 'en effet les trois$^{\textbf{rhème}}$ extra-terrestres traversaient la route', 'les' ne peut être exprimé en surface par un article car il effacerait du même coup le caractère rhématique de 'trois'. Comparons (11) et (12):

(11) **Hier, j'ai été très surpris de voir des extra-terrestres. En effet, les trois$^{\textbf{rhème}}$ extra-terrestres traversaient la route.*

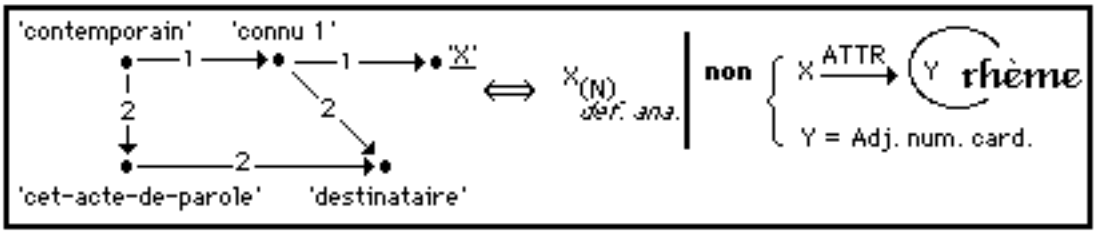
(12) *Hier, j'ai été très surpris de voir des extra-terrestres. En effet, trois$^{\textbf{rhème}}$ extra-terrestres traversaient la route.*

Tout comme la réalisation d'un temps grammatical est à la fois motivée par la présence, au niveau sémantique, de certains grammèmes de temps et par la réalisation de certaines configurations syntaxiques, l'apparition de l'article est elle-aussi contrôlée par d'autres facteurs que la seule application d'une règle grammatico-sémantique. On pourrait bien sûr, dans l'exemple qui vient d'être donné, rendre compte de ces phénomènes au moyen de conditions sur

---

[28] Nous faisons grâce au lecteur des RSém (graphiques). De plus, ici, 'les' et 'des' renvoient en fait aux significations grammaticales correspondantes.

l'application des règles. La règle prenant en charge le grammème de "défini anaphorique" pourrait ainsi être réécrite de la façon suivante:



mais, en fait, une condition ainsi formulée tente de régler au niveau de la correspondance RSém ⇔ RSyntP un problème qui relève strictement de la syntaxe. C'est le fait qu'il y ait un numéral rhématisé comme déterminant du nom qui rendrait incorrecte l'utilisation de l'article défini. Dans de tels cas, la présence de l'article doit se décider au moment de la construction de la RSyntS grâce aux règles de la composante syntaxique profonde du modèle.

4 LES REGLES DE CORRESPONDANCE SEMANTIQUE PROPREMENT DITES

Dans ce chapitre, nous avons jusqu'à présent introduit trois types de règles déclaratives qui appartiennent à la composante sémantique d'un MST:

- les règles de bonne formation, qui définissent la syntaxe des RSém et des RSyntP;
- les règles de paraphrase sémantique (appelées aussi "règles lexico-sémantiques"), qui prennent en charge la lexicalisation de la RSém;
- les règles grammatico-sémantiques, qui prennent en charge la grammaticalisation de la RSém.

Il nous reste maintenant à présenter les *règles de correspondance sémantique proprement dites* — que nous appellerons ici simplement *règles de correspondance sémantique*. Elles se distinguent des autres règles sémantiques par le fait qu'elles permettent

véritablement d'établir le lien entre les deux niveaux de représentation considérés: RSém et RSyntP. Dans un contexte de transition RSém ⟩ RSyntP, qui est au centre de la présente étude, les règles de correspondance sémantique permettront d'effectuer l'opération que nous avons appelée *arborisation*, qui correspond à la CONSTRUCTION proprement dite de l'arbre syntaxique profond (voir le Chapitre VI dans la Partie **2**).

4.1 Formalisation des règles de correspondance sémantique

La seule publication aisément accessible qui donne un ensemble COMPLET de règles de correspondance d'un MST est *Surface Syntax of English* (Mel'čuk & Pertsov 1987; dorénavant SSE), pour la composante syntaxique de surface. Mentionnons toutefois que Mel'čuk 1974 présente des règles de correspondance pour le russe, notamment pour la composante syntaxique profonde; des règles syntaxiques de surface pour le français ont aussi été formulées dans Apresjan *et al.* 1985. Les règles de correspondance RSém ⇔ RSyntP, quant à elles, ne sont même pas mentionnées dans les textes présentant le modèle. Ainsi SSE, dans sa partie introductive, a recours à des règles lexico-sémantiques mixtes qui effectuent une correspondance (cf. nos remarques en 2.1 sur la validité de telles règles). Ou encore, Mel'čuk 1978 ne donne, pour la composante sémantique, que des règles prenant en compte des opérations comme la lexémisation ou la grammaticalisation. Jusqu'à présent, il n'était pratiquement pas question d'un système de règles de correspondance bien isolé, équivalent dans son principe aux systèmes de règles de correspondance des autres composantes (RSyntP ⇔ RSyntS, RSyntS ⇔ RMorphP, etc.). Nous allons donc, dans un premier temps, résumer le formalisme des règles de correspondance déjà formulées pour des MST, puis introduire de façon analogique le formalisme que nous proposons pour les règles de correspondance RSém ⇔ RSyntP. Afin de caractériser complètement ce type de règle, nous serons amené à décrire un nouveau type de dépendance linguistique: la *dépendance communicative*.

4.1.1 Les règles de correspondance dans SSE[29]

Une règle de correspondance de la composante syntaxique de surface a la forme:

RSyntS ⇔ RMorphP | Conditions

où RSyntS est un sous-arbre minimal syntaxique de surface (au moins deux nœuds liés par une relation syntaxique de surface), RMorphP est une sous-représentation morphologique profonde spécifiant notamment la façon dont se linéarise la relation considérée, et Conditions, un ensemble de conditions d'application d'une telle règle. Examinons la figure II-9 qui présente une règle de SSE décrivant un des cas de correspondance pour la relation parenthétique (règle 13.10., p. 301):

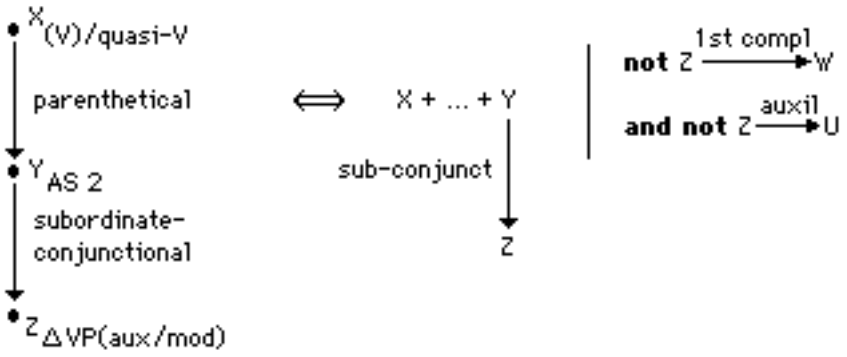


Figure II-9
Une règle RSyntS ⇔ RMorphP pour la relation parenthétique

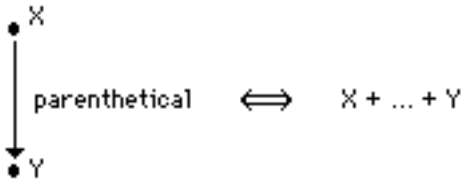
Un des exemples donnés pour cette règle est:

---

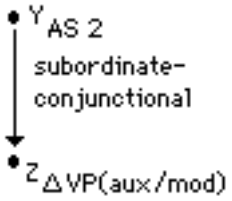
[29] En plus des règles de SSE et de Apresjan *et al.* 1985, nous avons pu examiner un certain nombre de règles de correspondance RSyntP ⇔ RSyntS de Mel'čuk 1974 grâce aux explications fournies par L. Iordanskaja. Nous y ferons appel à la fin de cette sous-section.

> To prevent the idea of games from becoming vulgarized[X], as[Y] so many terms are[Z], it should be emphasized once more that this is a very precise idea.

Le formalisme de règle qu'illustre la figure II-9 semble présupposer que l'on sépare explicitement les conditions d'application et la règle de correspondance comme telle. Cependant, tout ce qui n'est pas ici:



peut, d'une certaine façon, être considéré comme une condition d'application. Les spécifications des valeurs possibles des nœuds en fonction des classes syntaxiques ((V)/quasi-V), des syntagmes (ΔVP), ou des lexèmes (AS**2**) peuvent être interprétées comme des conditions. De la même façon, le sous-arbre:



qui spécifie le "contexte d'arborescence" (tree context) dans la partie gauche et droite de la règle, ne fait qu'introduire une condition contextuelle sur l'application de celle-ci. Dans les implémentations de la composante syntaxique de surface que nous avons effectuées (cf. Polguère 1988a et 1989), nous avons complètement dissocié les trois parties des règles syntaxiques de surface[30]:

1) <u>partie gauche</u>, qui indique le sous-arbre minimal devant coïncider avec une partie de l'arbre syntaxique de surface global;
2) <u>partie droite</u>, qui indique la chaîne linéaire à construire;
3) <u>partie conditions</u>, qui indique toutes les conditions possibles d'application de la règle.

---

[30] Nous spécifions ici ces trois parties en fonction d'une transition RSyntS ⇒ RMorphP; il est aisé d'établir ces spécifications en fonction d'une transition inverse.

Selon cette méthode, la règle de la figure II-9 se reformulerait ainsi:

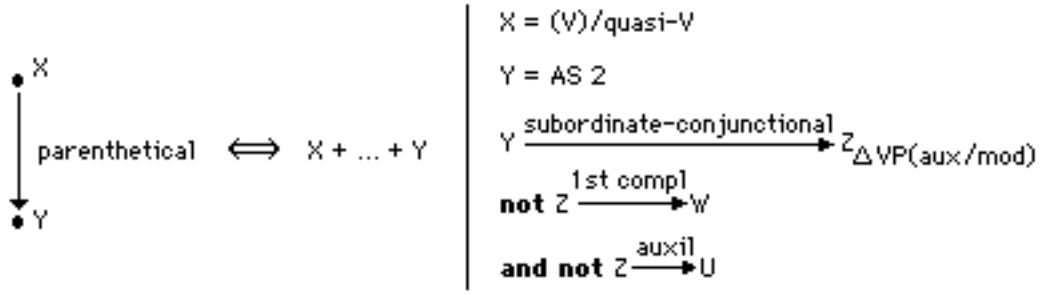


Figure II-10
Reformulation de la règle de la figure II-9

Une telle formulation correspond assez directement à l'interprétation procédurale qui doit être faite des règles de transition mais est certainement d'une lecture moins aisée. En décrivant les règles de correspondance RSém ⇔ RSyntP, nous ne chercherons donc pas nécessairement à donner un exact reflet de la façon dont nous les avons implémentées, mais plutôt à copier l'approche adoptée dans SSE pour la composante syntaxique de surface. Il s'agit aussi de refléter une certaine classification que l'on peut faire des conditions.

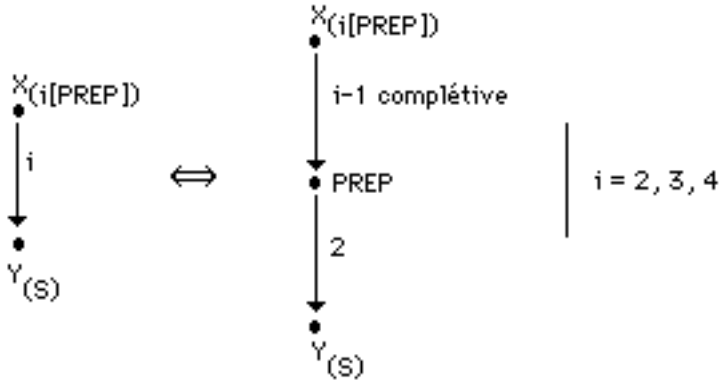
Apparaissent en général de part et d'autre du symbole "⇔" les conditions qui sont satisfaites directement:

- en allant examiner dans le dictionnaire les entrées des lexèmes étiquetant les nœuds: classe syntaxique du lexème (V), nom du lexème lui-même (AS**2**), etc.
- en testant la présence d'une certaine configuration: un syntagme (VP), etc.

Apparaissent dans la partie Conditions de la règle, les conditions qui sont plus complexes à exprimer et qui ne correspondent pas à de simples caractérisations de lexème; par exemple, les formules logiques: **si** A **alors** B, etc.

Cette distinction est particulièrement évidente lorsque l'on considère les règles RSyntP ⇔ RSyntS de Mel'čuk 1974. On y voit notamment que les conditions basées sur la consultation du schéma

de régime d'un lexème étiquettent les nœuds des représentations — comme, par exemple, dans la règle 10 (p. 244):



Dans une telle règle, $X_{(i[PREP])}$ limite la correspondance aux sous-arbres syntaxiques profonds ayant pour sommet un lexème dont le schéma de régime mentionne l'utilisation de la préposition (PREP) introduisant au niveau syntaxique de surface le i-ème actant syntaxique profond. Dans un contexte de transition RSyntP ⇒ RSyntS, on pourrait dire que cette condition permet en fait "d'introduire" PREP au niveau syntaxique de surface. Si l'on considère la règle d'un point de vue purement déclaratif, on voit bien que la condition correspond à une caractérisation (d'une partie du syntactique) d'un lexème. Il faut remarquer que, dans cet exemple, la formule qui est formellement présentée comme une condition ("i = 2, 3, 4"), serait, selon les nouvelles conventions adoptées dans SSE, introduite directement sous la règle de transition elle-même. Cette formule spécifie les valeurs possibles d'une variable et ne doit pas être considérée comme une véritable condition d'application.

4.1.2 Les règles de correspondance sémantique et la notion de dépendance communicative

La structure générale d'une règle de correspondance sémantique va être de la forme:

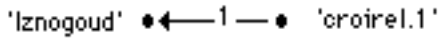
RSém ⇔ RSyntP | Conditions

Il reste maintenant à examiner ce que peut contenir chacune des trois composantes d'une telle règle.

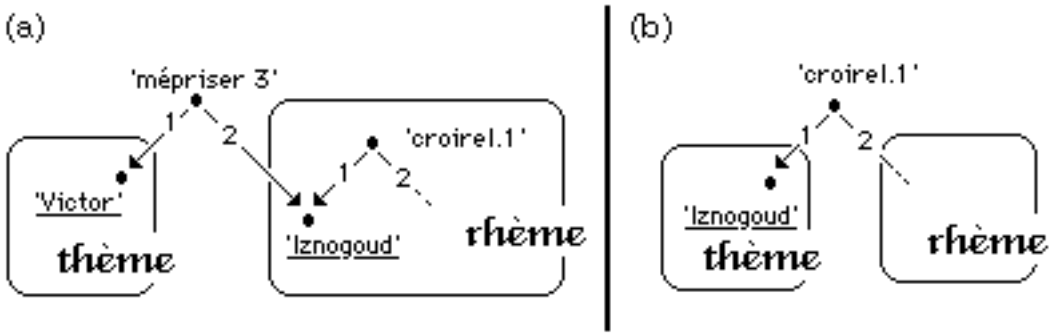
La RSém considérée devrait être un sous-réseau sémantique minimal comportant au moins deux nœuds sémantiques liés par une relation actancielle. La RSyntP sera un sous-arbre syntaxique profond et Conditions, un ensemble (qui peut être vide) de conditions restreignant l'application de la règle en fonction du contexte, de certaines particularités des lexèmes considérés, etc. Comme nous venons de le voir pour d'autres niveaux de correspondance, certaines conditions pourront en fait être spécifiées directement dans les parties gauche et droite de la règle.

Les relations sémantiques étant au nombre de six et les relations syntaxiques profondes au nombre de neuf, on peut prédire que les règles de correspondance sémantique d'un MST sont en nombre limité. On doit donc pouvoir les décrire exhaustivement; leur écriture pose néanmoins de multiples problèmes.

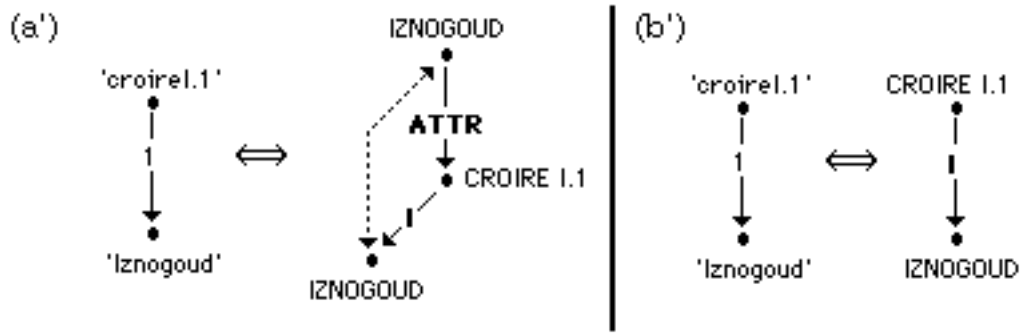
Tout d'abord, considérons ce sous-réseau sémantique extrait de la figure II-3:



Nous voudrions démontrer que l'on ne peut pas écrire de règle établissant une correspondance avec un sous-arbre syntaxique profond à partir d'un tel sous-réseau sémantique. Il manque, dans ce sous-réseau, une information nécessaire sur son contexte d'apparition dans le réseau global. Ainsi, la correspondance établie sera différente selon que l'on considère un réseau comme (a) ou (b).



(a) pourrait être la RSém d'une phrase comme *Victor méprise* ***Iznogoud qui croit*** *qqch.*, alors que (b) correspond simplement à ***Iznogoud croit*** *qqch.* En clair, (a) et (b) impliquent, respectivement, les correspondances suivantes (a') et (b'):



IL SEMBLERAIT DONC QUE L'APPARITION D'UNE RELATION COMME **ATTR** EN SYNTAXE PROFONDE PUISSE ETRE FONCTION DE LA STRUCTURE COMMUNICATIVE DU RESEAU GLOBAL D'OU LE SOUS-RESEAU EXAMINE EST EXTRAIT. La structure communicative de (a) nous permet d'identifier le nœud 'mépriser**3**' comme nœud dominant (rappelons que les règles de calcul du nœud dominant d'une RSém seront introduites dans la Partie **2**). Il est possible de déduire de cette dominance absolue d'un nœud sur tous les autres une certaine relation de dominance entre tous les nœuds de la RSém liés par une relation actancielle. Ainsi, dans le cas de (a), le nœud 'Iznogoud' devient "par transitivité" le nœud dominant de la paire 'Iznogoud'/'croire**I.1**'. Au contraire, en (b), la structure communicative du réseau isole 'croire**I.1**' comme nœud dominant; il domine donc dans ce cas 'Iznogoud'. Nous ne décrirons pas ici les règles qui nous permettent de déterminer ces relations de dominance entre les nœuds puisque cela est discuté en détail dans la Partie **2**. Constatons simplement pour l'instant que:

La structure communicative des RSém introduit une sorte de hiérarchie entre les nœuds sémantiques, hiérarchie dont l'importance est primordiale dans le fonctionnement de la composante sémantique d'un MST. Du fait de son origine, cette hiérarchie sera appelée *hiérarchie communicative*.

La structure communicative exerce en fait une sorte de "CONTRE-POUVOIR" à l'intérieur du réseau. Les exemples précédents montrent que la dépendance sémantique, introduite par Mel'čuk 1981b avec les notions de dépendances syntaxique et morphologique, cohabite, au niveau de la RSém, avec un autre type de dépendance pour former un système de dépendance mixte. Ce système comprend la dépendance sémantique actancielle (modélisée par une relation prédicat-argument) ainsi que la *dépendance communicative*, cette dernière pouvant être inverse de la première. Dans le réseau (a), 'Iznogoud' est un actant sémantique de 'croire**I.1**' mais il "vient communicativement avant" lui; c'est-à-dire que 'croire**I.1**' est communicativement présenté comme étant une propriété, une spécification, de 'Iznogoud' et non le contraire. Cela justifie les réalisations différentes en syntaxe profonde.

En résumé, deux types de dépendance existent entre les nœuds d'une RSém: (i) la *dépendance sémantique actancielle*, ou *dépendance sémantique*, et (ii) la *dépendance communicative*. Si la première dépendance est directement modélisée dans le formalisme des RSém par une relation prédicat-argument, la seconde doit être inférée, au niveau de la relation entre deux nœuds spécifiques, de la structure communicative globale d'une RSém.

A travers la dépendance communicative, on retrouve donc, en spécifiant les règles de correspondance sémantique, la notion de nœud dominant d'un réseau qui avait été introduite lors de l'examen

des règles de paraphrase sémantique. La détermination d'un nœud dominant est en fait la clé du fonctionnement des procédures d'activation de la composante sémantique. Par analogie, on peut dire que, de même qu'une RSém n'est pas complète si sa structure communicative n'est pas spécifiée, de même cette RSém n'est pas "équipée" pour la lexicalisation ou l'arborisation si un nœud dominant n'a pas été calculé.

Cela démontre que le processus de transition RSém ⇒ RSyntP doit comprendre une étape intermédiaire supplémentaire qui n'a pas été considérée jusqu'à présent. Il s'agit de la détermination d'une hiérarchie communicative complète dans les RSém:

> L'opération consistant à déterminer le nœud dominant d'une RSém puis à en déduire des relations de dépendance communicative entre tous les nœuds sémantiques, liés par une relation actancielle, sera appelée *hiérarchisation*[31].

L'opération de hiérarchisation, relevant elle aussi des aspects procéduraux de la transition RSém ⇒ RSyntP, sera décrite dans la Partie **2**.

Sur le plan du formalisme, ces considérations mettent en évidence la nécessité de spécifier, dans la partie gauche des règles de correspondance sémantique, TOUTES les relations de dépendance liant les deux nœuds majeurs considérés: la "réelle" dépendance à prendre en compte dans le cadre de la correspondance sémantique étant une résultante de la dépendance sémantique et de la dépendance communicative.

---

[31] Cette terminologie a été proposée pour la première fois dans Iordanskaja & Polguère 1988.

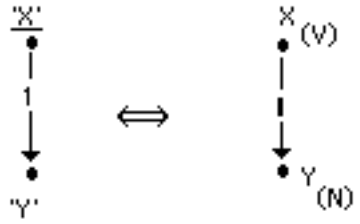
La dépendance sémantique est indiquée par l'orientation des arcs sémantiques (direction des flèches); formellement, nous spécifierons la dépendance communicative en soulignant le nœud dominant dans la partie gauche des règles de correspondance sémantique.

Nous allons examiner maintenant brièvement quelques cas concrets de règles de correspondance RSém ⇔ RSyntP.

Remarque: Ce qui a été dit plus haut pour les entrées lexicales données en exemple s'applique aussi ici pour nos règles de correspondance sémantique. Aucun travail complet n'ayant malheureusement été entrepris à ce propos, nous ne disposons même pas de règles déjà faites que l'on pourrait considérer comme valides. Nous n'avons donc ici que des illustrations à proposer. Afin d'étoffer tout de même notre présentation de ce nouveau type de règles, nous donnons dans l'Annexe de ce chapitre une ébauche d'un système de règles de correspondance sémantique du français — pour ce qui est de la mise en jeu de la relation **ATTR**. Ce système traite de façon exhaustive les cas de dépendance répertoriés, pour une telle relation, dans Apresjan *et al.* 1984 et 1985.

4.1.3 Quelques exemples

La relation de premier actant sémantique est un cas intéressant pour lequel nous allons proposer quelques ébauches de règles qui, bien sûr, sont loin de couvrir tous les cas de figure possibles.

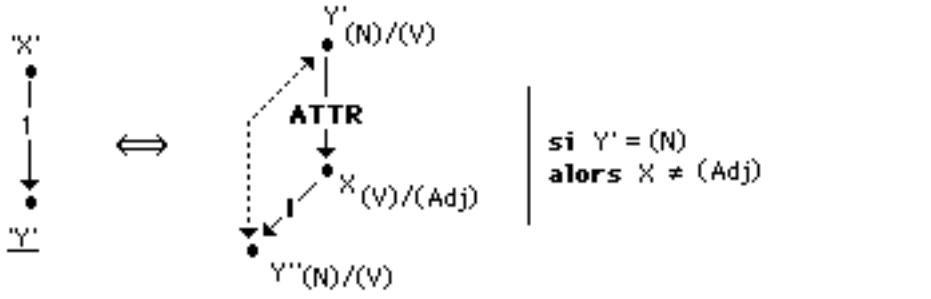


Exemple :

(1) De tremblants centimètres[Y] nous séparent[X] à peine
(Supervielle)

Figure II-11
Correspondance du type 1 pour le premier actant sémantique

Commentaires:
Cette règle très simple traite le cas où le prédicat sémantique domine communicativement son premier argument. Les conditions indiquées dans la partie droite de la règle précisent que, dans le sous-arbre syntaxique profond, le lexème réalisant le prédicat sémantique doit être un verbe et celui réalisant son premier actant doit être un nom. Il serait incohérent d'appliquer ce type de contrôle à la partie gauche de la règle puisque la notion de classe SYNTAXIQUE n'est pas pertinente au niveau de la représentation SEMANTIQUE.

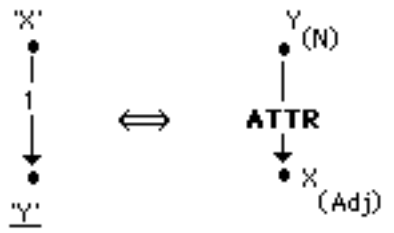


Exemples :

(1) Il pleure sans raison / Dans ce cœur[Y'] qui[Y''] s'écœure[X] (Verlaine)
(2) Tu pars[Y'], ce qui[Y''] est triste[X]

Figure II-12
Correspondance du type 2 pour le premier actant sémantique

Commentaires:
La figure II-12 présente un cas de dépendance communicative inverse. Cette nouvelle règle permet de prendre en compte, au niveau syntaxique profond, une relation attributive (**ATTR**) ainsi qu'une structure un peu complexe, avec lien de co-référence. Il est intéressant de constater que c'est uniquement la dépendance communicative inverse entre les nœuds dans la partie gauche qui conditionne l'apparition de la relation attributive. Cette dernière peut être considérée comme la marque, au niveau syntaxique profond, de la lecture qui a été faite sur le réseau à partir de son nœud dominant. Tout au long de ce chapitre, nous avons insisté sur cette notion de lecture, ou interprétation, d'un réseau sémantique en vue d'assurer l'arborisation. La structure générale de ce travail, séparation des aspects déclaratifs liés à la CORRESPONDANCE RSém ⇔ RSyntP (Partie **1**) et des aspects procéduraux liés à la TRANSITION RSém ⇒ RSyntP (Partie **2**), nous force pour l'instant à référer à cette lecture sans véritablement l'avoir décrite.



Exemple :

(1) Mourons pour des idées, d'accord, mais de mort[Y] lente[X] (Brassens)

Figure II-13
Correspondance du type 3 pour le premier actant sémantique

Commentaires:
L'adjectif en français est un mot qui est "fait" pour être dominé syntaxiquement par l'élément qu'il domine sémantiquement. La règle II-13 introduit donc le cas standard de l'adjectif en position épithète lorsqu'il est dominé communicativement au niveau sémantique. Du fait des caractéristiques propres au français, une règle du type de II-13 ne pourrait convenir lorsque l'adjectif domine communicativement son dépendant sémantique: il faudra dans ce cas établir une règle introduisant la copule (adjectif attribut, en grammaire traditionnelle; cf. la figure II-18 en 4.3.2). Il en va tout autrement pour les langues qui semblent admettre de véritables prédicats adjectivaux; par exemple, en mandarin:

tā hěn gāo
'3e pers.' 'très' 'grand'
Elle/Il est (*très*) *grande/grand.*

Malgré le caractère universel des relations sémantiques et des relations syntaxiques profondes, les règles de transition sémantique sont particulières à chaque langue. Elles constituent un système complexe que tout MST doit contenir.

Rappelons que le lecteur intéressé peut se reporter à l'Annexe du présent chapitre pour une ébauche complète des règles de correspondance RSém ⇔ RSyntP mettant en jeu la relation **ATTR**.

4.2 Le cas des relations APPEND et COORD

Dans la présentation qui vient d'être faite des règles de correspondance sémantique, nous n'avons envisagé que des règles introduisant soit les relations actancielles syntaxiques (**I**, **II**, ..., **VI**), soit la relation de modification attributive **ATTR**. Les deux autres relations universelles syntaxiques profondes devront elles aussi être introduites par ce type de règles:

APPEND, qui correspond à des constructions non parfaitement intégrées à la structure générale de l'arbre (expressions parenthétiques, apposition, etc.);

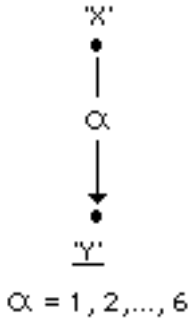
COORD, qui correspond aux divers types de coordinations (de mots, de syntagmes, de phrases, avec ou sans valeur de succession temporelle, etc.).

Nous ne pouvons décrire ici tous les cas de figure permettant, pour le français, d'introduire en syntaxe profonde de telles relations. Cela devra se faire dans le cadre de l'écriture d'une SEMANTIQUE complète de cette langue[32]. Il est toutefois possible d'observer une certaine particularité des règles introduisant les relations **APPEND** et **ATTR**:

Les relations **APPEND** et **ATTR** trouvent toutes deux leur origine dans l'examen d'une structure sémantique où la dépendance communicative est inverse de la dépendance sémantique.

[32] La "sémantique d'une langue" (par analogie avec "syntaxe de surface d'une langue") signifie ici, outre les règles de paraphrase sémantique, la collection de toutes les règles de correspondance sémantique mises en jeu par cette langue.

Cela signifie que les parties gauches des règles de correspondance sémantique introduisant **APPEND** et **ATTR** auront le profil suivant:



où, au niveau sémantique, le nœud 'X' domine sémantiquement le nœud 'Y' qui le domine communicativement.

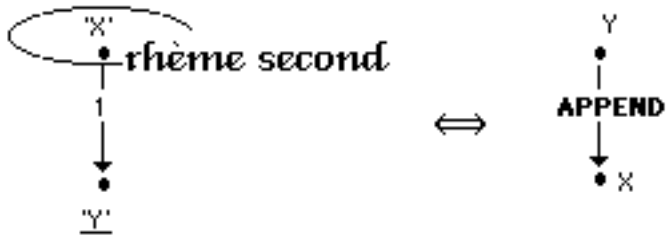
Le choix entre les deux relations peut se faire en fonction de multiples facteurs. Un cas intéressant est celui de phrases comme:

(13) *Les deux orphelines, nommées Rose et Blanche, aimaient le bon vieux Dagobert.*

(14) *Les deux orphelines qui s'appelaient Rose et Blanche aimaient le bon vieux Dagobert.*

Pour rendre compte adéquatement de la différence entre (13) et la phrase avec relative déterminative (14), il pourrait être nécessaire de considérer que (13) est en fait une phrase à rhèmes multiples, où l'information présentée par *"Rose et Blanche"* est une sorte de *rhème second*[33]. On pourrait considérer que, communicativement, (13) est la résultante de la fusion de deux messages: un message "principal" explicitant les sentiments des deux orphelines envers Dagobert, et un message "secondaire" donnant (rappelant) le nom des deux orphelines. (Dans le cas de (14), le contenu de ce dernier message fait partie intégrante du thème.) La règle prenant en charge la correspondance RSém ⇔ RSyntP avec rhème second serait alors:

[33] Nous ne faisons ici qu'effleurer cette notion de rhème second. Elle relève d'une problématique plus large, liée à la nécessité de posséder des structurations sémantique et communicative complexes, à plusieurs niveaux, pour les RSém.

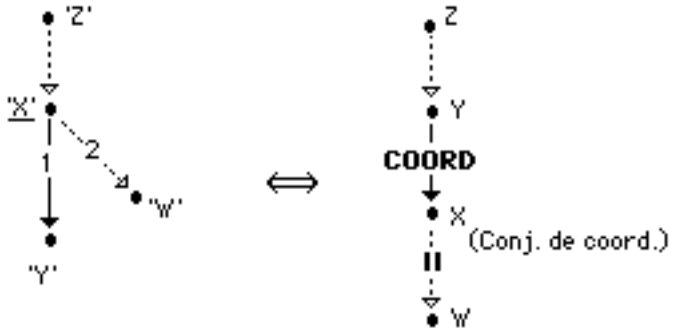


Exemple :

(1) Les deux orphelines[Y], nommées[X] Rose et Blanche, ...

Figure II-14
Exemple de règle de correspondance sémantique
impliquant la relation **APPEND**

Pour ce qui est du cas de la relation **COORD**, la correspondance s'effectue essentiellement en fonction de la classe syntaxique des lexèmes impliqués dans le sous-arbre syntaxique profond, comme l'illustre la règle de la figure II-15 qui met en jeu une conjonction de coordination.



Exemple :

(1) Au troisième chapitre, il perd[Z] à nouveau Rose[Y] et[X] Blanche[W]

Figure II-15
Exemple de règle de correspondance sémantique
impliquant la relation **COORD**[34]

Les règles introduisant **COORD** ne présentent pas nécessairement une harmonie entre dépendance sémantique et dépendance

[34] La spécification de l'argument 'W' ne sert ici qu'à faciliter la lecture de la règle.

communicative dans la partie gauche. Comme toutes les relations de dépendance syntaxique profonde, **COORD** correspond en fait à une famille de dépendances syntaxiques (explicitées au niveau syntaxique de surface) et peut être introduite à partir de configurations sémantiques très différentes. La figure II-16 donne un exemple de règle avec dépendances sémantique et communicative inverses:

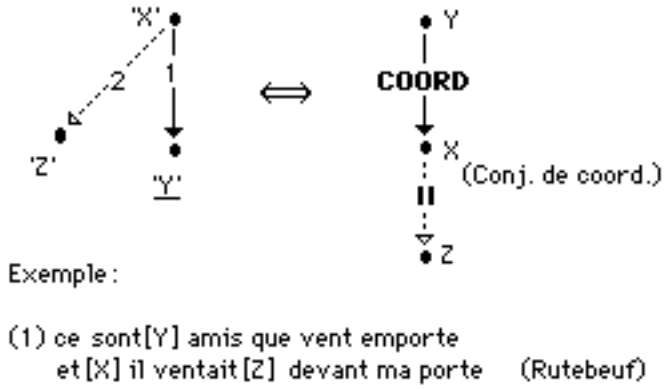


Figure II-16
Exemple de règle impliquant la relation **COORD** à partir de dépendances sémantique et communicative inverses

4.3 Introduction de lexèmes par les règles de correspondance sémantique

Les règles de correspondance sémantique peuvent introduire du matériel lexical au niveau syntaxique profond lorsque cela est nécessaire pour assurer la bonne formation de l'arbre[35]. Nous mentionnerons ici brièvement deux cas de règles de ce type portant sur:

1) la quantification des actions;
2) la construction du sommet de l'arbre syntaxique profond.

---

[35] Ceci est bien sûr la façon "synthétique" (RSém ⇒ RSyntP) de présenter de telles règles.

4.3.1 Quantification des actions

Il est possible de dire en français *deux moutons*, *quatre rêves*, etc. mais pas:

(15) **Il lut trois cette lettre avant de la comprendre.*

Une telle phrase est cependant exclue pour des raisons qui ne sont pas sémantiques. Ce n'est pas parce que LIRE**I.1** est sémantiquement un prédicat et non un objet (comme MOUTON**I.1** ou REVE**1**) qu'il ne peut être quantifié numériquement de façon directe. C'est uniquement son statut de verbe qui en est la cause. On peut ainsi paraphraser la phrase correcte (16) par (17) en utilisant le nom LECTURE**1** qui, sémantiquement, est un prédicat (équivalent à LIRE**I.1**).

(16) *Il lut trois fois cette lettre avant de la comprendre.*
(17) *Il fis trois lectures de cette lettre avant de la comprendre.*

Certaines configurations syntaxiques profondes sont donc rendues possibles uniquement par l'introduction d'éléments lexicaux. FOIS**I.1**, en (16), rend possible une certaine dépendance syntaxique profonde entre TROIS et le verbe LIRE**I.1**, en réalisant cette dépendance de façon indirecte. Ce lexème FOIS**I.1** fonctionne un peu comme les prépositions introduites dans la RSyntS à partir du régime d'un verbe ou d'un nom. C'est donc un lexème profond "outil" qui sera introduit par une règle du type:

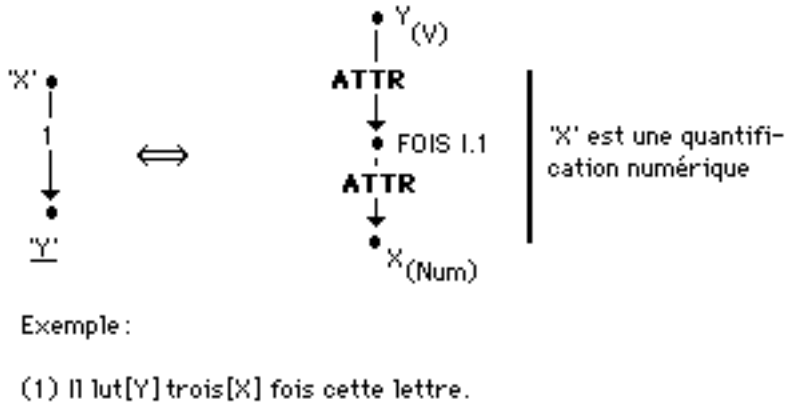


Figure II-17
Correspondance sémantique impliquant FOIS**I.1**

4.3.2 Sommet de l'arbre syntaxique profond

Un autre cas intéressant est celui des règles prenant en charge le sommet de l'arbre syntaxique profond. Un tel sommet doit être impérativement verbal sauf, bien sûr, dans le cas des Textes correspondant à des titres d'ouvrages, à certains cas de langue parlée, etc. Nous verrons aussi dans la Partie **2** que ce sommet est directement en correspondance avec le nœud dominant de la RSém considérée. Dans le cas où l'on isole comme nœud dominant un Sens qui ne peut se réaliser par un verbe, l'arbre syntaxique pourra tout de même être construit en utilisant certains moyens syntaxiques et lexicaux. Un des plus communs est la copule ETRE**1II** qui permet justement de construire un sous-arbre à sommet verbal, comme l'illustre la figure ci-dessous:

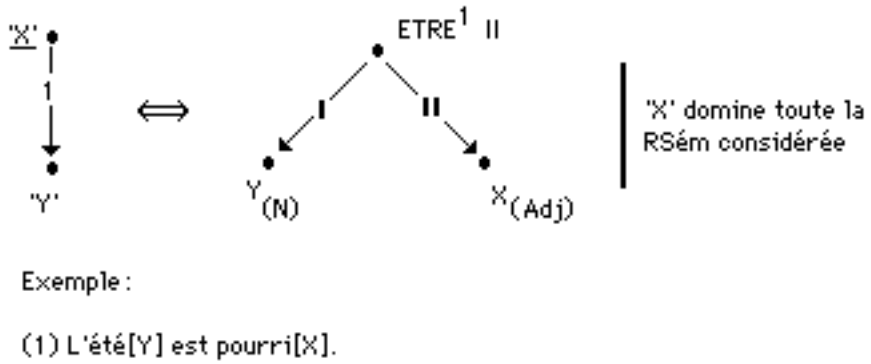


Figure II-18
Correspondance sémantique impliquant ETRE^{1}**II**

Nous avons maintenant terminé notre présentation des règles de correspondance sémantique et aussi terminé ce chapitre central sur la description du système complet de règles assurant cette correspondance.

Rappelons que nous avons ici reformulé les différents formalismes impliqués en tenant compte des impératifs de l'activation procédurale des MST. Nous avons aussi introduit les différentes opérations permettant d'effectuer la transition RSém ⇒ RSyntP et mis en lumière une nouvelle notion, centrale dans notre travail: la *dépendance communicative*.

ANNEXE DU CHAPITRE II

Ébauches de règles de correspondance sémantique mettant en jeu la relation **ATTR**

Comme l'annonce le titre de cette annexe, les règles données ici ne sont que l'ébauche des règles d'une réelle grammaire, ou partie de grammaire, du français. Nous tenons simplement à illustrer plus en détail ce que nous avons appelé, dans la Partie **1**, les *règles de correspondance sémantique proprement dites*. Nous allons utiliser un sous-ensemble du français afin de montrer qu'il est possible d'écrire de telles règles pour rendre compte de la correspondance entre certaines configurations sémantiques et certaines configurations syntaxiques profondes mettant en jeu la relation syntaxique profonde **ATTR**.

On trouve dans Apresjan *et al.* 1984 et 1985 un ensemble de règles de correspondance RSyntS ⇔ RMorphP pour le français. Bien qu'elles ne soient pas sémantiques, ces règles présentent un double intérêt pour nous:

1) Elles prennent en compte un sous-ensemble assez important du français et sont articulées autour d'exemples précis qui forment un corpus non trivial.
2) Elles sont regroupées selon les configurations syntaxiques profondes auxquelles elles correspondent.

Ce dernier aspect est particulièrement important. Nous pouvons trouver dans Apresjan *et al.* 1984, sous la rubrique "Relations syntaxiques de surface du type attributif", les configurations syntaxiques de surface qui pourraient correspondre à une configuration syntaxique profonde articulée autour de la relation **ATTR**. En conséquence, notre point de référence (c'est-à-dire notre corpus de règles et d'exemples) sera l'ensemble des relations numérotées de 17 à 33 dans Apresjan *et al.* 1984:48-51. Comme le texte sur lequel nous nous basons est d'un accès difficile, nous récapitulerons tout d'abord ces relations et les exemples qui les accompagnent — section 1. Nous présenterons ensuite — section 2 — notre ébauche des règles RSém ⇔ RSyntP correspondant au corpus de la section 1.

Notre description cherche à "coller" au maximum à celle proposée dans Apresjan *et al.* 1984 et 1985. Dans quelques cas, cependant, nous nous sommes vu obligé de ré-interpréter cette dernière afin de mieux la faire coïncider avec notre propre vision du système de dépendance syntaxique impliqué.

1 CORPUS TIRE DE APRESJAN *ET AL.* 1984

Les relations syntaxiques de surface ci-dessous sont énoncées dans l'ordre original avec la numérotation de Apresjan *et al.* 1984[1]. Les variables X et Y, dans les exemples, servent à indiquer les deux lexèmes unis par la relation syntaxique considérée. X renvoie au lexème dominant et Y au lexème dominé.

<u>17 - Relation "épithétique"</u>:

(17.1) *le premier* [Y] *homme* [X]
(17.2) *quelques propriétés* [X] *mécaniques* [Y]

---

[1] Nous nous sommes permis de modifier légèrement certains exemples dont le français ne nous semblait pas tout à fait correct.

(17.3) *l'ionisation* [X] *induite* [Y] *par le soleil*

18 - Relation "descriptive-épithétique":

(18.1) *Très fiables* [Y], *ces entraînements* [X] *sont utilisés dans les systèmes les plus complexes.*

19 - Relation "déterminative":

(19.1) *l'* [Y] *atelier* [X]
(19.2) *cet* [Y] *entraînement* [X]
(19.3) *chaque* [Y] *jour* [X]

20 - Relation "relative-restrictive":

(20.1) *l'étoile* [X] *dont la parallaxe a* [Y] *été mesurée*
(20.2) *plusieurs systèmes* [X] *ont été construits qui ont* [Y] *facilité l'analyse des données*
(20.3) *Il cherchait la main* [X] *à quoi s'accrocher* [Y].

21 - Relation "relative-qualificative":

(21.1) *cette étoile* [X], *dont nous connaissons* [Y] *la parallaxe*

22 - Relation "attributive":

(22.1) *la maison* [X] *à* [Y] *trois étages*
(22.2) *les isolants* [X] *en* [Y] *plastique*
(22.3) *l'astre dont* [Y] *nous voyons la faible lueur* [X]
(22.4) *rouge* [X] *de* [Y] *colère*
(22.5) *environ* [Y] *trente* [X] *personnes*

Remarque 1:

La relation de surface mise en jeu dans (22.5) serait peut-être mieux décrite en tant que "restrictive" plutôt qu'"attributive" (cf. Mel'čuk & Pertsov 1987:314, où la relation "restrictive" est appliquée à l'équivalent anglais de ce cas de figure). Dans (22.6), ci-dessous, nous associons la variable X à *tard* et non à *plus*, comme c'est le cas dans l'exemple original.

(22.6) *une heure* [Y] *plus tard* [X]
(22.7) *théoriquement* [Y] *possible* [X]

(22.8) *le gamin* [X] *lisant* [Y] *un journal d'aujourd'hui*

23 - Relation "quantitative":

(23.1) *trente* [Y] *kilomètres* [X]
(23.2) *une* [Y] *heure* [X] *et demie*

24 - Relation "appositive":

(24.1) *Paris* [X], *capitale* [Y] *de la France*
(24.2) *cette impulsion*[X], *cas*[Y] *important d'une impulsion rectangulaire*
(24.3) *Victor* [X] *Hugo* [Y]

Remarque 2:

Peut-être "*Victor* [Y] *Hugo* [X]" serait-il plus approprié ici.

25 - Relation "numérique-appositive":

(25.1) *paragraphe* [X] *vingt* [Y]
(25.2) *les années* [X] *soixante* [Y]
(25.3) *numéro* [X] *15* [Y]

26 - Relation "compositive-appositive":

(26.1) *moteur* [X]-*alternateur* [Y]
(26.2) *transformateurs* [X]-*redresseurs* [Y]
(26.supl.) *aide* [X]-*cuisinier* [Y]

Remarque 3:

Les exemples (26.1) et (26.2) ne seront pas pris en compte dans l'ensemble de règles sémantiques donné ci-dessous. Nous ne pensons pas que la structure syntaxique interne de ces termes techniques doive être décrite par les règles générales de la grammaire. Dans la mesure où ce sont des idiomes, leur structure syntaxique sera présentée directement dans le lexique. (Cette remarque s'applique également aux exemples (27.1) et (27.2) ci-dessous: *vecteurs espace* et *ingénieur système*.) Nous avons ajouté ici un exemple supplémentaire qui n'est pas idiomatique et qui pourra être traité par une de nos règles sémantiques (règle VI de la section 2).

27 - Relation "compositive":

(27.1) *vecteurs* [X] *espace* [Y]
(27.2) *l'ingénieur* [X] *système* [Y]
(27.3) *des cheveux* [X] *châtain* [Y] *clair*

28 - Relation "proleptique":

(28.1) *Ce livre* [Y], *je l'* [X] *ai déjà lu.*
(28.2) *Ta femme* [Y], *est-elle* [X] *jeune?*
(28.3) *Lui* [Y], *il* [X] *n'est pas encore venu.*
(28.4) *D'où viennent-ils* [X], *ces deux* [Y]-*là?*

Remarque 4:

Nous considérerons provisoirement que de telles extrapositions sont basées sur une relation **ATTR** produite, au niveau de la RSyntP elle-même, par "dislocation" d'un nœud de l'arbre syntaxique — ce qui permet la réalisation d'un thème emphatique. La relation **ATTR** étant ainsi introduite par UNE REGLE DE TRANSFORMATION DE L'ARBRE SYNTAXIQUE PROFOND, nous ne proposons pas de règles sémantiques particulières traitant ces exemples.

29 - Relation "circonstancielle":

(29.1) *ils peuvent être définitivement* [Y] *établis* [X]
(29.2) *écrire* [X] *en* [Y] *français*

Remarque 5:

(29.2) correspond à une configuration actancielle syntaxique profonde, provenant elle-même d'une configuration actancielle sémantique: 'X écrit**I.1** Z **dans la langue Y**'. Il n'y a donc pas lieu de considérer ici cet exemple.

(29.3) *Il travaille* [X] *du* [Y'] *matin au* [Y''] *soir.*
(29.4) *Si* [Y] *on est prudent, on obtiendra* [X] *ce résultat.*
(29.5) *Ne pouvant* [Y] *sortir de ce bois, nous y avons* [X] *campé.*
(29.6) *Le dimanche* [Y], *le magasin est* [X] *fermé.*

30 - Relation "durative":

(30.1) *il travaillait* [X] *toute la nuit* [Y]

31 - Relation "parenthétique":

(31.1) *Il est* [X] *évidemment* [Y] *déraisonnable de faire cela.*
(31.2) *Comme* [Y] *nous l'avons déjà dit, ce matériau possède* [X] *une bonne stabilité.*

32 - Relation "délimitative":

(32.1) *Il ne* [Y_1] *marche* [X] *pas* [Y_2].
(32.2) *on n'entend que* [Y] *des cris* [X]
(32.3) *il y a seulement* [Y] *les alliages* [X]

33 - Relation "conclusive":

(33.1) *Il peut* [X] *être facilement mis en forme, ce* [Y] *qui nous permet de l'utiliser dans l'installation.*

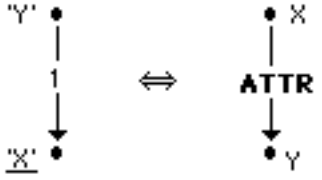
2 REGLES RSEM ⇔ RSYNTP METTANT EN JEU LA RELATION SYNTAXIQUE PROFONDE ATTR

Les règles données ici sont numérotées en chiffres romains de I à VIII. Elles sont toutes accompagnées d'exemples dont la numérotation donne accès aux relations syntaxiques de surface de la section 1 et aux exemples correspondants. Ainsi, la phrase numérotée (II.29.3) pour la règle II, correspond au troisième exemple de la relation syntaxique de surface 29. Le lecteur est invité à consulter les deux sections simultanément au moyen de cette numérotation. Soulignons toutefois que le lien entre les règles présentées ici et les configurations introduites en 1 est nécessairement indirect. Il faut tenir compte de la composante syntaxique profonde Sens-Texte qui prend en charge la correspondance intermédiaire RSyntP ⇔ RSyntS: c'est-à-dire, dans le cas présent, la correspondance entre les

dépendances **ATTR**ibutives et les différentes relations syntaxiques de surface qui peuvent lui être associées. Nous référons, dans les remarques, à ce niveau de correspondance mais il n'est pas directement considéré dans cette Annexe.

Rappel: Le nœud souligné dans la partie gauche des règles de correspondance sémantique est le nœud communicativement dominant. On peut vérifier ici que toutes les configurations sémantiques correspondant à des configurations **ATTR**ibutives sont fondées sur des dominances sémantique et communicative inverses. (Cette caractéristique est étudiée en II:4.1.)

Règle I:



X = N
Y = Adj/Adj_{Num}/Art/Prép/Adv[2]
ou
X = V/Adj/Adj_{Num}
Y = Adv/Prép
ou
X = V
Y = Prép/Conj

(I.17.1) *le premier* [Y] *homme* [X]
(I.17.2) *quelques propriétés* [X] *mécaniques* [Y]
(I.18.1) *très fiables* [Y], *ces entraînements* [X] *sont...*

Remarque 6:

Dans (I.18.1), la relation de dépendance mise en jeu au niveau syntaxique de surface ("descriptive-épithétique") est le reflet d'une certaine intention communicative. On peut considérer que, pour un tel exemple, 'fiable' est spécifié comme élément d'un sous-réseau communicatif particulier (**rhème second**, par exemple). Les règles de correspondance sémantique doivent permettre de mettre en correspondance la structure communicative de la RSém et celle de la RSyntP (cf. Chapitre VI). Les règles de correspondance syntaxique profonde tiennent compte de cette information de type communicatif

[2] Voir Remarque 10.

pour spécifier la correspondance entre **ATTR** et les différentes dépendances syntaxiques de surface qui peuvent lui être associées.

(I.19.1) *l'* [Y] *atelier* [X]

Remarque 7:

Dans le cas où l'article français est considéré comme un grammème du nom, il n'apparaît pas comme un dépendant **ATTR** du nom au niveau syntaxique profond mais comme un marqueur grammatical sur les nœuds de la RSyntP. Il est introduit en tant que nœud lexical séparé au niveau syntaxique de surface (voir la discussion de ce problème en II:3.2).

(I.19.2) *cet* [Y] *entraînement* [X]
(I.19.3) *chaque* [Y] *jour* [X]
(I.22.1) *la maison* [X] *à* [Y] *trois étages*
(I.22.2) *les isolants* [X] *en* [Y] *plastique*

Remarque 8:

Nous postulons ici que les prépositions pleines comme À**IV.2** et EN**I.2** doivent être sémantiquement des prédicats à deux arguments dont le premier est la chose caractérisée et le second la valeur de la caractérisation. Bien sûr, nous ne nous intéressons pas ici au Sens véritable (c'est-à-dire, à la définition) de ces lexèmes.

(I.22.5) *environ* [Y] *trente* [X] *personnes*
(I.22.7) *théoriquement* [Y] *possible* [X]
(I.23.1) *trente* [Y] *kilomètres* [X]
(I.23.2) *une* [Y] *heure* [X] *et demie*

Remarque 9:

Contrairement à ce qui est impliqué ici, nous ne sommes pas du tout convaincu que des expressions quantitatives telles que *trente kilomètres* doivent se voir associer une RSém de même structure que des expressions telles que *les quatre modernisations*. Nous serions plutôt tenté d'établir, du point de vue de la RSém, une nette distinction entre (i) les expressions quantitatives mettant en jeu de simples noms communs déterminés par un adjectif numéral et (ii) les expressions quantitatives mettant en jeu des unités de mesure. Dans ce dernier cas, l'unité de mesure pourrait être, au niveau sémantique, un prédicat spécifiant la "qualité" de la chose mesurée (*trente kilomètres* se paraphrasant ainsi par quelque chose comme *une*

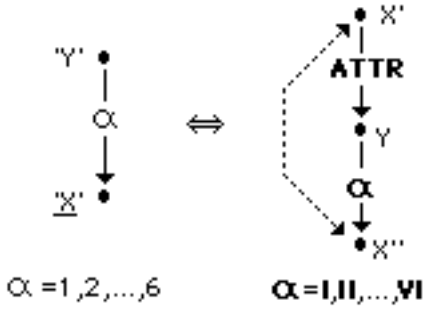
distance de trente mesurée en kilomètres plutôt que par *trente fois un kilomètre*).

(I.29.1) *ils peuvent être définitivement* [Y] *établis* [X]
(I.29.3) *il travaille* [X] (*durant* [Y]) *du matin au soir*
(I.29.4) *Si* [Y] *on est prudent, on obtiendra* [X] *ce résultat.*
(I.29.5) (*parce que* [Y]) *ne pouvant sortir du bois, nous y avons campé* [X]
(I.29.6) (*durant* [Y]) *le dimanche, le magasin est fermé* [X]
(I.30.1) *il travaillait* [X] (*durant* [Y]) *toute la nuit*
(I.31.1) *Il est évidemment* [Y] *déraisonnable* [X] *de faire cela.*
(I.31.2) *Comme* [Y] *nous l'avons déjà dit, ce matériau possède* [X] *une bonne stabilité.*
(I.32.1) *Il ne* [Y_1] *marche* [X] *pas* [Y_2].
(I.32.2) *on n'entend que* [Y] *des cris* [X]
(I.32.3) *il y a seulement* [Y] *les alliages* [X]

Remarque 10:

Nous avons adopté le traitement de l'exemple (I.32.3) proposé par Apresjan *et al.* 1984[3] même s'il est douteux que la dépendance syntaxique unisse ici le nom et l'adverbe. Il nous semblerait plus approprié de rattacher ce dernier directement au sommet de l'arbre syntaxique: *il y a* [X] *seulement* [Y]...

Règle II:



X', X'' = N
Y = V

(II.17.3) *l'ionisation* [X'] *induite* [Y] *par le soleil*
(II.20.2) *plusieurs systèmes* [X'] *ont été construits qui* [X''] *ont facilité* [Y] *l'analyse des données*
(II.20.3) *Il cherchait la main* [X] *à quoi* [X''] *s'accrocher* [Y].
(II.22.8) *le gamin* [X] *lisant* [Y] *un journal d'aujourd'hui*

Remarque 11:

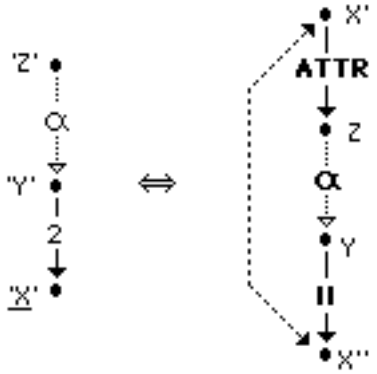
La correspondance entre a et **a** signifie qu'à la relation actancielle sémantique **1** correspond la relation actancielle syntaxique profonde **I**,

---

[3] Le même traitement a été proposé pour ce type d'exemple dans Mel'čuk 1974:226.

qu'à la relation **2** correspond la relation **II**, etc. Remarquons que le composant X'' n'a pas nécessairement de réalisation de surface (cf. les constructions participiales).

Règle III:



X', X'' = N
Y = N
Z = V

Remarque 12:

Une telle règle ne fonctionne pas de façon autonome. Elle fait appel à une autre règle de correspondance sémantique pour établir la correspondance entre a et **a**. Dans un cadre de transition RSém › RSyntP, la relation sémantique entre 'Z' et 'Y' peut être interprétée comme la spécification du "contexte sémantique" d'application de la règle; la relation de dépendance syntaxique en pointillé (dans la partie droite) peut, elle, être interprétée comme un appel récursif à la grammaire.

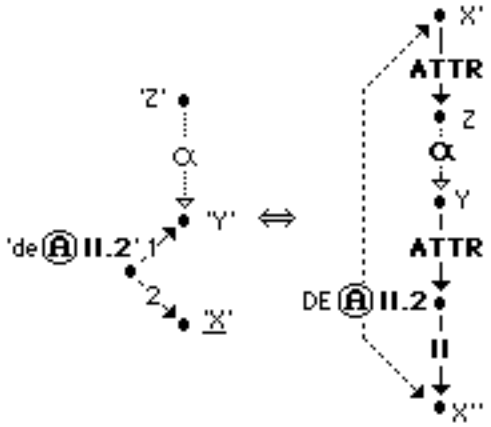
(III.20.1) *l'étoile* [X'] *dont* [X''] *la parallaxe* [Y] *a été mesurée* [Z]
(III.21.1) *cette étoile* [X'], *dont* [X''] *nous connaissons* [Z] *la parallaxe* [Y]

Remarque 13:

Notre but n'étant pas de décrire la correspondance RSyntP ⇔ RSyntS, nous ne tenons pas compte de l'information communicative qui doit être véhiculée pour établir une distinction entre les relations syntaxiques de surface "relative-restrictive" *vs.* "relative-qualificative" qui sont mises en jeu dans ces deux exemples (cf. Remarque 6).

PARALLAXE est considéré sémantiquement dans (III.21.1) comme un prédicat à deux arguments. Le premier argument correspond à la valeur de cette mesure qu'est la parallaxe; le second argument correspond à l'objet dont la parallaxe est donnée.

Règle IV:



X', X'' = N
Y = N
Z = V

Remarque 14:

À la différence de À**IV.2** et EN**I.2** (cf. Remarque 8), le Sens du lexème DEⒶ**II.2** est considéré ici comme un prédicat à deux arguments dont le premier est la chose déterminée sémantiquement, ou qualifiée, et le second la détermination sémantique, ou la qualité. Il existe en fait plusieurs prépositions qui, comme DEⒶ**II.2**, pourraient apparaître ici. Dans une description grammaticale et lexicale plus complète, on sera amené à regrouper les prépositions pleines du français en sous-classes — selon les règles syntaxiques qui les mettent en jeu.

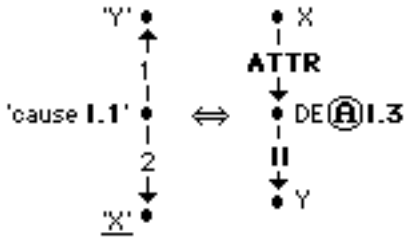
(IV.22.3) *l'astre* [X'], *dont* [DEⒶ**II.2 0** X''] *nous voyons* [Z] *la faible lueur* [Y]

(IV.supl.) *la pièce* [X'], *dont* [DEⒶ**II.2 0** X''] *Iznogoud cassait* [Z] *toutes les chaises* [Y]

Remarque 15:

Nous proposons un exemple supplémentaire pour cette règle car LUEUR**1**, en (IV.22.3), devrait peut-être être considéré sémantiquement comme un prédicat prenant pour actant la "source" de la lueur (ici, le Sens de ASTRE**1**).

Règle V:



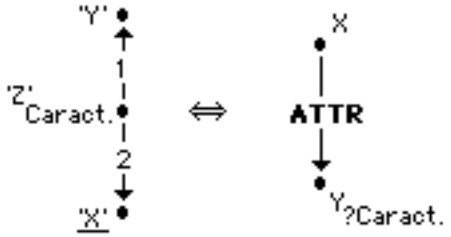
X = Adj/V/N
Y = N/V

(V.22.4) *rouge* [X] *de colère* [Y]

Remarque 16:

On peut citer comme autres exemples de ce type de construction, mettant en jeu le lexème DE(A)**I.3**, *malade de trop avoir mangé*, *mourir de faim*, *trois décès de maladie*, etc. Cependant, comme pour *rouge* dans (V.22.4), il est possible que ces constructions contiennent en réalité sémantiquement des prédicats à deux arguments: 'W être rouge du fait de Y', etc.

<u>Règle VI</u>:



X = Adv/N
Y = N/Adj_{Num}/V[4]

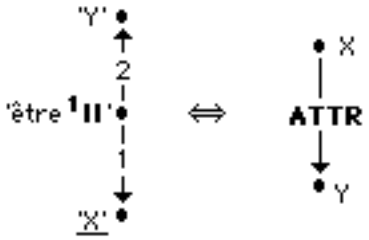
Remarque 17:

Le trait "Caract." ("caractéristique") sert à identifier une classe de Sens de lexème. Certains de ces Sens figurent entre parenthèses dans les exemples ci-dessous. "?Caract." indique qu'il est vraisemblablement nécessaire de transmettre le trait "Caract." à Y — en tant que trait contextuel dans la RSyntP — pour permettre le calcul de la relation syntaxique de surface appropriée. Il est intéressant de voir ici comment le recours à une relation très générale comme **ATTR** permet de "mettre à plat" certaines nuances, comme celle exprimée par le Sens 'Z'$_{Caract.}$, qui ne sont pas nécessairement pertinentes pour les niveaux plus superficiels de représentation.

(VI.22.6)	*une heure* [Y] ('quantité**1**' ['Z']) *plus tard* [X]
(VI.24.3)	*Victor* [X] ('patronyme' ['Z']) *Hugo* ['Y']
(VI.25.1)	*paragraphe* [X] ('numéro**1**' ['Z']) *vingt* [Y]
(VI.25.2)	*les années* [X] ('numéro**1**' ['Z']) *soixante* [Y]
(VI.26.supl.)	*aide* [X]- ('profession**II.1**') *cuisinier* [Y]
(VI.27.3)	*cheveux* [X] ('couleur**I.1**' ['Z']) *châtain* [Y] *clair*

[4] Pour une application de cette règle avec un Y verbal, voir Chapitre VI (Partie **2**).

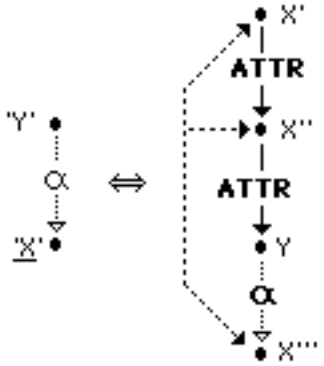
Règle VII:



X = N
Y = N

(VII.24.1) *Paris* [X], ('être**1II**' ['Z']) *capitale* [Y] *de la France*
(VII.24.2) *cette impulsion* [X], ('être**1II**' ['Z']) *cas* [Y] *important d'une impulsion rectangulaire*
(VII.25.3) *numéro* [X] ('être**1II**' ['Z']) *15* [Y]

Règle VIII:



X', X'', X''' = V
Y = V

(VIII.33.1) *Il peut* [X'] *être facilement mis en forme, ce* [X''] *qui* [X'''] *nous permet* [Y] *de l'utiliser dans l'installation.*

Remarque 18:

Cette règle fait un appel récursif à une autre règle afin de déterminer la correspondance a-**a**.

Partie 2

La Transition Sémantique

CHAPITRE III

Principes généraux de transition sémantique

Nous entamons maintenant la seconde partie de ce travail qui s'attache à décrire la TRANSITION RSém ⇒ RSyntP, c'est-à-dire l'activation procédurale de la composante sémantique Sens-Texte dans un cadre de synthèse.

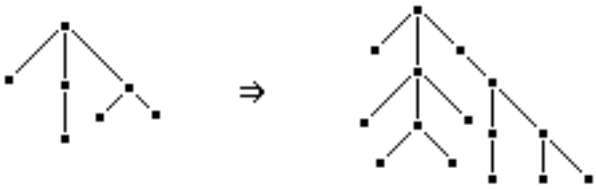
Le présent chapitre se divise en trois sections:

1) présentation très générale de la transition sémantique, par comparaison, notamment, avec les autres transitions contenues dans une synthèse linguistique Sens-Texte;
2) récapitulation des différentes opérations de transition impliquées par la composante sémantique d'un MST — ces opérations ont toutes déjà été évoquées dans le cours de l'exposé et s'appuient sur les règles décrites de façon détaillée dans la Partie **1**;
3) introduction de la méthode générale proposée qui doit permettre de mettre en œuvre procéduralement la transition sémantique tout en s'appuyant sur des notions ayant une pertinence linguistique.

1 EXPOSE DU PROBLEME A TRAITER

Si l'on considère les différentes représentations auxquelles fait appel un MST D'UN POINT DE VUE STRICTEMENT FORMEL, en faisant abstraction des données mêmes qu'elles véhiculent, la transition RSém ⇒ RSyntP semble, de toutes les transitions impliquées par le modèle, la plus difficile à opérer. Pour justifier cette thèse, passons tout d'abord brièvement en revue les autres transitions impliquées par le modèle. Chaque mention d'un type donné de transition sera suivie d'un petit diagramme symbolisant le problème formel posé par cette transition.

RSyntP ⇒ **RSyntS**:



Il s'agit de transformer un arbre de dépendance en un autre arbre de dépendance. L'ensemble des relations unissant les nœuds de la représentation cible (40 proposées pour l'anglais dans SSE et l'on peut s'attendre à un nombre sensiblement équivalent pour le français) est potentiellement plus grand que celui des relations de syntaxe profonde (un nombre universel de 9: **I**, **II**, **III**, **IV**, **V**, **VI**, **ATTR**, **COORD** et **APPEND**). Ce type de transition ne pose pas de difficulté majeure dans la mesure où les règles statiques de la grammaire effectueront l'essentiel de cette tâche. Il n'existe cependant pas pour le français d'exemple de telles règles bien formalisées et publiées. Seul le russe a été décrit à ce niveau (cf. Mel'čuk 1974); une description très partielle de l'anglais a été faite dans le cadre du projet GOSSIP de même qu'une description du sous-langage de la météorologie marine (français et anglais) dans le cadre du projet FoG (cf. Bourbeau *et al.* 1989). À fin d'illustration, la figure III-1 ci-dessous donne un exemple de règle syntaxique de correspondance RSyntP ⇔ RSyntS pour le français.

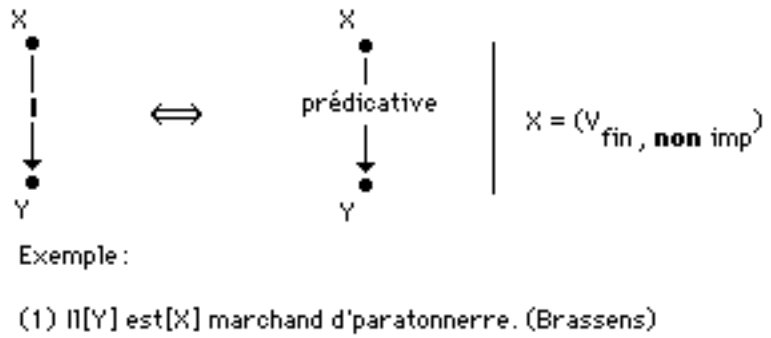
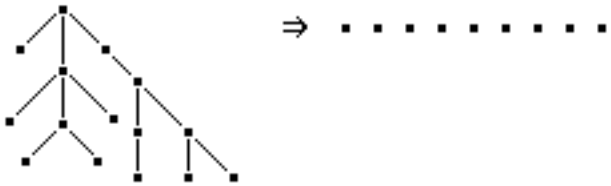


Figure III-1
Exemple de règle syntaxique RSyntP ⇔ RSyntS

RSyntS ⇒ **RMorphP**: 

Il s'agit d'effectuer la linéarisation de l'arbre syntaxique de surface en spécifiant, pour chaque élément de la représentation créée, les valeurs de toutes les variables morphologiques. L'opération en cause ici est en partie équivalente à un ordonnancement des branches des arbres syntaxiques et permet d'introduire la représentation de la linéarité de la phrase (qui, rappelons-le, fait déjà partie intégrante des arbres syntaxiques dans les modèles génératifs). Cependant, les éléments de la représentation produite ne portent ici que de l'information intéressant les niveaux morphologiques. Là encore, les règles statiques du modèle (telles que décrites pour l'anglais dans SSE) joueront un rôle central dans l'opération de transition. Il faut tout de même remarquer que des méta-règles traitant des alternatives dans l'ordre des syntagmes, des phénomènes de non-projectivité, de l'influence de la structure communicative, etc. vont être utilisées. Cette transition est nécessairement plus difficile à implémenter que la précédente puisqu'elle met en correspondance deux représentations formellement très différentes (un arbre de dépendance et une chaîne linéaire).

Remarque: Nous ne nous intéressons ici au problème de la transition que du point de vue des différentes représentations de la phrase, sans prendre en considération les représentations intermédiaires des mots-formes qui la composent. Ainsi, dans la RMorphP, si la structure considérée pour la représentation de la phrase est parfaitement linéaire, les noms de LEXEMES COMPOSES peuvent prendre la forme d'un "mini" arbre de dépendance. Un tel type de représentation est nécessaire pour les lexèmes composés de certaines langues qui utilisent la composition de façon très régulière: la structuration syntaxique interne du lexème composé y aura une incidence sur le calcul de sa RMorphS ainsi que sur son interaction avec les autres lexèmes de la phrase (voir Mel'čuk à paraître:Sixième Partie, Chapitre I, §2).

RMorphP ⇒ **RMorphS**, ▪ ▪ ▪ ▪ ▪ ▪ ▪ ⇒ ▪ ▪ ▪ ▪ ▪ ▪ ▪
RMorphS ⇒ **RPhonP** et
RPhonP ⇒ **RPhonS**:

Ces transitions ne posent pas de problèmes FORMELS majeurs puisqu'elles concernent toutes des chaînes linéaires. Nous n'avons donc ici aucune remarque spéciale à faire à leur sujet.

Comparée à toutes les transitions qui viennent d'être brièvement caractérisées, la transition RSém ⇒ RSyntP présente certains aspects originaux qui la rendent particulièrement difficile à implémenter. Nous avons examiné, dans la partie précédente (I:4.2), l'écart formel séparant les RSém et les RSyntP. En mettant au point des procédures rigoureuses pour implémenter la transition sémantique, nous avons cherché, en quelque sorte, à réduire cet écart. Il semble en effet que la transition sémantique, pour être bien comprise, doive passer par une étape intermédiaire où l'écart formel entre réseaux sémantiques et arbres syntaxiques profonds est réduit au maximum. Nous avons pu constater que, lorsque nous examinions un réseau sémantique pour lui associer un arbre syntaxique profond,

nous cherchions tout d'abord à faire une certaine "interprétation" du réseau. Une telle interprétation nous permettait alors de faire jouer un ensemble fini de règles élémentaires de transition.

Soulignons que la première implémentation de la transition sémantique, cf. Boyer 1984:26, se basait sur un traitement très différent des réseaux sémantiques:

> "La construction de la structure syntaxique profonde se fait en utilisant un dictionnaire mettant en correspondance des réseaux appelés à être sous-réseaux d'une structure sémantique et des schémas d'arbre qui, fusionnés, formeront la structure syntaxique profonde."

Il s'agissait d'utiliser les définitions de lexèmes pour engendrer, à partir du réseau, des sous-arbres syntaxiques profonds que l'on connectait ensuite les uns aux autres. Dans le cas où l'on doit traiter des réseaux très complexes à l'aide de lexiques consistants, ce type d'approche de la transition sémantique peut poser des problèmes d'explosion combinatoire[1]. De plus, cette première implémentation de la transition sémantique fait prendre en charge la presque totalité du processus de construction de la RSyntP par les règles lexicales, or, nous avons vu (Chapitre II) que nous avons des raisons théoriques pour ne pas adopter cette approche.

<u>Remarque</u>: Un problème particulier étudié, au niveau de l'implémentation, dans Boyer 1984 est celui de la mise en jeu des *fonctions lexicales* pour effectuer des transformations sur un arbre syntaxique "de première main". Nous verrons que notre approche ne s'attaque justement pas à ce problème. Nous n'apportons ici aucune lumière nouvelle quant à la mise en jeu des *fonctions lexicales* dans le cadre de la transition sémantique. Sur la manipulation automatique des *fonctions lexicales*, on pourra consulter, en plus de Boyer 1984 et Boyer &

---

[1] L'explosion combinatoire se produit lorsqu'un algorithme génère, durant son exécution, un nombre de solutions au problème traité qui croit exponentiellement.

Lapalme 1985, les propositions faites dans Décary 1986 à propos de l'implémentation d'un éditeur du DEC.

2 OPERATIONS DE TRANSITION IMPLIQUEES PAR LA COMPOSANTE SEMANTIQUE D'UN MST

2.1 Répertoire des opérations mises en jeu

Les opérations répertoriées ici ne tiennent pas compte de la mise en jeu des *règles de paraphrase lexicale* et de *paraphrase syntaxique*. Ces règles permettent d'effectuer des transformations sur des configurations d'arbres syntaxiques profonds qui:

- soit, préservent le Sens exprimé par ces configurations, par exemple:
  "Le fakir a vendu un tapis à Iznogoud" ≡ *"Iznogoud a acheté un tapis au fakir"*;
- soit, rendent compte de certaines implications sémantiques:
  "Telesphore a augmenté la vitesse de son avion" ⊢ *"La vitesse de l'avion a augmenté"*.

Les règles de ce type, qui font pourtant partie de la composante sémantique d'un MST, n'ont pas été considérées ici pour les deux raisons suivantes:

1) Dans le cadre d'un processus de synthèse linguistique, où l'on se focalise sur la transition RSém ⇒ RSyntP, les opérations effectuées par ces règles sont plutôt considérées comme des transformations d'un arbre syntaxique profond initial obtenu à partir d'une RSém donnée.
2) Cet aspect de la composante sémantique, bien que méritant une recherche en soi, peut être artificiellement laissé de côté sans que cela n'entrave la progression de notre exposé.

On trouvera dans Mel'čuk 1988d une présentation du système de paraphrase d'un MST. Pour notre part, nous avons considéré ici que les règles correspondant aux définitions de lexèmes étaient des règles de paraphrase sémantique; ce sont les seules règles de paraphrase que nous prenons en compte dans notre étude de la transition RSém ⇒ RSyntP.

Nous avons isolé dans la Partie **1**, lors de l'examen des règles sémantiques, quatre opérations constituant la transition sémantique: la *lexicalisation (au sens propre)*, la *grammaticalisation*, l'*arborisation* et la *hiérarchisation*.

Dans ce qui suit, nous récapitulons ces opérations en les considérant du point de vue de la théorie, par rapport aux règles Sens-Texte qu'elles manipulent. Ceci signifie qu'il n'est pas question ici des contraintes qui peuvent découler d'une implémentation véritable. Par exemple, lorsque nous disons que telle opération met en jeu tel niveau de représentation, nous envisageons un traitement strictement séquentiel de la transition sémantique. Dans la pratique cependant (c'est-à-dire dans le cadre du développement d'un véritable système de génération de texte "intelligent") ces opérations de transition devraient peut-être être considérées en parallèle[2].

2.1.1 La *lexicalisation (au sens propre)*

Cette opération fonctionne au seul niveau de la RSém. Elle est basée sur l'activation des *règles de paraphrase sémantique* (présentées en II:2), avant tout pour substituer un Sens de lexème (appelé aussi "sémantème") à un sous-réseau sémantique. La lexicalisation vise donc à sélectionner les Sens qui seront directement exprimés par des lexèmes au niveau syntaxique profond. En quelque sorte, elle permet de définir la profondeur de la décomposition sémantique dont rend

---

[2] On peut noter que ces opérations sont effectuées dans le système GOSSIP de façon non strictement séquentielle; ainsi, il n'existe pas d'étape de hiérarchisation puisque cette opération se trouve imbriquée dans la lexicalisation, la grammaticalisation et l'arborisation.

compte la RSyntP (et, par voie de conséquence, le Texte final). Rappelons (cf. le principe énoncé en II:2.2) qu'un réseau sémantique devrait être réduit au maximum pour permettre la production de phrases stylistiquement élégantes. On peut d'ailleurs considérer que si la lexicalisation fonctionne "à l'envers", c'est-à-dire en décomposant les Sens plutôt qu'en les réduisant, les règles sémantiques utilisées permettent alors d'établir des paraphrases valides d'un point de vue strictement théorique et qui ne sont pas fonctionnelles lorsque l'on envisage le niveau du Texte lui-même.

Tel qu'annoncé dans l'Introduction, nous ne consacrons pas de chapitre particulier à cette opération très complexe dont l'étude systématique reste encore à faire. On trouvera dans cette partie des remarques sur la lexicalisation qui complètent les observations faites dans la Partie **1**.

2.1.2 La *grammaticalisation*

Cette opération met en jeu les niveaux sémantique et syntaxique profond en prenant en charge l'activation des *règles grammatico-sémantiques* (présentées en II:3). Elle présuppose le même type de traitement de la RSém que l'opération de lexicalisation: c'est-à-dire, une analyse du réseau en vue d'identifier certains sous-réseaux particuliers. La grammaticalisation, cependant, présente les deux caractéristiques suivantes:

1) Elle se base sur un ensemble restreint de règles grammatico-sémantiques puisque, pour une langue donnée, il existe très peu de significations grammaticalisables (comparé au très grand nombre de significations directement exprimables par un lexème profond).
2) Les règles grammatico-sémantiques mettant en correspondance le niveau sémantique et le niveau syntaxique profond, la grammaticalisation, contrairement à la lexicalisation au sens propre, met en jeu un niveau de représentation en construction.

La première de ces caractéristiques montre qu'il sera plus facile d'organiser méthodiquement la grammaticalisation que la lexicalisation, vu le petit nombre de règles à mettre en jeu[3]. Néanmoins, la seconde caractéristique montre que la grammaticalisation est une opération conceptuellement plus "sophistiquée" puisqu'elle joue sur deux niveaux de représentation simultanément.

Malgré ces différences, nous avons en général, dans nos recherches, considéré ensemble la lexicalisation et la grammaticalisation; cela explique que cette dernière opération, comme la première, ne soit pas étudiée ici en détail au niveau des procédures de transition sémantique.

2.1.3 L'*arborisation*

Cette opération met en jeu les niveaux sémantique et syntaxique profond de représentation par l'activation des *règles de correspondance sémantique proprement dites* (présentées en II:4). L'arborisation réalise véritablement la construction de la RSyntP. C'est une sorte d'opération de TRADUCTION qu'il faut décrire comme telle au niveau des micro-structures mises en jeu, c'est-à-dire les sous-réseaux sémantiques et les sous-arbres syntaxiques profonds apparaissant dans les règles de correspondance[4]. Rappelons que ce travail introduit de véritables règles de correspondance sémantique; jusqu'à présent, les MST permettaient que cette correspondance soit prise en charge par les définitions lexicales elles-mêmes. Nous avons montré en II:2.1 que les règles de correspondance sémantique devaient être spécifiées séparément des entrées lexicales. La prise en compte de telles règles permet de rétablir une certaine homogénéité

---

[3] Il ne faut pas perdre de vue que, dans le contexte de réelles applications ou implémentations, on ne s'intéresse qu'à une sous-partie de la grammaire et du lexique des langues considérées. En général, on plafonne ainsi très vite dans la prise en compte des phénomènes liés à la grammaticalisation alors que la construction du lexique reste souvent une activité "ouverte".

[4] Pour une présentation des MST en tant que modèles basés sur la notion de traduction, voir Mel'čuk 1978.

dans le fonctionnement des MST puisque, de cette façon, toutes les composantes de ces modèles mettent en jeu des règles de correspondance spécialisées. Ceci isole aussi, par une conséquence logique, cette opération particulière que nous avons appelée arborisation.

Notons que l'arborisation ne se décrit pas seulement par l'activation des règles de correspondance sémantique. Cette opération doit être (i) ordonnée — pour savoir dans quel ordre effectuer les traductions des différents liens actanciels sémantiques, et (ii) contrôlée — pour assurer que "l'assemblage" de l'arbre syntaxique profond s'effectue correctement. Dans cette partie, Chapitre VI, nous présenterons nos propositions quant à l'organisation de l'arborisation à travers le traitement d'un exemple.

2.1.4 La *hiérarchisation*

La hiérarchisation est une opération qui ne s'appuie a priori sur aucune règle déclarative Sens-Texte. En ce sens, elle ne pouvait pas être déduite directement de l'examen des différents éléments de la composante sémantique d'un MST mais elle découle logiquement de deux aspects importants de celle-ci:

1) Les nœuds dominants apparaissant dans les sous-réseaux sémantiques impliquent conceptuellement une notion de hiérarchie entre les nœuds de ces représentations.
2) L'étude de l'activation des règles sémantiques a permis de mettre à jour une certaine corrélation entre la notion de dépendance communicative véhiculée par les nœuds dominants des sous-réseaux et l'organisation hiérarchique des éléments de la RSyntP.

Pour être plus précis, la hiérarchisation considérée ici devrait être appelée *hiérarchisation communicative*. En effet, l'opération d'arborisation correspond en quelque sorte elle aussi à une hiérarchisation. Elle permet la construction d'une représentation qui, en tant que

représentation syntaxique, se caractérise par une organisation hiérarchique des nœuds la constituant. On peut ainsi noter que la grande différence entre l'arrangement "spatial" des réseaux sémantiques et celui des arbres syntaxiques profonds est que ces derniers sont organisés verticalement (du haut vers le bas) alors que, pour les premiers, il n'existe pas a priori d'axe de construction. La construction des RSyntP présuppose donc une hiérarchisation syntaxique.

Nous montrerons, Chapitres V et VI, que ce que nous définissons comme la hiérarchisation communicative (qui est sémantique, puisqu'elle s'effectue au niveau de la RSém) contient en germe la production de la hiérarchie syntaxique profonde de la phrase.

La hiérarchisation que nous allons décrire dans cette seconde partie se décompose en deux opérations logiques:

1) le choix d'un nœud de la RSém qui sera placé au sommet de la hiérarchie — ce nœud est appelé le *nœud dominant* de la RSém;
2) la détermination du reste de la hiérarchie communicative à partir du nœud dominant l'ensemble de la RSém.

Ces deux opérations, relevant de systèmes de règles complètement différents, seront présentées séparément: respectivement, dans les Chapitres IV et V.

2.2 Dans quel ordre effectuer les opérations de transition sémantique?

Non seulement les opérations de transition sémantique doivent-elles être décrites en elles-mêmes, mais il convient aussi de déterminer les différentes contraintes sur l'ordre dans lequel elles s'effectuent. Il est assez aisé de montrer qu'il existe un ordre

théorique logique pour l'exécution des quatre opérations constituant la transition sémantique.

Soit le tableau de la figure III-2 qui récapitule ces opérations en spécifiant (i) les niveaux de représentation qu'elles mettent en jeu, (ii) les règles sur lesquelles elles s'appuient, et (iii) leur finalité:

Opérations	Niveaux	Règles activées	But
lexicalisation	RSém	règles de paraphrase sémantique	choix des Sens exprimés lexicalement dans le Texte
grammaticalisation	RSém et RSyntP	règles grammatico-sémantiques	choix des Sens exprimés grammaticalement dans le Texte
arborisation	RSém et RSyntP	règles de correspondance sémantique	construction de l'arbre syntaxique profond
hiérarchisation	RSém	règles procédurales	détermination d'une hiérarchie communicative dans la RSém

Figure III-2
Tableau récapitulatif des quatre opérations de transition sémantique

Les trois premières opérations mentionnées ici s'appuient toutes sur des règles déclaratives de la composante sémantique des MST. Ces règles, qui viennent d'être décrites dans la Partie **1**, font toutes appel à des configurations sémantiques à "matcher" sur la RSém et ces configurations doivent toujours comprendre la spécification d'un nœud dominant[5]. On peut donc déduire de ceci que la hiérarchisation communicative d'une RSém doit avoir été effectuée avant que les trois opérations de lexicalisation, grammaticalisation et arborisation puissent l'être à leur tour.

Nous avons donc isolé la première opération à effectuer. Nous pouvons remarquer ensuite que l'opération d'arborisation, qui permet la construction comme telle de l'arbre syntaxique profond, doit s'appuyer sur une RSém déjà lexicalisée. En effet, une fois qu'une

---

[5] Cette nécessité de la spécification d'un nœud dominant n'a pas été mentionnée explicitement dans le cas des règles grammatico-sémantiques. Cependant, la partie gauche de ces règles présentant un sous-réseau sémantique à matcher sur la RSém, ce sous-réseau doit posséder un nœud dominant qui désigne en fait quel Sens est "déterminé" par la signification grammaticalisable (cf., dans le cas des grammèmes, les exemples présentés en II:3.3). La contrainte formelle énoncée en I:2.1 nous assure que TOUS les sous-réseaux sémantiques de ce type, apparaissant dans des règles d'un MST, posséderont un nœud dominant.

configuration sémantique aura été associée à une configuration syntaxique profonde, il sera trop tard pour effectuer une lexicalisation (au sens propre) mettant en jeu les mêmes significations. De la même façon, la grammaticalisation devra précéder l'arborisation pour ne pas traduire en un sous-arbre syntaxique profond une configuration sémantique, c'est-à-dire une signification, qui devrait être exprimée grammaticalement dans le Texte. On voit donc que l'arborisation sera logiquement la dernière opération effectuée.

Reste maintenant à ordonner les deux opérations de lexicalisation et de grammaticalisation. Nous n'avons pas trouvé de raison logique pour effectuer l'une avant l'autre. Il a toutefois été signalé plus haut que la grammaticalisation s'appuyait sur un ensemble de règles beaucoup plus restreint que la lexicalisation. Il peut donc être justifié, d'un point de vue strictement pratique, d'effectuer en premier lieu l'opération plus aisée de grammaticalisation afin de réduire au maximum la RSém à traiter lors de la lexicalisation, et ainsi de diminuer le nombre de règles de paraphrase sémantique à examiner. En effet, il est évident que moins une RSém met en jeu de Sens, plus le nombre de règles de paraphrase sémantique potentiellement concernées par cette représentation est petit. Le fait d'effectuer la grammaticalisation avant la lexicalisation permet ainsi de rendre plus performante l'exécution de la seconde opération. Rappelons cependant qu'a priori cet ordre pourrait être interverti.

Maintenant que toutes les opérations sémantiques fondamentales ont été introduites, il est possible de construire un schéma complet présentant l'enchaînement des opérations de transition RSém ⇒ RSyntP. La figure III-3 ci-dessous visualise ces opérations. Dans cette figure, nous avons cherché à représenter le fait que plusieurs hiérarchisations différentes peuvent, dans certains cas, conduire à la même lexicalisation. Il apparaît aussi que l'opération de grammaticalisation, du fait de son caractère obligatoire, entraîne moins d'options que la lexicalisation. Nous allons montrer, en 3, que la représentation servant de point de départ à l'opération

d'arborisation peut en fait être considérée comme une *RSém réduite* (voir l'introduction de cette notion en II:2.3). La figure III-3 indique que l'arborisation associe à chaque RSém réduite une et une seule RSyntP. (Rappelons que cette propriété distingue justement la RSém réduite de la RSém véritable.)

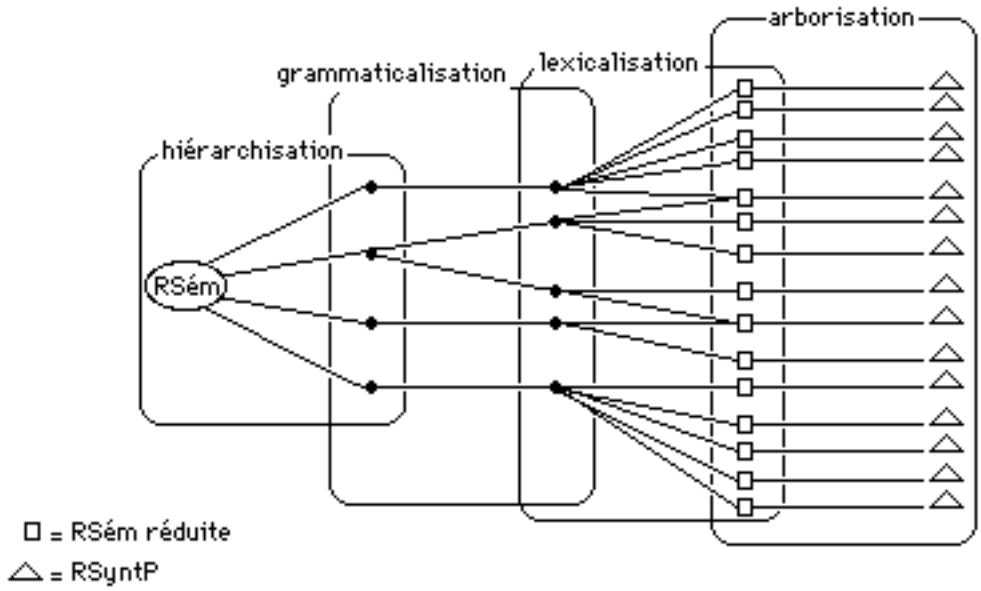


Figure III-3
Enchaînement des opérations de transition RSém ⇒ RSyntP

L'examen des caractéristiques de la composante sémantique d'un MST nous a non seulement permis d'isoler quatre opérations fondamentales définissant la transition RSém ⇒ RSyntP mais nous a aussi permis d'ordonner logiquement ces opérations. Le fait d'avoir reformulé les règles sémantiques et isolé les différentes opérations les mettant en jeu est un premier pas dans la démonstration de la thèse annoncée dans l'Introduction; à savoir qu'il existe une méthode rigoureuse et générale d'organisation de la transition RSém ⇒ RSyntP basée sur les règles qui viennent d'être décrites et implémentant les quatre opérations associées. Pour que la méthode que nous avons développée soit réellement présentée, il faut encore que les mécanismes de fonctionnement des quatre opérations proposées aient été décrits, justifiés et, dans la mesure du possible, que leur

validité ait été prouvée. La hiérarchisation, notamment, doit apparaître de plus en plus, à ce stade de l'exposé, comme la clef de voûte de notre démonstration. Or, cette opération et la notion qu'elle sous-tend de dépendance communicative resteront encore très floues tant que n'aura pas été véritablement introduit le système de procédures d'activation de la composante sémantique.

3 METHODE PROPOSEE ET NOTION DE RSEM REDUITE

Nous postulons ici que la transition sémantique peut être contrôlée par des principes linguistiques généraux, principes qui permettent une implémentation efficace et linguistiquement pertinente. Nous verrons plus loin que ces principes reposent sur deux idées-clé:

La structure communicative d'une RSém joue un rôle majeur dans le contrôle du processus de transition sémantique.

La transition sémantique doit se faire à partir d'un parcours complet et systématique du réseau qui modélise une certaine lecture (interprétation) de celui-ci.

Ce parcours du réseau sera appelé *parcours sémantique*. Il est utile de préciser maintenant la terminologie que nous emploierons pour manipuler formellement les réseaux sémantiques. Cette terminologie est la plus proche possible de celle standard utilisée dans les travaux liés à la théorie des graphes orientés[6]. (On peut se référer à Sowa 1984:Annexe A pour une introduction succincte aux différentes notions mises en jeu dans la manipulation des graphes.)

---

6 Rappelons que les réseaux sémantiques ont été définis comme des graphes orientés connectés en I:2.1.

Nous définissons tout d'abord, dans un graphe orienté, la notion de *chemin*. Elle désigne un ensemble NON ORDONNE d'arcs connectés les uns aux autres DE FAÇON A FORMER UNE CHAINE; un arc entre deux nœuds n_1 et n_2 peut être noté $<n_1,n_2>$. Un même chemin ne *traverse* qu'une seule fois un nœud donné; c'est-à-dire qu'un nœud ne peut apparaître dans un chemin qu'une seule fois à la jonction entre deux arcs. Par voie de conséquence, un même arc ne peut apparaître qu'une seule fois dans un chemin. On peut définir dans le graphe de la figure III-4, ci-dessous, de multiples chemins dont C_1, C_2 et C_3 mais, puisque notre notion de chemin n'inclut pas la notion d'ordonnancement, C_3 est formellement identique à C_2:

C_1 = { <A,B> , <B,E> , <E,D> }
C_2 = { <C,B> , <B,E> , <E,F> }
C_3 = { <F,E> , <E,B> , <B,C> }

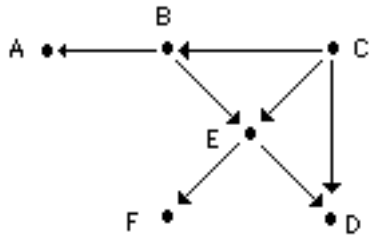


Figure III-4
Exemple de graphe orienté

Un *parcours* se définit comme UN ENSEMBLE ORDONNE DE COUPLES DE NOEUDS correspondant à des arcs connectés. Les arcs sont désignés dans un parcours en utilisant des couples de nœuds, notés (n_1,n_2), pour rendre compte d'une orientation donnée du parcours de l'arc. Pour simplifier, on pourra cependant considérer que le parcours est aussi un ensemble d'arcs. À part l'ordonnancement et l'orientation, deux caractéristiques distinguent la notion de parcours de celle de chemin:

1) La connexion entre les arcs impliqués dans un parcours est tout à fait libre; c'est-à-dire qu'arrivé à une étape donnée, le parcours peut se poursuivre à partir de n'importe quel nœud déjà atteint et sur n'importe quel arc pourvu qu'il n'ait pas encore été parcouru. Le parcours ne définit donc pas nécessairement une chaîne.
2) La notion de parcours ne renvoie pas uniquement à un ensemble d'arcs; elle englobe aussi un aspect dynamique. Le parcours peut désigner le processus permettant de construire, par étapes successives, cet ensemble d'arcs.

Le parcours P_1, ci-dessous, est un exemple de parcours du graphe de la figure III-4:

P_1 = { (E,F) , (E,D) , (E,C) , (C,B) , (B,E) , (B,A) , (C,D) }

Dans le cadre plus spécifique du traitement procédural de la transition RSém ⇒ RSyntP, nous définissons la notion de *parcours sémantique* de la façon suivante:

Un *parcours sémantique* est un parcours (i) effectué sur une RSém (ii) tel qu'il emprunte une et une seule fois chacun des arcs de cette RSém, dans le sens de la relation actancielle, (n_1,n_2) sur $n_1 \rightarrow n_2$, ou à contre-sens, (n_2,n_1) sur $n_1 \rightarrow n_2$.

Ainsi, l'orientation du parcours sémantique ne respecte pas nécessairement celle des arcs (qui représentent les relations sémantiques prédicat-arguments). Par contre, un tel parcours doit être COMPLET, c'est-à-dire qu'il recouvre toute la superficie du réseau: il comprend tous les arcs du réseau.

Comme tout parcours, le parcours sémantique doit trouver avant tout son point de départ. Ce point de départ joue un rôle capital dans les différentes opérations composant la transition sémantique (qui ont été récapitulées dans la section précédente). Nous l'avons appelé

le *nœud dominant* du réseau et l'on peut d'ores et déjà signaler qu'il est en correspondance directe avec le sommet de l'arbre de syntaxique profond que nous cherchons à bâtir. Rappelons que la notion de nœud dominant d'une définition de lexème, puis, par extension, d'une RSém, a été introduite dans la partie précédente au moment de l'étude des règles de paraphrase sémantique (cf. II:2.1 et II:2.3).

Si nous voulons résumer les remarques faites à propos de la transition RSém ⇒ RSyntP, nous dirons que le problème posé est double:

1) problème de l'écart formel entre les deux types de représentation qui implique ce que nous avons convenu d'appeler une hiérarchisation du réseau sémantique;
2) problème de la non-unicité du lien entre RSém et RSyntP qui fait qu'à une RSém peut correspondre de multiples RSyntP.

Le second problème peut être modélisé comme suit: supposons que le fait d'associer un arbre syntaxique profond à une RSém corresponde à une sorte d'"interprétation" de celle-ci; on dira qu'une RSém est toujours vague vis-à-vis de cette interprétation puisque l'on peut lui associer plusieurs RSyntP. Pour réaliser l'arborisation, il va donc falloir, pour chaque arbre syntaxique profond, trouver un moyen de sélectionner une interprétation de la RSém.

Nous avons rappelé au début de ce chapitre que, pour effectuer la transition RSém ⇒ RSyntP, nous proposons de transiter par une étape de représentation intermédiaire où l'écart formel entre RSém et RSyntP se trouve limité. La représentation intermédiaire que nous postulons a déjà été brièvement introduite en II:2.3 et est appelée *représentation sémantique réduite* (*RSémR*). La RSémR d'un Texte se caractérise par les trois aspects suivants:

1) La structure sémantique d'une RSémR est figée. Ceci implique qu'aucune lexicalisation au sens propre (application d'une règle

de paraphrase sémantique) ou grammaticalisation (application d'une règle grammatico-sémantique) ne peut être effectuée.

2) Une RSémR possède un nœud dominant, identifié dans les figures au moyen du soulignement, qui sera le point de départ du processus d'arborisation.
3) Une RSémR est en relation univoque avec un arbre syntaxique profond donné. Cette caractéristique, déjà signalée dans la Partie **1** à propos des definiens décrits comme des réseaux sémantiques réduits, est une conséquence des deux caractéristiques formelles qui viennent d'être mentionnées.

Il est très important de noter que LES RSEMR N'EXISTENT PAS, DANS LE CADRE DE LA TST, EN TANT QUE VERITABLES REPRESENTATIONS LINGUISTIQUES. Elles correspondent à une sorte d'artifice formel qui permet de décomposer un niveau de transition en mettant en évidence une étape logique de cette transition. En ce sens, si l'on considère une équivalence RSém ⇔ RSyntP comme une simple MISE EN CORRESPONDANCE de deux niveaux de représentation, seules comptent, en tant que représentations linguistiques, la RSém et la RSyntP. Si RSém ⇔ RSyntP est considérée dans le cadre d'une opération de TRANSITION (RSém ⇒ RSyntP), alors, la prise en compte d'une RSémR va permettre de mieux décrire ce processus de transition comme tel. Il est ainsi possible de concevoir le rôle de la RSémR, dans le processus de transition RSém ⇒ RSyntP, de la façon suivante:

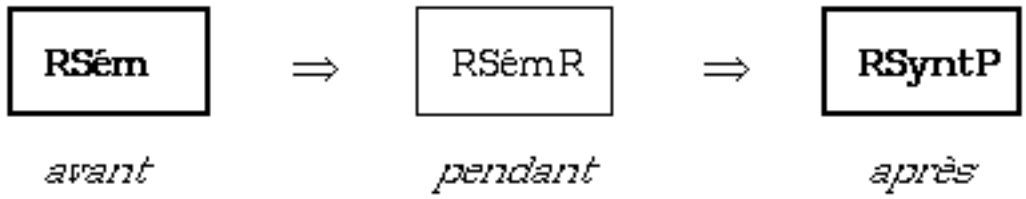


Figure III-5
Rôle de la RSémR dans la transition RSém ⇒ RSyntP

D'une certaine façon, on pourrait être tenté de dire que la RSémR est une sorte de RSém de Surface, par opposition à la RSém qui serait, elle, une RSém Profonde. Rappelons que la dichotomie profond/de surface, dans l'approche Sens-Texte, est interprétée de la façon suivante:

- profond, qui est plus directement lié à ce qui est exprimé (le Sens);
- de surface, qui est plus directement lié à ce qui exprime (le Texte).

La RSémR, qui spécifie (i) la sélection des Sens qui seront effectivement exprimés, ainsi que (ii) un nœud dominant impliquant une hiérarchie communicative dont nous verrons qu'elle est un reflet de la hiérarchie syntaxique, peut donc être vue comme plus directement liée à LA FACON D'EXPRIMER le Sens qu'au Sens lui-même. Il y aurait cependant une sorte d'incohérence à parler de RSém Profonde dans la mesure où la RSém EST le Sens et n'est liée à rien d'autre de plus profond à l'intérieur du modèle linguistique. Il est donc préférable de conserver à la RSémR son statut de représentation sémantique intermédiaire et de garder un seul niveau de représentation sémantique véritable.

Il est intéressant de noter que Y. Apresjan, dans les recherches menées à Moscou sur la TST (cf. Apresjan 1980), a avancé l'idée d'une dichotomie profond/de surface jouant au niveau sémantique. Mais alors que notre RSémP serait grosso modo la RSém classique de la TST, la proposition de Y. Apresjan visait à "aller encore plus en profondeur". La RSém serait ainsi une RSémS, distincte d'une autre représentation sémantique beaucoup plus proche des niveaux conceptuels et des sens non-linguistiques (non liés à une langue donnée): selon Apresjan, la RSémP relèverait d'une "sémantique universelle" alors que la RSémS relèverait d'une "sémantique nationale". Soit:

- l'APPROCHE 1, la réattribution des rôles des différentes représentations sémantiques dans le cadre d'une dichotomie sémantique profond/de surface dérivée de notre approche;
- l'APPROCHE 2, la dichotomie inspirée des propositions de Apresjan.

La différence entre ces deux approches pourrait être visualisée de la façon suivante:

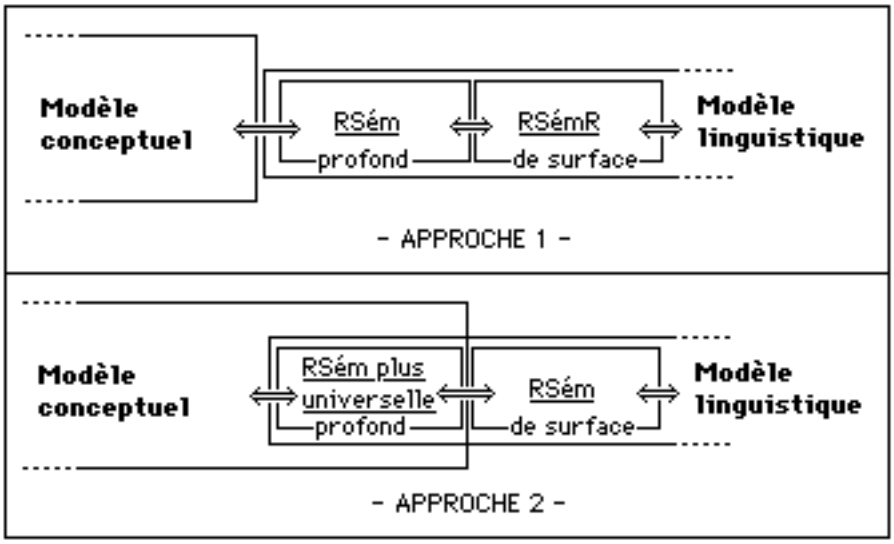


Figure III-6
Deux approches de la dichotomie profond/de surface pour le niveau sémantique

Nous pensons a priori qu'il est préférable de laisser à la notion de sémantique son caractère purement linguistique en fonction duquel la RSém vise à représenter uniquement l'invariant de paraphrase langagière (cf. Chapitre I) — ce que ne ferait pas une RSém "trop profonde". Nous laisserons cependant de côté ce débat puisque, de toute façon, notre but n'est pas ici d'effectuer une remise en question de la démarche théorique initiale sur laquelle nous nous basons.

Remarque: Dans un cadre de TRANSITION RSém ⇒ RSyntP, le recours au niveau intermédiaire de RSémR fait que les *règles*

grammatico-sémantiques, qui ont été présentées dans la Partie précédente (cf. II:3) comme des règles de correspondance RSém ⇔ RSyntP, fonctionneront en fait comme des règles de transition RSém ⇒ RSémR. Les *règles de correspondance sémantique proprement dites* (cf. II:4), quant à elles, prendront en charge la transition RSémR ⇒ RSyntP.

L'utilisation d'un niveau de représentation intermédiaire, pour décomposer formellement une correspondance entre deux niveaux, n'est pas une étrangeté dans le cadre de l'approche Sens-Texte. On trouve dans le *Cours de Morphologie Générale* (Mel'čuk à paraître, Sixième Partie, Chapitre I, §2) une justification du recours à deux niveaux intermédiaires de ce type, entre la RMorphS et la RPhonP, qui sont:

- la Représentation Morphique;
- la Représentation Morphonologique.

La justification de ces sous-niveaux présuppose tout un travail préparatoire de définition du modèle morphologique proposé. Nous renvoyons donc le lecteur au *Cours de Morphologie Générale* pour plus de détails. Remarquons cependant que ce même texte mentionne la possibilité de considérer un niveau SYNTAXIQUE intermédiaire entre le niveau de RSyntS et le niveau de RMorphP. Cette représentation prendrait la forme d'un arbre syntaxique partiellement ordonné: seuls les composants des groupes syntaxiques seraient linéarisés, sans que ces groupes soient eux-mêmes ordonnés les uns par rapport aux autres.

Ces niveaux intermédiaires sont présentés dans Mel'čuk (à paraître) comme véritablement linguistiques. C'est-à-dire que leur motivation est étrangère au problème de l'activation des MST mais relève de la seule méthodologie de description linguistique. Dans l'état actuel de notre recherche, nous ne sommes pas convaincu que l'on puisse en dire autant d'un niveau tel que celui de la RSémR que nous avons initialement développé dans un contexte de synthèse linguistique. Nous laisserons cependant cette question en suspens

pour l'instant. Il est intéressant de constater que, dans le cadre des implémentations GOSSIP et FoG, nous avons utilisé une étape intermédiaire dans la transition RSyntS ⇒ RMorphP, étape où les éléments de la RSyntS se voient attribuer une "position relative" avant que ne soit calculée, pour les éléments de la RMorphP, une "position absolue". Cette approche, de même que l'idée de transiter (pour ce qui est de la composante sémantique) par une RSémR, s'est imposée à nous sans que nous ayons eu alors connaissance des thèses présentes dans Mel'čuk (à paraître). Ceci est un indice supplémentaire du fait que de ces sous-niveaux additionnels sont (comme il est affirmé dans cet ouvrage) présupposés par le MST. Deux autres travaux peuvent être cités à ce propos.

Nous avons trouvé un équivalent de la RSémR dans Beringer 1988, qui est une thèse en informatique traitant de l'analyse automatique des phrases françaises au moyen d'un modèle de type Sens-Texte. H. Beringer définit de la façon suivante un niveau qu'il appelle "niveau de représentation présémantique":

> "(...) la représentation présémantique est une représentation sémantique de l'énoncé précédant toute exploitation des relations de synonymie" (p.188)

Compte tenu du fait qu'il considère une orientation Texte ⇒ Sens, on peut remarquer que sa définition pourrait correspondre à une vision analytique de ce que nous appelons ici RSémR.

St-Germain 1988, dans une étude Sens-Texte des verbes causatifs du français, énonce le principe suivant:

> "la représentation sémantique finale doit être en (ou doit tendre vers une) correspondance biunivoque avec la représentation syntaxique profonde." (p.247)

L'orientation envisagée ici est bien sûr synthétique. Nous adhérons complètement au principe ci-dessus. Le travail de J. St-Germain vise

entre autre à développer les connaissances sur ce que nous appellerons plus loin la *topographie des réseaux sémantiques* (voir IV:2) afin de permettre une application simple des règles sémantiques. La modélisation des procédures de transition RSém ⇒ RSyntP que nous proposons ici relève, selon nous, du même type d'approche.

Nous avons jusqu'à présent essentiellement cherché à justifier et expliquer le recours à la RSémR comme sous-niveau de représentation intermédiaire. Si nous acceptons de considérer pour le moment le principe d'une telle représentation comme étant acquis, il nous reste à expliciter les trois caractéristiques des RSémR énumérées plus haut. Cette explicitation doit être à la fois fonctionnelle — rôle joué dans le processus de transition sémantique — et formelle. Nous avons déjà décrit brièvement les opérations de lexicalisation et de grammaticalisation qui servent à la construction de la RSémR. Dans le chapitre qui suit, nous allons étudier en détail la notion de nœud dominant (qui a seulement été esquissée dans la Partie **1**) et montrer comment le nœud dominant d'un réseau peut être calculé. Cette notion sera décrite en liaison avec celle de hiérarchisation communicative des RSém, ce qui nous permettra de donner plus de corps à la troisième caractéristique des RSémR: le lien d'unicité avec une hiérarchie syntaxique profonde donnée. La hiérarchie communicative étant calculée à partir d'un *parcours sémantique* (c'est-à-dire complet) de la RSém, nous décrirons, au Chapitre V, comment s'effectue le parcours sémantique en nous basant notamment sur un exemple précis de réseau relativement complexe.

Ceci terminera notre caractérisation des RSémR et des notions liées de *nœud dominant*, *hiérarchie communicative* et *parcours sémantique*. Rappelons que pour terminer cette étude des aspects procéduraux de la transition sémantique, nous devrons encore présenter (Chapitre VI) les procédures d'activation des règles de correspondance sémantique qui ont été décrites en II:4.

CHAPITRE IV

Hiérarchisation: calcul du nœud dominant des réseaux sémantiques

La notion de nœud dominant n'est pas entièrement nouvelle. Sans que cette notion ait véritablement été définie au préalable, on trouve dans la littérature récente sur la TST une utilisation plus ou moins systématique du soulignement des nœuds majeurs, ou dominants, dans les RSém. Ce que nous proposons ici, c'est à la fois une théorisation de cette notion et une description de son rôle dans les différentes étapes de la transition sémantique. Nous tenons à souligner que ce qui suit doit énormément au travail effectué en collaboration avec L. Iordanskaja (Iordanskaja & Polguère 1988) et I. Mel'čuk (Mel'čuk & Polguère à paraître).

1 LA NOTION DE NOEUD DOMINANT

Une des différences formelles majeures entre une RSém et une RSyntP qui peut lui être associée est que la seconde représentation rend visible une *hiérarchie stricte* qui n'existe pas dans la première.

Nous entendons par hiérarchie stricte une hiérarchie telle qu'elle peut être représentée au moyen d'un arbre.

Ainsi, dans une RSyntP, tous les nœuds sont des dépendants directs d'un seul nœud et toute la hiérarchie syntaxique est dominée par un seul et unique nœud: le sommet (ou la racine) de l'arbre syntaxique. Pour faire le lien entre une RSém et une RSyntP, nous proposons une méthode permettant d'introduire, au niveau sémantique, une hiérarchie qui est une sorte de "prémonition" de la hiérarchie syntaxique que nous cherchons à construire: c'est ce que nous avons appelé la *hiérarchie communicative* (cf. II:4.1.2). La "beauté" de notre approche, si nous osons nous exprimer ainsi, réside dans les trois points suivants:

1) Elle se base sur une notion déjà présente dans la TST, même si elle n'avait jamais été clairement étudiée: la notion de nœuds sémantiques dominants[1].

2) Elle met à jour le fait que la notion de nœud dominant est en réalité implicitement basée sur celle, très générale, de dépendance communicative. Cette dernière notion permet de mieux comprendre les deux opérations de *lexicalisation* et d'*arborisation*.

3) Elle est une véritable méthode de construction des RSyntP et, comme telle, confère un caractère déterministe au processus de transition RSém ⇒ RSyntP.

Tout au long de ce travail, nous cherchons à introduire avec le maximum de détails les différentes notions essentielles à la bonne compréhension de la méthode proposée. Cette Partie **2**, plus particulièrement, précise et complète cet ensemble de notions en

---

[1] La notion de nœuds dominants des sous-réseaux d'une RSém n'était pas présente initialement dans la TST mais a été introduite plus ou moins explicitement dans les recherches plus récentes en lexicographie ou en sémantique Sens-Texte.

présentant une sorte de "mode d'emploi" de la transition sémantique. Il est cependant normal que la portée réelle de chaque notion introduite n'apparaisse pas immédiatement de façon claire. Cela tient au fait qu'une même notion, pour être comprise dans sa totalité, doit être envisagée aussi bien d'un point de vue purement statique que d'un point de vue procédural. C'est le cas, par exemple, des notions de dépendance communicative et de nœud dominant qui, introduites toutes deux dans la partie précédente, sont réexaminées ici dans le cadre de l'étude des procédures de transition sémantique proprement dites. À la fin de cette Partie **2**, le rôle des principales notions présentées dans le processus de transition sémantique devrait apparaître beaucoup plus clairement.

Rappelons que notre but est de montrer comment construire une hiérarchie entre les nœuds sémantiques. Pour mieux comprendre la façon dont nous procédons, imaginons que nous disposions d'une RSém dont la structure sémantique serait en fait un ensemble de billes de verre (les nœuds), qui ne pourraient être déplacées que verticalement vers le haut, et qui seraient reliées entre elles par des baguettes élastiques en forme de flèches (les arcs). Une façon "mécanique" de créer une hiérarchie entre les nœuds d'un tel réseau est de saisir une des billes et de soulever le réseau en interprétant la position relative de deux nœuds liés (du nœud situé le plus haut vers celui situé le plus bas) comme une relation de hiérarchie; cette manipulation est illustrée ci-dessous:

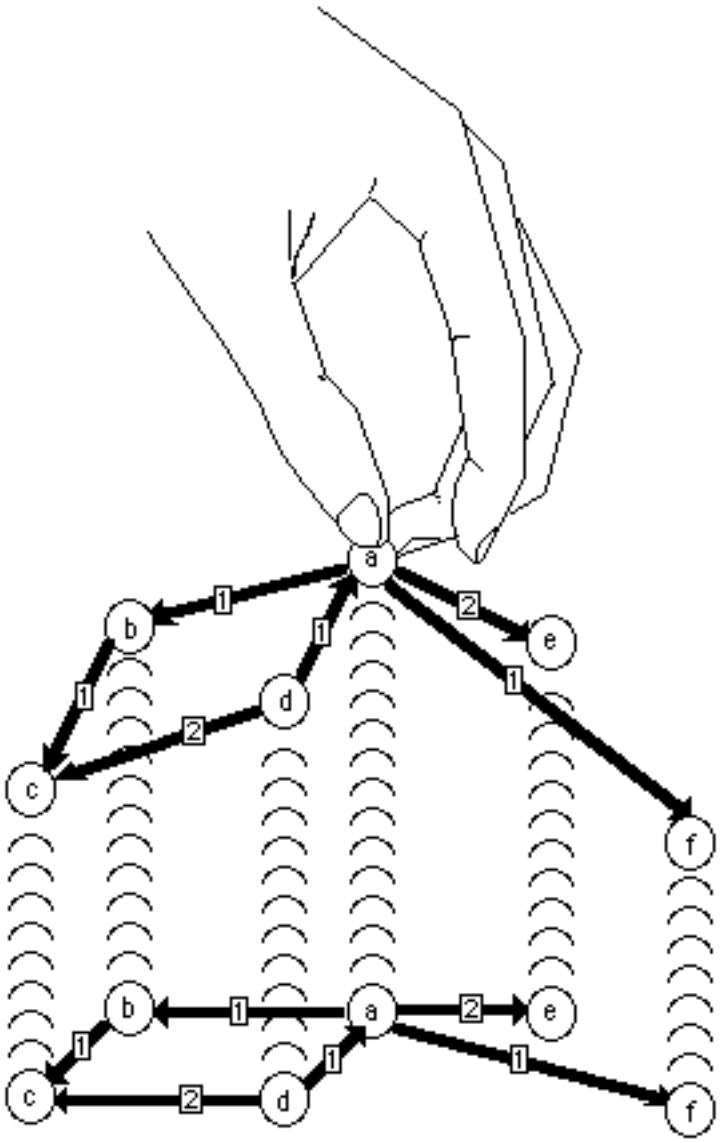


Figure IV-1
Métaphore physique de la hiérarchisation d'un réseau

Dans l'exemple de la figure IV-1, nous avons soulevé le réseau en saisissant la bille "a". De ce fait, elle se trouve placée nécessairement au-dessus de toutes les autres billes; elle domine donc, selon notre interprétation, la hiérarchie ainsi créée. Ce nœud auquel est suspendu tout le reste du réseau représente en fait le nœud dominant de notre hiérarchie. IL ACQUIERT IMMEDIATEMENT SON STATUT

DE NOEUD DOMINANT PAR LE SIMPLE FAIT QU'IL EST SELECTIONNE, SINGULARISE ET PLACE AU SOMMET DE LA HIERARCHIE. La spécification d'un nœud dominant est donc une question de choix. Notre métaphore montre aussi que le reste de la hiérarchie est créé en fonction d'un ensemble de données et de lois physiques: longueur et élasticité des arcs, ordre dans lequel ils se trouvent soulevés, etc. La suite de l'opération, une fois effectué le choix du nœud dominant, est donc une affaire de calcul basé sur la structure même du réseau. Cette métaphore illustre assez bien les deux types de problèmes que nous allons résoudre dans la description du processus de hiérarchisation:

1) LA DETERMINATION DU NOEUD DOMINANT LA HIERARCHIE COMMUNICATIVE, qui est une question de sélection d'un nœud donné parmi un ensemble de nœuds possibles et est décrite dans le présent chapitre;
2) LA HIERARCHISATION PROPREMENT DITE, qui est basée sur une procédure de calcul des liens de hiérarchie communicative à partir (i) du nœud dominant initialement sélectionné, (ii) des liens de dépendance sémantique existant entre les nœuds, et (iii), comme nous le verrons, de la structure communicative du réseau; la hiérarchisation proprement dite sera décrite au Chapitre V.

Pour compléter notre description métaphorique de la notion de nœud dominant, nous pouvons encore constater que lorsque nous examinons un réseau sémantique, afin de comprendre ce qu'il représente, nous effectuons une lecture selon un certain parcours du réseau. Ce parcours ne commencera pas n'importe où, ou bien, si c'est le cas, il faudra alors bien souvent revenir en arrière dans la lecture car le parcours entamé ne nous permet pas d'assimiler correctement le contenu sémantique du réseau. En fait, notre premier souci, avant d'amorcer cette lecture, est d'identifier un nœud particulier qui pourra nous servir de point de départ. Ce nœud possède un statut spécial dans la représentation. Nous ne pouvons simplement dire qu'il est sémantiquement le plus important, ce qui

serait trop vague; il est un de ceux qui (avec les nœuds dominants des différents sous-réseaux **thème**, **rhème**, etc.) occupent une place stratégique vis-à-vis de la structure communicative.

Le *nœud dominant* du réseau est ce point de départ de la lecture. Sa spécification n'est pas nécessaire dans la construction d'une RSém, mais relève de l'utilisation de la RSém dans le processus de transition RSém ⇒ RSyntP.

Nous avons défini dans la partie précédente la notion de *dépendance communicative* (II:4.1.2); cette notion, à son tour, nous permet de définir la notion de nœud dominant:

Définition de la notion de *nœud dominant*:

Le nœud dominant d'une RSémR est le seul nœud de cette représentation qui domine communicativement tous les autres nœuds. En d'autres termes, c'est le seul nœud sémantique qui ne dépende communicativement d'aucun autre nœud.

Trois remarques importantes doivent être faites à propos de cette définition:

1) La notion de nœud dominant est liée à celle de RSémR (examinée en détail au chapitre précédent). Par conséquent, lorsque nous parlerons de "nœud dominant d'une représentation sémantique", il faudra toujours comprendre qu'il s'agit en fait d'une RSémR.
2) La définition fait aussi appel à la notion de dépendance communicative. Il convient de rappeler qu'un nœud peut tout à fait dominer communicativement un autre nœud dont il dépend sémantiquement (c'est-à-dire, dont il est un actant sémantique). Le nœud dominant d'une RSémR sera donc lui-même presque toujours un dépendant sémantique d'un ou plusieurs autres nœuds.

3) Dans notre définition du nœud dominant, nous présupposons une certaine réciprocité entre les notions de dominance et de dépendance communicatives. Nous simplifions ici volontairement les faits car nous serons en mesure de montrer plus loin (en V:2) qu'il peut être utile d'employer ces deux notions pour référer à des phénomènes différents. Notons simplement ici que la dominance communicative correspond au phénomène de dépendance communicative CONSIDERE A L'ECHELLE DU RESEAU SEMANTIQUE. Nous pouvons cependant ignorer temporairement cette distinction.

Il est intéressant aussi de remarquer que la notion de nœud dominant (souligné) pour les sous-réseaux constituant une RSém avait été introduite dans l'appareillage formel de la composante sémantique d'un MST sans que la notion liée de dépendance communicative ait réellement été rendue explicite. Cela peut être expliqué par le fait que, si la notion de nœud dominant se définit à partir de celle de dépendance communicative (voir ci-dessus), la hiérarchie communicative, elle, ne peut être établie sur une représentation sémantique qu'une fois singularisé un nœud dominant. Nous sommes en présence d'une notion qui se définit à partir d'une autre notion qui, pourtant, en dépend fonctionnellement. (Pour pouvoir spécifier une hiérarchie communicative dans un réseau, il faut d'abord choisir un nœud dominant.)

En conclusion de cette section, il nous reste à considérer la question suivante: pourquoi le nœud dominant global n'est-il pas une donnée de la structure communicative de la RSém comme le sont les nœuds dominants des **thème**, **rhème** et autres sous-réseaux (cf. Partie **1**)?

La réponse à cette question n'est pas simple. Il faut tout d'abord accepter le postulat implicite de la TST qui veut que deux RSém qui ne diffèrent que par leur structure communicative sont tout de même des RSém distinctes. Cela justifie le fait de spécifier non seulement la structure communicative des RSém sous forme de sous-réseaux mais

aussi le fait d'équiper ces sous-réseaux de nœuds dominants "locaux" pour indiquer clairement ce que disent les Textes correspondant à une RSém et à propos de quoi cette chose est dite. C'est ce que nous pourrions appeler une spécification minimale des données communicatives. Il en va tout autrement du nœud dominant global d'une RSém. Il apparaît en effet que l'ensemble des paraphrases prises en compte par une RSém possédant un nœud dominant serait trop restrictif: puisque la spécification d'un tel nœud conditionne une hiérarchie communicative qui, à son tour, conditionne une hiérarchie syntaxique donnée. La RSém déjà équipée d'un nœud dominant deviendrait donc tellement "superficielle" qu'elle perdrait presque sa raison d'être comme niveau autonome de représentation.

2 NOEUD DOMINANT ET STRUCTURE COMMUNICATIVE

Nous avons signalé plus haut que le nœud dominant d'un réseau sémantique devait être choisi notamment en fonction de la structure communicative de ce réseau. Considérons un exemple qui est une RSém présentant plusieurs options possibles quant à la structure communicative:

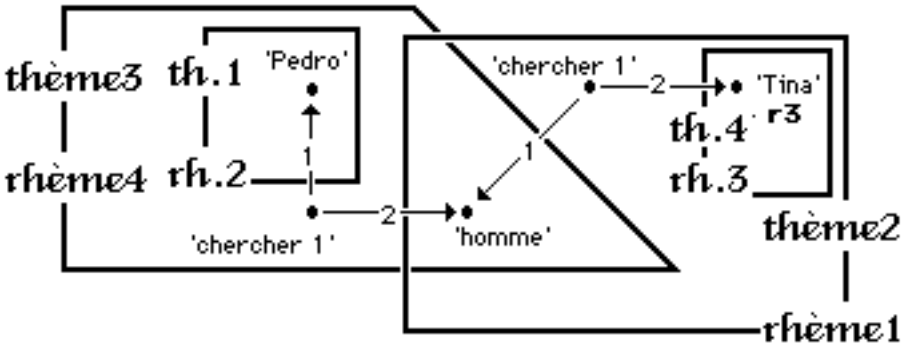


Figure IV-2
Différentes structures communicatives possibles
dans une même RSém

Tout en gardant la même structure sémantique dans la RSém, les différentes options sur la structure communicative permettent de

représenter sémantiquement des phrases communicativement très différentes, comme l'illustrent les exemples (1)-(4) ci-dessous.

(1) *Pedro* <u>*cherche*</u> *un homme qui cherche Tina.* (**th. 1** et **rh. 1**)
← 'Pedro cherche qui?'

(2) *Un homme qui cherche Tina* <u>*est cherché*</u> *par Pedro.* (**th. 2** et **rh. 2**)
← 'Un homme qui cherche Tina est cherché par qui?'

(3) *Un homme que cherche Pedro* <u>*cherche*</u> *Tina.* (**th. 3** et **rh. 3**)
← 'Un homme que cherche Pedro cherche qui?'

(4) *Tina* <u>*est cherchée*</u> *par un homme que cherche Pedro.* (**th. 4** et **rh. 4**)
← 'Tina est cherchée par qui?'

Chacun de ces exemples est associé à une question hypothétique permettant de mieux mettre en valeur la structure communicative de l'énoncé. (Nous avons déjà fait usage de la technique du questionnement dans la Partie **1**.)

Dans (1)-(4), nous avons souligné l'élément de la phrase correspondant au nœud dominant proposé. Ces exemples simplistes permettent de mettre en évidence le fait que le nœud dominant d'une RSémR occupe une place stratégique dans le réseau: il fait, en quelque sorte, la jonction entre le thème et le rhème. De façon plus générale, nous dirons que:

Le nœud dominant d'une RSémR est un des pivots de l'articulation communicative du réseau.

C'est à dessein que nous disons "UN DES pivots" et non "LE pivot". Nous verrons plus loin qu'un même réseau peut présenter plusieurs candidats valides pour le "poste" de nœud dominant. Il est aussi question ici d'articulation communicative et pas simplement de structure thème/rhème. En effet, comme nous l'avons souligné dans la Partie **1** (cf. I:2), tout ce qui est dit dans ce travail sur la structure

communicative concerne en fait avant tout l'opposition thème/rhème (la structure thématique). En conséquence, nous reprenons à notre compte l'expression "articulation communicative" souvent employée dans les travaux linguistiques traitant de la structure communicative. Pour des raisons pratiques, nous adoptons aussi le point de vue suivant énoncé dans Sgall 1987:179, sans toutefois y souscrire entièrement:

> "We have found no reason to distinguish topic/focus and theme/rheme or any similar two articulations; the system of language appears to embody just a single hierarchy of this kind, though this hierarchy is not a simple one."

Il existe bien évidemment des raisons pour distinguer, dans certains cas, plusieurs hiérarchies interagissant à l'intérieur d'un même Texte et nous verrons d'ailleurs (en V:4.2) que nous faisons un usage bien différent d'un système d'opposition comme posé/présupposé. Néanmoins, nous retiendrons ici l'idée qu'il est possible de considérer une seule articulation communicative dont la structure, bien que complexe, permet cependant de faire apparaître une claire partition minimale des énoncés — partition à laquelle nous référons tout au long de ce travail comme l'opposition thème/rhème.

Sans nous engager sur le chemin périlleux d'une véritable étude des notions de thème et de rhème, il est clair qu'une mise au point sur ces notions s'impose ici. En effet, les règles de calcul du nœud dominant qui vont être introduites étant basées sur une analyse de la structure communicative, la seule perception intuitive de ce que cette dernière structure recouvre pourrait être insuffisante pour évaluer les propositions qui seront faites ici.

Nous voudrions, avant d'entamer une courte exploration de la notion de structure communicative, signaler que cette notion, telle que nous la manipulons, aura nécessairement un statut spécial. Rappelons les deux faits suivants:

1) Un MST rend compte de l'association entre Sens et Texte en se basant sur des représentations concernant le niveau de la phrase prise en isolation.
2) La structure communicative d'une phrase est un reflet de ce qui, dans cette phrase, est présenté comme plus ou moins lié au contexte d'énonciation. Il est donc souvent difficile de percevoir la pertinence d'une structuration communicative d'une phrase prise en isolation comme le montrent les exemples suivants où le même Texte (à la surface), placé dans des contextes différents, semble présenter une organisation communicative différente:

   Tiens! Telesphore arrive. <u>Ce glouton a mangé toutes les confitures.</u>

   A - *Qui a mangé toutes les confitures?*
   B - *<u>Ce glouton a mangé toutes les confitures</u>.*

Il résulte de ces deux faits que la représentation, au niveau sémantique, de la structure communicative d'un énoncé peut, dans bien des cas, présenter un caractère arbitraire. L'intuition que l'on peut avoir du Sens d'un énoncé ne suffit alors plus pour juger de la pertinence d'une RSém donnée si l'on ne situe pas cet énoncé dans un véritable contexte d'énonciation. Ce problème est particulièrement sensible lorsque l'on étudie des langues comme le français ou l'anglais, qui manifestent souvent la structuration communicative d'un énoncé par des moyens plus discrets que l'ordre linéaire (qui, lui, correspond à une donnée directement perceptible).

Nous pensons cependant – et c'est justement un des buts de ce travail que d'apporter des éléments de preuve à ce propos – que la notion de structure communicative joue un rôle crucial dans la construction de tout modèle linguistique et dans la description du lien entre Sens et Texte.

On peut citer, pour appuyer cette thèse, la réflexion extrêmement pertinente de M. Halliday (Halliday 1974:44; c'est nous qui soulignons):

> "I will assume here that FSP [= Functional Sentence Perspective; c'est-à-dire, grosso modo, ce que nous entendons par structure communicative d'un Texte] is a universal phenomenon. There were at one time suggestions that it might be specific to certain languages, at least in its fully developed form; <u>but these were founded, in my opinion, on a confusion between FSP as the organization of meaning, on the one hand, and the means by which it is realized on the other</u>."

La suite de la présente discussion devrait permettre de justifier, du moins en partie, la thèse avancée ici.

On cite généralement V. Mathesius (cf. Mathesius 1939) comme l'instigateur des travaux linguistiques sur la partition communicative des énoncés[2]. (Pour un véritable rappel historique des notions abordées ici, on peut consulter Firbas 1974.) Selon Mathesius, un énoncé standard est communicativement divisé en deux parties complémentaires:

- La première partie joue le rôle de POINT DE DéPART de l'énonciation. Elle contient l'information présentée par le locuteur comme connue et correspond grosso modo à ce que l'on a appelé le thème ou le topique.
- La seconde partie est l'ELEMENT COMMUNICATIVEMENT CENTRAL (LE NOYAU) de l'énoncé. Elle contient l'information présentée par le locuteur comme nouvelle, ce que le locuteur dit à propos du point de départ, et correspond grosso modo aux notions de rhème ou focus.

---

[2] Nous employons ici le terme "communicative" pour qualifier cette partition des énoncés même si cette appellation n'est pas nécessairement adoptée dans tous les travaux auxquels nous ferons référence dans cette section.

Cette partition a été largement exploitée ensuite dans le cadre de la linguistique fonctionnelle du Cercle de Prague ainsi que dans d'autres approches, comme celle de M. Halliday. Malheureusement, on peut dire qu'il existe presque autant de points de vue sur ce qu'est la structure communicative d'un énoncé qu'il y a de linguistes ayant étudié la question. On assiste ainsi à une multiplication des théorisations, qui s'accompagne, bien évidemment, d'une multiplication des outils terminologiques[3] et des débats théoriques sur la question. Iordanskaja 1989, citant Morgan & Sellner 1980:170-171 à propos de la notion de structure communicative:

> "The main problem with this approach... is that crucial, intuitively appealing notions such as topic, given/new, theme, rheme, and so forth, ... are not clearly and explicitly defined or described, so that it is not clear how to apply them in any but the simplest cases."

note que l'opposition thème/rhème est une donnée sémantique extrêmement subjective. Elle n'est pas nécessairement liée à une quelconque réalité dénotationnelle. Ainsi, le thème est ce que le locuteur PRESENTE comme donné, mais il importe peu, dans bien des cas, que cette information soit réellement connue du destinataire. Cela est particulièrement vrai lorsque l'on considère des phrases prises isolément. L'opposition thème/rhème ne doit donc pas être réduite à une opposition connu/inconnu. L'importance de l'opposition thème/rhème varie aussi suivant les langues considérées. Il semblerait ainsi que le français soit beaucoup plus tolérant que le tchèque (ou que les langues slaves en général, qui ont été le premier terrain d'étude de la structure communicative) quant à la possibilité de modifier la structure communicative d'une phrase dans un contexte d'énonciation donné. Ou plus exactement, il semblerait que la structure communicative en tchèque soit plus directement reflétée à

---

[3] F.Daneš, cité dans Firbas 1974:11, relève les équivalents anglais suivants pour référer à ce que nous appelons la structure communicative: "functional sentence perspective", "contextual sentence (utterance) organization", ou les oppositions "theme-rheme" et "topic-comment". La liste n'est même pas exhaustive; on pourrait par exemple ajouter "theme-tail" – cf. Dik 1978.

la surface qu'elle ne l'est en français puisque, pour cette première langue, elle est un facteur important de détermination de l'ordre linéaire. Ainsi, Gebruers 1987:113 remarque à propos de l'articulation communicative établie par V. Mathesius:

> "This articulation is held to co-determine the actual order of words. In Czech, and indeed in many other languages that have relatively “free” word order, it may be the leading word-order principle. In other languages, such as English and German, it may be outranked by other principles."

Malgré tout le vague qui entoure encore cette notion, il reste que la structure communicative sera pour nous une composante essentielle de la RSém car elle permet de rendre compte de certaines différences linguistiques majeures entre des énoncés. Notamment, toute recherche abordant le problème de la génération de texte multi-phrastique doit, pour pouvoir rendre compte de la structuration logique interne d'un texte et de ses implications sur la construction de chaque phrase le composant, aborder cette notion de structure communicative.

Il est donc relativement aisé de percevoir, même intuitivement, l'intérêt présenté par la description de la structure communicative d'un Texte. Il est cependant beaucoup plus problématique, comme le signale la citation de Morgan & Sellner 1980 ci-dessus, de déterminer des règles précises d'utilisation de cette description. Nous pensons que notre travail, en mettant en évidence le rôle joué par la structure communicative d'une RSém dans la composante sémantique d'un MST, peut apporter une contribution originale en ce sens. Il n'apporte cependant pas de réponse sur la correspondance que l'on pourrait établir entre la structure communicative et la réalité extra-linguistique d'où elle tire son origine. Des éléments de réponse peuvent être trouvés dans les travaux sur la génération de texte qui se sont justement attachés à étudier cette question (cf. Iordanskaja 1989 et Carcagno & Iordanskaja 1989) ainsi que dans les travaux en linguistique proprement dite, notamment ceux du Cercle de Prague (cf. Firbas 1964 et Daneš 1974). Davidse 1987 met en évidence un certain

rapprochement entre les travaux des linguistes du Cercle de Prague (F. Daneš, P. Sgall, etc.) et ceux des linguistes systémiques (cf. M. Halliday) dans l'importance qu'ils attribuent à la description de ce que les premiers appellent "la perspective fonctionnelle de la phrase" et les seconds "la structure thématique".

Nous voulons considérer dans ce travail une définition relativement générale de l'information véhiculée par la structure communicative. Il faut avant tout remarquer que la structure communicative est un système d'oppositions entre deux blocs d'information sémantique. Ainsi, pour utiliser la terminologie de Firbas 1964, le thème, en tant qu'élément du Texte qui possède le plus faible degré de "dynamisme communicatif", s'oppose au rhème qui est, dans le Texte, l'élément informatif par excellence. Ce système d'oppositions peut être très complexe et le même Texte peut contenir plusieurs micro-structures communicatives (ou des composantes thématique et rhématique elles-mêmes subdivisées). De la même façon, tout le contenu d'un Texte n'est pas nécessairement concerné par le jeu des oppositions communicatives. Cela implique, formellement, que toute la "surface" des RSém n'est pas nécessairement caractérisée communicativement: certains nœuds sémantiques peuvent se trouver à l'extérieur des sous-réseaux **thème** et **rhème**. J. Firbas, cité dans Davidse 1987:65, propose justement des structurations communicatives relativement complexes pour les énoncés, comme le montre l'exemple suivant:

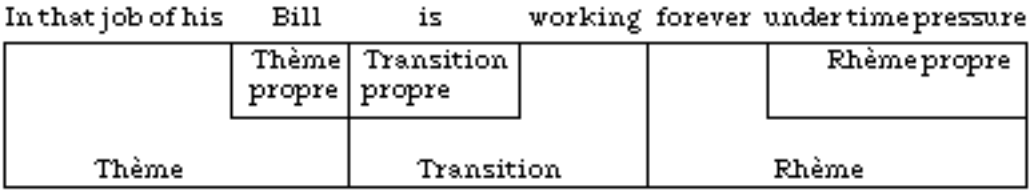


Figure IV-3
Structuration communicative complexe d'un énoncé
(proposée par J. Firbas)

Une telle description présente pour nous un intérêt certain même si nous ne pouvons pas l'adopter telle quelle pour au moins deux raisons.

Tout d'abord, elle segmente communicativement la phrase à partir d'un élément comme *is* qui est ici un verbe auxiliaire sans corrélat sémantique direct. Or, à supposer que la notion de "transition propre" soit pertinente pour la description linguistique, elle ne pourrait pas concerner le niveau sémantique puisqu'elle semble classifier des éléments de la phrase qui ne se représentent pas comme tels au niveau sémantique. (Ce sont les éléments de la phrase que J. Firbas (1965:170), appelle "the temporal and modal exponent of the finite verb" ou TME.) Il serait bien sûr possible de considérer que, dans un exemple comme celui de la figure IV-3, c'est une combinaison de grammèmes de temps et d'aspect qui se trouve jouer le rôle de transition propre. Cependant, cette combinaison de grammèmes est réalisée ici sous une forme discontinue, *is* ⊕*-ing*, et nous pensons que le fait d'isoler *is* est surtout pertinent d'un point de vue syntaxique: spécification de l'élément dominant la hiérarchie syntaxique. Notons, cependant, qu'en principe on peut tout à fait imaginer une structuration communicative isolant des grammèmes dans une phrase. Ainsi, dans l'exemple suivant où la majuscule en gras indique une forte accentuation à l'oral:

(5)

A: "Quand est-ce qu'Iznogoud doit acheter son tapis?"

B: "Iznogoud **A** acheté son tapis."

il est probable que, DANS CE CONTEXTE PRECIS, la prononciation accentuée de l'auxiliaire est en réalité un moyen de rhématiser un grammème du passé (plus qu'une simple mise de l'emphase sur ce grammème); ce serait bien sûr une chose toute différente (et une incorrection) que de dire qu'ici **A** EST le rhème de la phrase.

Un autre inconvénient de la représentation de la figure IV-3 est qu'elle occulte l'OPPOSITION thème/rhème au profit d'une structuration

tripartite où toutes les composantes sont mises sur le même plan: thème (thème propre) *vs.* transition (transition propre) *vs.* rhème (rhème propre). L'approche de Firbas (qui se trouve résumée dans Firbas 1987) est, contrairement à l'approche Sens-Texte, analytique plutôt que synthétique. Firbas prend comme point de départ de son étude la phrase, sur laquelle il cherche à repérer l'élément thématique en utilisant non pas le principe d'une claire opposition entre deux blocs communicatifs, mais une GRADATION PROGRESSIVE de ce qu'il appelle le *dynamisme communicatif* (qui est, grosso modo, la propriété d'être informatif). Cette approche le conduit à attribuer des valeurs de dynamisme communicatif à chaque élément apparaissant en surface (valeurs représentées par des entiers). Les effets (à notre avis) négatifs de cette approche analytique sont évidemment accentués par le fait que Firbas ne considère pas simultanément plusieurs niveaux de description des phénomènes communicatifs. (À l'opposé, une approche Sens-Texte permettra d'opérer une nette distinction entre une structure communicative de niveau sémantique et une structure communicative de niveau syntaxique profond.)

Malgré ces importantes réserves, nous voudrions retenir comme essentielle ici la notion de TRANSITION entre les deux blocs thématique et rhématique. Dans notre approche, la partition de la RSém mettant en évidence deux sous-réseaux **thème** et **rhème** va être exploitée dans le processus de hiérarchisation. Elle permet de repérer une composante particulière que nous avons caractérisée plus haut comme étant un pivot de l'articulation de ces deux blocs communicatifs mais qui, à un niveau plus superficiel, jouera effectivement le rôle d'élément de transition entre les composantes thématique et rhématique de la phrase[4].

---

[4] La citation suivante de Firbas 1974:13 montre à quel point notre notion de nœud dominant est liée à celle de transition dans l'approche des fonctionnalistes du Cercle de Prague:

> "By transition Mathesius understands elements that actually belong to the rheme, but occur at its periphery and in this way intermediate between theme and rheme."

Avant de mieux préciser comment se calcule le nœud dominant à partir des composantes thématique et rhématique, nous adopterons les définitions suivantes de ces composantes; ces définitions sont approximatives mais suffisantes pour nos buts présents:

Définitions approximatives des notions de *thème* et *rhème*:

Le *thème* est la composante de la RSém d'un énoncé correspondant à ce que cet énoncé, en tant que Texte, présente comme étant le plus directement lié au contexte d'énonciation.

Le *rhème* est la composante de la RSém correspondant à ce que l'énoncé, en tant que Texte, présente comme étant ce qu'il est véritablement destiné à communiquer.

La notion de contexte d'énonciation, dont nous faisons usage dans la définition du thème, englobe non seulement ce qui a été dit avant l'énoncé lui-même (le "donné" du Texte) mais aussi tout l'environnement non nécessairement linguistique: telles que les connaissances "actives" dans l'esprit des participants à l'échange linguistique au moment de cet échange, etc. Nous sommes bien sûr conscient du fait que de telles définitions ne closent pas le débat sur la nature véritable de l'opposition thème/rhème, mais elles nous permettent tout de même de rendre explicites les notions que nous considérerons ici comme non définissables dans le cadre de ce travail — notamment, les notions de communication, de liaison au contexte d'énonciation et d'énoncé.

Dire qu'un nœud dominant est un pivot de l'articulation communicative d'une RSémR implique que ce nœud sera déterminé à

La parenté entre ces notions apparaîtra encore plus clairement une fois qu'auront été présentées les règles de calcul du nœud dominant (cf. 3, ci-dessous).

partir d'une analyse de la *topographie du réseau sémantique*. Cette topographie prend en compte non seulement la structure sémantique du réseau mais aussi sa fragmentation en sous-réseaux, comme celle décrite par la structure sémantico-communicative. Nous allons maintenant proposer des règles permettant de calculer le nœud dominant en fonction d'une telle analyse topographique.

3 REGLES DE CALCUL DU NOEUD DOMINANT

Notre présentation des règles de calcul du *nœud dominant* (dorénavant ND) doit permettre de répondre aux deux questions: "Quand calculer le ND?" et "Comment calculer le ND?".

3.1 Quand calculer le ND?

Nous avons vu précédemment que la transition RSém ⇒ RSyntP peut être décomposée en quatre opérations successives: (i) la hiérarchisation, (ii) la grammaticalisation, (iii) la lexicalisation, puis (iv) l'arborisation. Nous avons signalé aussi que la hiérarchisation elle-même commence par une première étape de sélection d'un ND puis se poursuit par une étape de calcul d'une hiérarchie communicative portant sur toute la RSém et qui est héritée du ND.

> La détermination du ND précède donc les autres opérations qui constituent la transition RSém ⇒ RSyntP. Cette opération est l'étape majeure de ce qui a été appelé la *hiérarchisation* car le choix du ND conditionne toute la hiérarchie communicative de la RSém.

Rappelons que la hiérarchisation n'est pas basée sur l'utilisation de règles déclaratives formant la composante sémantique d'un MST. Elle est calculée à partir:

1) d'un ensemble de règles de détermination du ND (cf. 3.2);

2) d'un ensemble de procédures de *parcours* des RSém qui permet de spécifier les relations de *dépendance communicative* entre nœuds sémantiques (cf. Chapitre V).

On voit donc que, si la notion de dépendance communicative est utilisée dans les règles de correspondance sémantique d'un MST[5], l'opération qui permet de déterminer cette dépendance (la hiérarchisation) relève, elle, strictement des procédures de mise en jeu du modèle.

3.2 Comment calculer le ND?

Nous allons proposer cinq règles de calcul du ND. Nous donnerons auparavant deux contraintes formelles absolues, C_1 et C_2, sur le choix du ND qui s'appliquent quelle que soit la règle utilisée. Les cinq règles, R_1-R_5, sont les règles de choix proprement dites et représentent QUATRE ALTERNATIVES INTRODUITES PAR ORDRE DE PREFERENCE.

Les contraintes C_1 et C_2 doivent être respectées pour n'importe quel choix opéré au moyen d'une des règles R_1-R_5 ci-dessous. Le système de règles proposé ici doit donc être interprété selon la formule suivante:

$$C_1 \wedge C_2 \wedge (R_1 \vee R_2 \vee R_3 \vee R_4 \vee R_5).$$

3.2.1 Contrainte C_1 sur le choix du ND

C_1 Pour que le Sens 'σ' puisse être un ND, il faut que le lexème profond L dont 'σ' est le Sens soit un verbe ou qu'il puisse être rendu verbal par des moyens lexicaux ou syntaxiques.

La contrainte formelle établie ici découle logiquement de ce qui a déjà été dit dans ce chapitre et dans le chapitre précédent à propos

---

[5] Cf. la description des règles de correspondance sémantique en II:4.

des ND et des RSyntP: le ND va correspondre au sommet de la RSyntP ET ce sommet est nécessairement verbal, DONC, le ND correspond au Sens d'un lexème verbal ou "verbalisable". Cette contrainte ne s'applique bien sûr que pour la construction de phrases "standards" — les Textes correspondant à des titres, à certaines exclamations, etc. ne possédant pas nécessairement une RSyntP avec un sommet verbal.

Les moyens lexicaux ou syntaxiques utilisables pour verbaliser un lexème sont basés sur:

1) l'utilisation de la dérivation ou des verbes supports — par recours aux *fonctions lexicales*; par exemple:
   $\mathbf{V_0}$: DÉCISION**1** → DÉCIDER**I.1**
   $\mathbf{Oper_1}$: DÉCISION**1** → PRENDRE**I.c.2** ⊕ DÉCISION**1**$_{une/la}$
2) l'utilisation de moyens syntaxiques de la langue, comme la construction "attribut du sujet" des adjectifs, déjà évoquée en II:4.3.2:
   PRINTEMPS ⊕ POURRI**3** →
   PRINTEMPS ⊕ ETRE**1II** ⊕ POURRI**3**

3.2.2 <u>Contrainte C_2 sur le choix du ND</u>

C_2 Pour que le Sens 'σ' puisse être un ND, il faut qu'il soit situé sur un *chemin communicatif* unissant le nœud dominant du sous-réseau correspondant au rhème au nœud dominant du sous-réseau correspondant au thème.

Cette contrainte s'appuie sur la notion de *chemin communicatif*. C'est une notion essentielle qu'il nous faut absolument définir au risque de lasser le lecteur en augmentant encore la concentration des définitions dans cette partie de l'exposé.

Définition de la notion de *chemin communicatif*:

> Soit un *chemin*[6] reliant deux nœud n_i et n_j dans un réseau sémantique. On l'appelle *chemin communicatif* si l'on ne peut pas trouver parmi les nœuds qu'il traverse deux nœuds entretenant déjà une relation de dépendance communicative dans le réseau.

Dans la figure IV-4 ci-dessous, un des chemins communicatifs valides entre les deux nœuds n_1 et n_2 est indiqué en gras. Le chemin en pointillé n'est pas valide car il traverse n_5 et n_2. Or, n_5 étant un élément du rhème dont n_2 est le nœud dominant, ce premier est nécessairement communicativement dépendant du second. Ce parcours pointillé ne correspond donc pas à un chemin communicatif entre n_1 et n_2, ce qui fait que n_9, par exemple, ne pourrait pas être sélectionné comme ND.

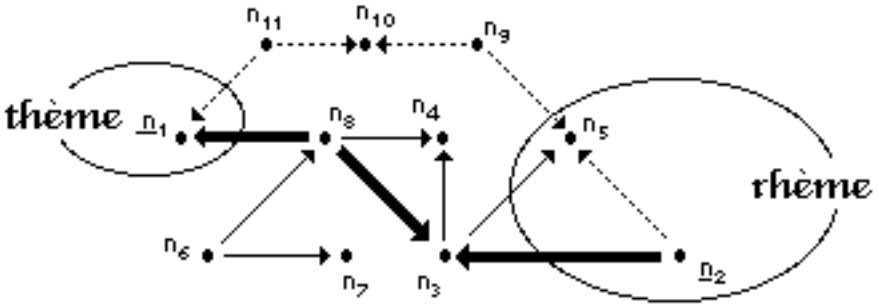


Figure IV-4
Exemple de chemin communicatif unissant deux nœuds n_1 et n_2

Il est nécessaire que le ND soit sur un chemin communicatif unissant les nœuds dominants du thème et du rhème afin qu'il puisse effectivement être le pivot de l'articulation communicative du réseau. Ainsi, dans la figure IV-4, n_6 ne devrait pas être choisi comme

---

6 Rappelons que la notion de *chemin défini sur les graphes* a été présentée en III:2.

ND car il se trouve dans un "cul-de-sac" et il ne peut aucunement jouer ce rôle central dans l'articulation communicative.

Remarque: Dans la figure IV-4, les seuls nœuds qui peuvent se trouver sur un chemin communicatif valide (et pourraient donc être sélectionnés comme des ND) sont n_1, n_2, n_3, n_4 et n_8.

3.2.3 Règle R_1 de choix du ND

R_1 Le ND d'une RSém est le nœud dominant de son rhème.

Nous avons signalé dans la Partie **1** que les sous-réseaux correspondant au thème et au rhème doivent chacun posséder un nœud dominant dans la RSém initiale. Le nœud dominant du rhème peut être considéré comme un bon premier choix comme ND de la RSém. En effet, notre définition du ND, fondée sur la notion de dépendance communicative, implique que ce nœud est tel qu'il semble communicativement "venir avant" les autres; on peut considérer que l'information introduite par le rhème, et donc par son nœud dominant, est communicativement centrale dans une RSém (selon les termes de J. Firbas, elle possède le plus haut degré de *dynamisme communicatif*). On pourra ainsi articuler la structure du Texte associé autour de ce nœud et de ses correspondants aux autres niveaux de représentation.

3.2.4 Règle R_2 de choix du ND

R_2 Le ND d'une RSém est un foncteur tel qu'un de ses actants est le nœud dominant du rhème.

Cette règle permet notamment de traiter les cas où le nœud dominant du rhème n'est pas verbalisable au niveau syntaxique, et, en particulier, lorsqu'il n'est pas lui-même un foncteur. Dans de tels cas, on peut sélectionner comme ND un foncteur dont le nœud

dominant du rhème est un actant. La contrainte C_2 assure que ce nœud n'est pas rhématisé (ne fait pas partie du rhème); l'importance de cette contrainte devient très claire ici. Lorsque nous sommes en train de sélectionner un ND, il n'existe pas encore de dépendance communicative généralisée au niveau de la RSém: la RSém n'est pas encore hiérarchisée. Cependant, les sous-réseaux spécifiés dans la RSém sont, eux, hiérarchisés puisqu'ils possèdent un nœud dominant. On ne pourrait donc pas choisir comme ND global un nœud rhématique qui ne serait pas le nœud dominant du rhème sous peine de créer une incohérence communicative: en tant que ND global, il dominerait communicativement le nœud dominant du rhème, mais, en tant qu'élément du rhème, il dépendrait communicativement de ce nœud dominant. La contrainte C_2, en fonction de la définition que nous avons donnée du chemin communicatif, nous assure que le ND que nous sélectionnons grâce à la règle R_2 n'appartiendra pas lui-même au rhème car, dans ce cas, il ne se situerait pas sur un chemin communicatif valide.

3.2.5 Règle R_3 de choix du ND

R_3 Le ND d'une RSém est le nœud dominant du thème.

Il est aussi possible d'attribuer la fonction de ND de la RSém au nœud dominant du thème. Cela est présenté comme un troisième choix mais une telle approche pourrait être discutée. Peut-être, notamment, la formulation de ces règles et leur ordonnancement devraient-ils être ajustés selon les langues considérées[7].

3.2.6 Règle R_4 de choix du ND

R_4 Le ND d'une RSém est un foncteur tel qu'un de ses actants est le nœud dominant du thème.

---

[7] Voir ce qui est dit à propos de ces règles dans la conclusion de ce chapitre.

Cette règle est le pendant de la règle R_2 pour les cas où l'on cherche à déterminer le ND à partir de la composante thématique.

3.2.7 Règle R_5 de choix du ND

R_5 Le ND d'une RSém peut être n'importe quel nœud de la RSém pourvu que les contraintes C_1 et C_2 soient respectées.

Cette règle permet, "en dernier recours", de déterminer un ND en ne privilégiant ni le domaine rhématique ni le domaine thématique, mais en respectant les deux contraintes de base.

Prenons maintenant un exemple pour illustrer le fonctionnement de ces règles.

Soit le texte suivant, tiré de Sue 1845:194-195:

> Quelquefois, bien rarement, car elle y mettait la plus excessive discrétion, la pauvre fille avait été se promener avec Agricol et sa mère; pour la couturière ça avait été des fêtes sans pareilles, elle avait veillé bien des nuits, jeûné bien des jours pour pouvoir s'acheter un bonnet passable et un petit châle, afin de ne pas faire honte à Agricol et à sa mère; ces cinq ou six promenades, faites au bras de celui qu'elle idolâtrait en secret, avaient été les seuls jours de bonheur qu'elle eût jamais connus.

et soit la figure IV-5 qui est une RSém de la phrase que nous avons soulignée:

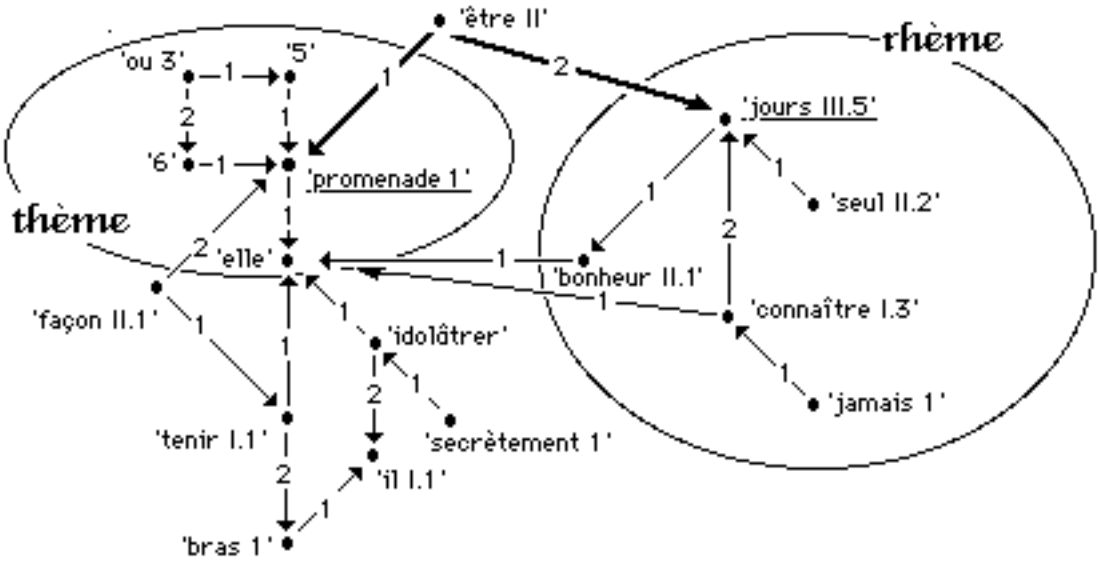


Figure IV-5
RSém avec chemin communicatif

Remarque: Pour simplifier cette figure, nous n'y faisons pas figurer les composantes de la RSém qui sont grammaticalisées dans le Texte.

Un tel réseau, compte tenu de ses structures sémantique et communicative, ne possède qu'un seul chemin communicatif valide, indiqué ici en gras. La phrase soulignée pourrait avoir été produite en sélectionnant le nœud 'être**II**' comme nœud dominant, c'est-à-dire en utilisant les règles R_2 ou R_4 (ou, par voie de conséquence, R_5). Une autre possibilité serait d'utiliser la règle R_3, ce qui pourrait donner une phrase comme (6) où la réalisation à la surface du ND (qui se trouve alors être en même temps le nœud dominant du thème) a été soulignée:

(6) *Elle s'était promenée cinq ou six fois, au bras de celui qu'elle idolâtrait en secret, ce qui avait été les seuls jours de bonheur qu'elle eût jamais connus.*

4 CONSEQUENCES POUR LA BONNE FORMATION DES RSEM

Nous allons maintenant examiner les liens existant entre les règles qui viennent d'être formulées et la notion de *règle sémantique de bonne formation* (présentée en II:1). Nous ferons aussi une incursion dans le domaine de l'étude topographique des RSém, pour ce qui a trait à leur structure communicative.

4.1 Structure communicative et bonne formation des RSém

Une caractéristique importante de l'ensemble de contraintes et règles qui vient d'être introduit est que l'on peut en déduire des contraintes sur la structuration communicative des RSém. Un réseau qui, compte tenu de sa structure sémantique et de sa structure communicative, ne permettrait pas de déterminer au moins un ND en utilisant les règles R_1-R_5 et en respectant les deux contraintes C_1 et C_2 ne serait pas une RSém valide. Ainsi, la figure IV-6 reprend la structure sémantique de la RSém présentée dans la figure IV-5 en lui appliquant une structure communicative non valide:

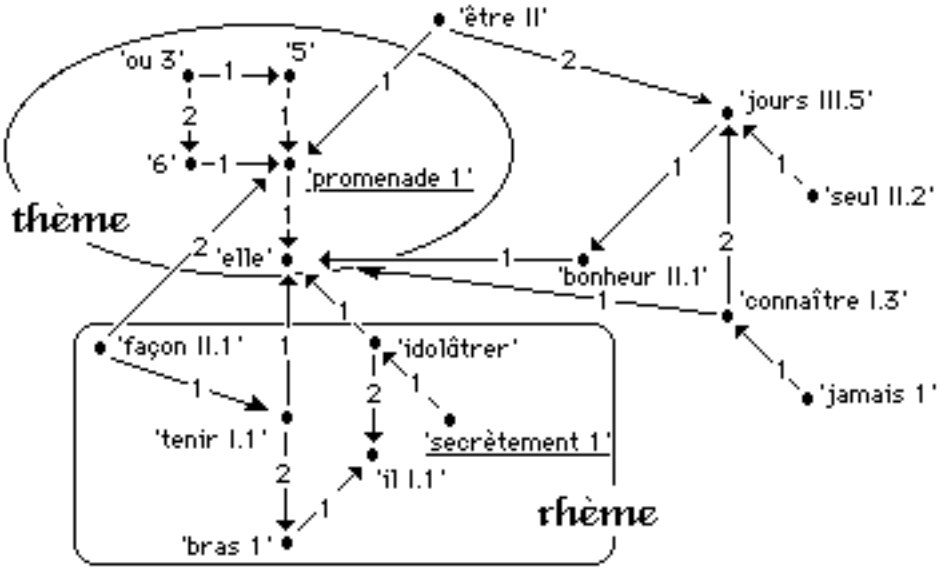


Figure IV-6
RSém avec structure communicative non valide

Aucune des règles R_1-R_5 ne permet de calculer un ND pour cette RSém car elle est construite de telle façon qu'il n'existe pas de chemin communicatif entre le nœud dominant du rhème ('secrètement**1**') et celui du thème ('promenade**1**'). DANS CETTE FIGURE, IL N'Y A PAS D'HARMONIE ENTRE LES STRUCTURES SEMANTIQUE ET COMMUNICATIVE. On pourrait dire que, dans un tel cas, la structure communicative n'aide en rien pour déterminer la construction de tel ou tel arbre syntaxique profond.

Il est possible de déduire de C_1, C_2 et R_1-R_5 quelques règles de bonne formation des RSém. Il est clair, notamment, que toute RSém devra posséder au moins un chemin communicatif unissant les nœuds dominants du thème et du rhème et sur lequel on peut isoler au moins un nœud qui pourrait résulter en un verbe au niveau syntaxique profond. Cela représente une contrainte assez forte sur la structure communicative, contrainte qui nous force à écarter des cas comme la RSém de la figure IV-6. Il faut bien voir, cependant, que les règles énoncées permettent au modèle de construire, à partir de la RSém, une phrase ayant une structure "standard". Il est toujours possible, cependant, de produire un Texte qui n'est pas nécessairement une phrase type, un Texte, notamment, dont la RSyntP ne possède pas de racine verbale. Ouvrons une parenthèse à propos de la "validité" des RSém.

On pourrait discuter le fait de savoir si l'on doit véritablement parler de RSém non valide dans le cas où il n'est pas possible d'associer à une telle représentation une RSyntP au moyen des règles de la grammaire. Ne serait-ce pas plutôt un cas de "cul-de-sac" comme on peut en rencontrer beaucoup en faisant fonctionner le modèle, à travers tous ses niveaux de représentation, pour produire une correspondance Sens $\Leftrightarrow$ Texte? En effet, il est tout à fait normal de considérer que la mise en jeu des règles de grammaire permet de parvenir, dans bien des cas, à des configurations qui ne peuvent donner lieu à une correspondance avec un niveau adjacent. Ce type de problème est somme toute très théorique et nous nous en tenons,

quant à nous, à un point de vue très simple: une soi-disant "représentation" R_n, d'un quelconque niveau, qui ne peut trouver sa place dans une correspondance donnée $Sens_n \Leftrightarrow Texte_n$ n'est pas considérée comme une véritable représentation car justement elle ne REPRESENTE rien (aucun $Texte_n$). On peut donc distinguer trois cas de figure possibles pour ce qui est des différents niveaux de représentation:

1) les configurations (sémantiques, syntaxiques profondes, etc.) bien formées et qui peuvent prendre place dans une correspondance donnée $Sens_n \Leftrightarrow Texte_n$ — elles sont, à ce titre, des représentations valides du $Texte_n$;
2) les configurations bien formées mais qui ne peuvent prendre place dans une correspondance $Sens_n \Leftrightarrow Texte_n$ — elles ne sont que de simples constructions théoriques d'un niveau donné (réseaux sémantiques, arbres syntaxiques profonds, etc.) sans être de véritables représentations;
3) les configurations qui ne respectent pas les règles de bonne formation — elles sont donc, par nature, invalides.

Les règles statiques de la grammaire ne doivent pas nécessairement être équipées pour éviter de produire le deuxième type de configuration mentionné ci-dessus. Ces règles jouant au niveau de "micro-structures" (dépendance sémantique ou syntaxique entre deux nœuds, etc.), elles ne peuvent prendre en charge les configurations globales que forme la combinaison de ces micro-structures. En réalité, c'est aux règles procédurales du modèle de contrôler au maximum la validité des représentations construites à partir des règles de grammaire. Quand on se situe au niveau des procédures, on sera donc toujours confronté à certaines options qu'il faut pouvoir caractériser pour choisir, dans un cas donné, la meilleure solution possible compte tenu des exigences de la grammaire impliquée. De même, dans un cadre de génération, il est certain qu'un module prélinguistique devrait être capable de construire une RSém comme celle de la figure IV-5 de préférence à celle de la figure IV-6. Un tel cas

pourra être traité sans trop de difficulté dans la mesure où la notion de chemin communicatif nous donne un critère assez solide de contrôle de la validité d'une RSém; nous ne prétendons pas, bien sûr, qu'une telle notion suffise à régler tous les problèmes de bonne formation de la structure communicative des RSém.

Jusqu'à présent, les articulations communicatives que nous avons considérées étaient basées sur une partition très claire des RSém, utilisant deux sous-réseaux, thématique et rhématique, absolument disjoints. Il est cependant logiquement possible d'isoler cinq autres cas de structuration communicative des RSém: 1) les RSém sans thème, 2) les RSém sans rhème, 3) les RSém où le rhème est inclus dans le thème, 4) les RSém où le thème est inclus dans le rhème, et enfin, 5) les RSém dont le thème et le rhème possèdent une intersection. Nous allons examiner ces cas particuliers à la lumière des nouvelles contraintes sur la structure communicative que nous avons introduites. Notons que nous partirons ici d'exemples de phrases, et non des RSém elles-mêmes, pour examiner les différents cas de figure possibles. Cela ne doit cependant pas masquer le fait que ces phrases ne sont que les réalisations des configurations de Sens, qui seules nous intéressent ici. Fidèle à notre approche synthétique, nous nous préoccupons en fait de trouver quelles phrases pourraient exprimer certaines articulations communicatives et non de trouver quelles articulations communicatives se trouvent exprimées par telle ou telle phrase.

4.2 Examen des différentes configurations communicatives possibles

4.2.1 Absence de thème

Certaines RSém sont soit entièrement rhématiques soit ont une composante qui, sans être thématique, ne s'oppose pas communicativement au rhème. De telles RSém correspondent à des phrases ne présentant aucune dichotomie communicative, des phrases où toute l'information semble posséder une importance

maximale au niveau de la communication. Firbas 1987:150 cite comme exemple d'un tel cas:

(7) *Le vase!*

qui, quelle que soit la portée contextuelle véritable de l'exclamation (désir de savoir où se trouve le vase, désir de manifester sa colère de voir que quelqu'un a peint le vase de Chine en vert, etc.) ne présente effectivement pas d'opposition thème/rhème évidente. Une telle phrase SANS VERBE exprime bien le contenu d'une RSém particulière, sans articulation communicative fondée sur la dichotomie thème/rhème puisque toute l'information est rhématisée. Cela ne signifie pas que la RSém de (7) ne possède AUCUNE articulation communicative d'une quelconque nature. Certains composants de cette RSém, qui sont implicites dans (7), devraient aussi être considérés (cf. une signification complexe comme 'Sachant que le destinataire sait pourquoi je dis ceci, je m'exclame X'). On peut noter qu'une phrase comme:

(8) *Je voudrais bien savoir pourquoi un imbécile a peint ce vase en vert.*

présente, elle, une articulation communicative (thème/rhème) très claire. Le fait que, dans un contexte d'énonciation donné, (7) puisse véhiculer la même information n'est pas pertinent pour nous ici: (7) n'est pas une *paraphrase langagière* de (8), ou, en d'autres termes, la RSém de (7) est différente de la RSém de (8). Ces deux phrases pourraient être ce que nous avons appelé dans la Partie **1** des *paraphrases cognitives*; dans ce cas, les informations liées aux significations exprimées dans (8) et non exprimées dans (7) seraient accessibles par un autre médium que la seule langue. Il n'y a donc rien de gênant à considérer que ces deux phrases présentent des articulations communicatives de nature très différente[8].

---

[8] L. Iordanskaja met très sérieusement en doute le fait que l'on puisse parler de structure communicative pour des énoncés correspondant à des exclamations (pour lesquels le critère de questionnement isolant le rhème peut difficilement être

4.2.2 Absence de rhème

Une RSém entièrement thématisée semble impossible à imaginer. Comment concevoir en effet qu'un thème puisse exister s'il n'est pas communiqué afin de contraster avec un rhème? Le thème, en tant qu'élément le moins important communicativement, le plus lié au contexte d'énonciation, ne se conçoit que par opposition avec le rhème. Puisqu'un énoncé a pour finalité de communiquer, le rhème ne saurait être un élément manquant dans la structure communicative. Il existe cependant des cas limites, comme l'illustre ce passage de *Cyrano de Bergerac* d'E. Rostand (Acte premier, Scène IV):

(9) CYRANO
Tournez!
– Ou dites-moi pourquoi vous regardez mon nez.
LE FACHEUX, *ahuri.*
Je ...
CYRANO, *marchant sur lui.*
Qu'a-t-il d'étonnant?

Dans cet exemple, la phrase soulignée, incomplètement réalisée, est privée de sa composante rhématique. Cette phrase est à la fois syntaxiquement et communicativement incomplète. Elle s'insère cependant très bien dans le dialogue car, bien que l'énoncé soit en apparence agrammatical, un certain message est transmis; ce message se caractérise justement par une absence d'élément rhématique ou plutôt par la présence d'un Sens incomplet. La dichotomie communicative existe donc ici sous une forme un peu particulière: c'est-à-dire qu'il nous faut considérer une opposition thème/rhème interrompu. Ce rhème interrompu est signalé dans notre exemple par les points de suspension. Pour être exact, il faut

---

appliqué). Cela pourrait concerner toutefois les exclamations "véritables". Pour les besoins de la présente discussion, nous pouvons considérer que le caractère exclamatif de la phrase (7) est ici secondaire. On pourra notamment la voir comme une réponse possible à *"Qu'est-ce qui cloche, mon petit Telesphore?"*

tenir compte du fait que cet élément du dialogue, tel que présenté par E. Rostand, est ambigu. Le Fâcheux peut véritablement avoir été interrompu dans sa phrase, ou il peut s'être interrompu lui-même volontairement car il n'avait rien de plus à dire que *"Je..."*. Une telle distinction est nettement perçue à l'oral et un acteur jouant Cyrano devra faire un choix entre une des deux interprétations possibles. Si le Fâcheux a véritablement été interrompu, on peut considérer que nous avons là une absence pure et simple de rhème. Il n'y a donc pas, dans ce cas, de RSém valide pour une telle phrase. Par contre, si le Fâcheux a planifié cette interruption, nous avons un RHèME ZéRO et une phrase communicativement complète possédant une RSém valide qui pourrait être (sous une forme simplifiée):

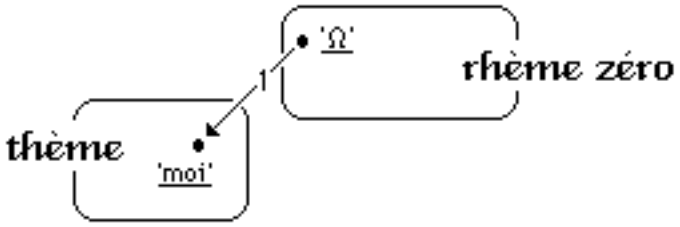


Figure IV-7
RSém avec rhème zéro

Quel que soit le point de vue adopté quant à la description de tels exemples, nous préférons les ignorer et considérerons la présence du rhème comme nécessaire. Même si de tels cas peuvent finalement trouver une description à partir des outils fournis par un modèle linguistique, ils ne doivent pas, à cause de leur caractère anormal, prendre une importance démesurée lorsque l'on cherche à déterminer les caractéristiques de l'objet de notre description linguistique.

4.2.3 Inclusion du rhème dans le thème

Il n'est pas vraiment difficile de trouver des exemples illustrant une telle configuration qui est rendue à la surface par une construction réflexive:

(10)

Q: *Et savez-vous, mon bon C, à qui ces malotrus font du tort en agissant de la sorte?*
C: *Ma foi, Monsieur Q, ...*
Q: *Ces gens sans éducation nuisent à eux-mêmes! Tout simplement!*

Nous pensons que la phrase soulignée peut légitimement être représentée au niveau sémantique de la façon suivante:

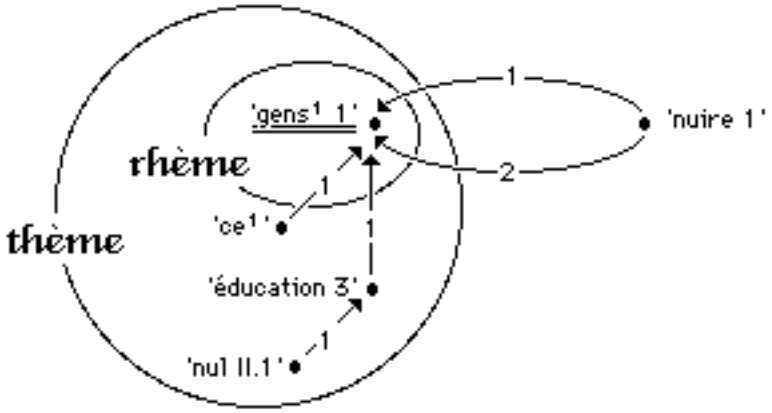


Figure IV-8
RSém présentant une inclusion du rhème dans le thème

Une telle configuration ne doit pas surprendre. Rappelons que nous situons l'étude de la structure communicative au niveau de la RSém. Il est certain que si nous nous situions au niveau de la phrase comme telle, il n'y aurait ici aucune inclusion à envisager puisque les éléments thématiques et rhématiques sont "éclatés" lors de l'arborisation (cf. la co-référence de *gens* et *eux-mêmes*). L'absence d'inclusion "linéaire" du rhème dans le thème n'est donc pas un aspect pertinent pour nous; elle n'est qu'un reflet de la richesse des moyens linguistiques utilisés pour refléter à la surface la structure

communicative du Sens. Seule compte ici la véritable inclusion sémantique. Or, nous pensons que notre exemple, énoncé dans le contexte approprié, manifeste clairement un cas d'inclusion du rhème dans le thème.

Admettre comme possible une telle inclusion pose un sérieux problème pour ce qui a trait à notre utilisation de la notion de chemin communicatif. En effet, en cas d'inclusion, le nœud dominant du rhème est nécessairement dominé communicativement par le nœud dominant du thème. Au mieux, comme c'est le cas dans notre exemple, un seul et même nœud dominera à la fois le thème et le rhème. On peut cependant admettre que nous sommes ici en présence d'un des cas limites où, si les deux extrémités du chemin communicatif entretiennent une relation de dépendance communicative, il est tout de même possible de trouver un élément de transition entre les deux composantes thématique et rhématique du réseau.

Comment, alors, rendre compte de cas comme la figure IV-9 ci-dessous, qui ne semble pas présenter de transition entre la composante thématique et la composante rhématique mais que l'on pourrait vouloir associer à la phrase soulignée de (11)?

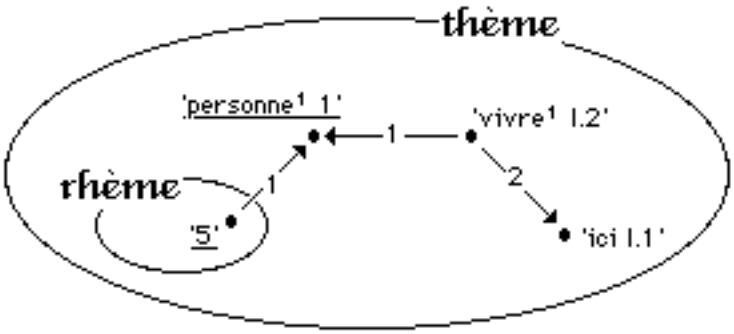


Figure IV-9
RSém sans chemin communicatif

(11)

Q: *Combien de personnes vivent dans cette maison?*
C: *Cinq personnes vivent ici!*

Dans un tel cas, que le rhème soit considéré ou non comme une partie du thème, la RSém de la phrase soulignée ne permettra de toute façon pas de sélectionner le nœud 'vivre^{1}**I.2**' comme ND tout en respectant les conditions C_1 et C_2 de choix du ND. On voit clairement dans cette RSém que, même si les deux sous-réseaux **thème** et **rhème** étaient dissociés, le nœud 'vivre^{1}**I.2**' ne serait pas situé sur un chemin communicatif unissant leurs nœuds dominants. Pourtant, c'est le Sens 'vivre^{1}**I.2**' qui correspond au nœud syntaxique sommet de la RSyntP de (11). (Rappelons qu'entre autres caractéristiques, le ND d'une RSémR est en relation directe avec le sommet de l'arbre syntaxique profond associé.) Pour résoudre le problème posé par la RSém de la figure IV-9, il faut noter que la structure syntaxique de notre exemple ne reflète tout simplement pas la structure sémantico-communicative considérée. *Cinq* peut être perçu comme rhème à cause du contexte d'énonciation, ce qui est tout à fait normal en français, mais seule une phrase comme (12) serait un cas de réelle manifestation lexicale et grammaticale de l'opposition thème/rhème présente dans la RSém. Nos procédures de calcul du ND nous permettraient de produire la hiérarchie communicative correspondante, pour peu que l'on considère une RSém légèrement modifiée: avec un rhème dissocié du thème.

(12) *Les personnes vivant ici sont au nombre de cinq.*

Nous avons ici une illustration typique de la nature particulière des règles de choix du ND.

La transgression des règles de choix du ND n'entraîne pas la production de phrases agrammaticales; elle n'entraîne pas non plus NECESSAIREMENT une incohérence communicative au niveau du Texte. Contrairement aux règles sémantiques "impératives" de la grammaire, qui spécifient les bonnes correspondances RSém ⇔ RSyntP, ce sont des règles "de préférence", qui permettent d'utiliser au mieux l'information communicative. LES CONSIGNES QU'ELLES DONNENT DOIVENT ETRE RESPECTEES DE LA MEILLEURE FAÇON POSSIBLE.

4.2.4 Inclusion du thème dans le rhème

Un tel cas pourrait être illustré par la figure IV-10 qui est la RSém de la phrase (13).

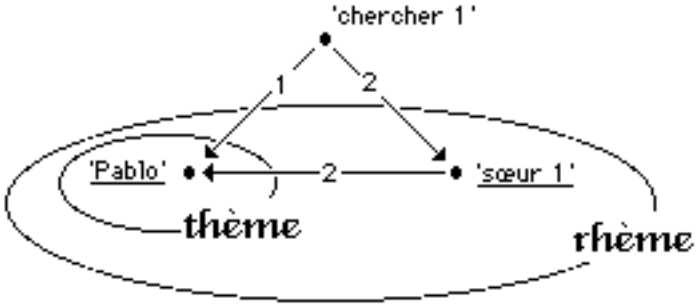


Figure IV-10
RSém avec inclusion du thème dans le rhème

(13) *Pablo cherche sa sœur.*
← 'Pablo cherche qui?'

Là-encore, comme dans le cas de figure précédent, il convient de distinguer l'INCLUSION au niveau sémantique et l'ENCHâSSEMENT au niveau (linéaire) du Texte. L'inclusion sémantique du thème dans le rhème est tout à fait possible et, comme dans le cas inverse, la contrainte sur le chemin communicatif doit être ici interprétée de façon un peu libérale pour permettre le choix du ND approprié.

Les quelques cas examinés jusqu'à présent montrent que la notion de niveau de représentation, qui joue un rôle tellement important dans l'approche Sens-Texte, intervient aussi pour ce qui a trait à la structuration communicative des énoncés. Le fait de considérer une stratification de la structure communicative permet de mieux rendre compte des conflits pouvant exister entre les réelles intentions communicatives du locuteur, rendues au niveau de la RSém, et les différentes contraintes qu'imposent la grammaire et le lexique de la langue. Pour un Texte donné, ce type de conflit peut se résoudre par une grande divergence entre la structuration sémantico-communicative et la structuration syntaxico-communicative.

4.2.5 Intersection du thème et du rhème

On peut illustrer ce dernier cas en complexifiant légèrement l'exemple précédent; cf. la figure IV-11 qui est la RSém de la phrase (14):

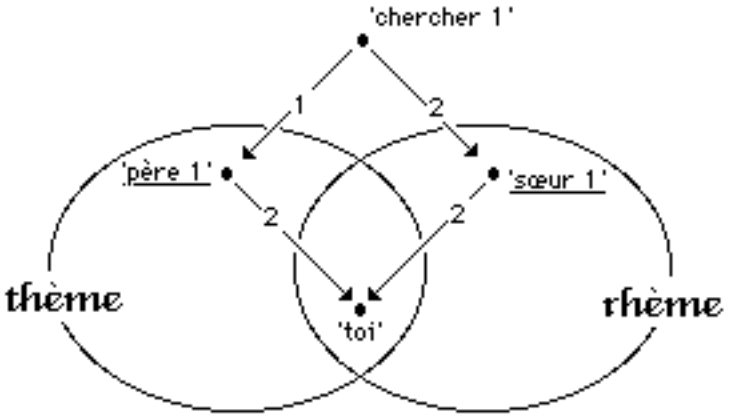


Figure IV-11
RSém avec intersection du thème et du rhème

(14) *Ton père cherche ta sœur.*
← 'Mon père cherche qui?'

Notre étude de l'utilisation des données communicatives dans le processus de transition sémantique ne fournit qu'une petite partie des informations nécessaires pour mieux représenter la structure communicative des Textes. Dans un cas comme celui-ci, par exemple, nous avons encore des doutes sur la nécessité de faire figurer une intersection des sous-réseaux **thème** et **rhème**. On notera ainsi que, dans notre exemple de la figure IV-5, un des nœuds ('elle') qui pourrait logiquement appartenir aussi bien au thème qu'au rhème est localisé uniquement dans le thème, car il est un actant sémantique du nœud dominant du thème. Dans ce cas, nous avons voulu avant tout simplifier la représentation mais il semblerait en fait que nous soyons là-aussi en présence d'une intersection thème-rhème.

4.2.6 Autres problèmes topographiques

Ce dernier exemple clôt notre bref survol des différentes configurations communicatives que l'on peut logiquement rencontrer au niveau de représentation sémantique. Nous sommes bien sûr conscient que la discussion ne devrait pas s'arrêter là et qu'il conviendrait d'aller plus avant dans la spécification des règles de bonne formation communicative des RSém. Notre travail, qui tente d'expliciter une partie du rôle joué par la structure communicative dans la transition RSém ⇒ RSyntP, devrait apporter des éléments utiles à une telle recherche. Ainsi, un autre cas intéressant, qui mériterait une étude approfondie, est celui des RSém à structures communicatives multiples qui correspondent à des cas de conjonctions de phrases; examinons (15), qui est une adaptation de l'exemple utilisé plus haut en 3.2.7:

(15) Ces cinq ou six promenades, faites au bras de celui qu'elle idolâtrait en secret, avaient été les seuls jours de bonheur qu'elle eût jamais connus, mais même de si simples joies lui étaient maintenant à jamais interdites.

Une telle phrase correspond à la conjonction de deux phrases possédant chacune leur propre structure communicative. Le

comportement SYNTAXIQUE de la conjonction est clair. Il peut être décrit comme suit:

1) soit Δ_1, l'arbre syntaxique profond de la première phrase S_1 *"Ces cinq ou six promenades..."*;
2) soit Δ_2, l'arbre syntaxique profond de la seconde phrase conjointe S_2 *"même de si simples joies..."*;
3) la RSyntP Δ de la phrase (15), au complet, sera:

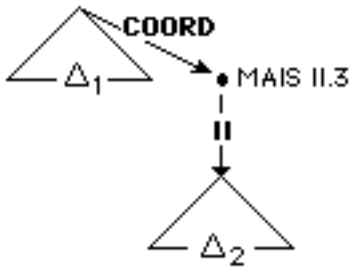


Figure IV-12
Représentation de la conjonction de phrases
au niveau syntaxique profond

MAIS**II.3** est donc un dépendant **COORD** du sommet de Δ_1 (RSyntP de S_1) et a comme second actant syntaxique profond le sommet de Δ_2 (RSyntP de S_2).

Au niveau sémantique, cependant, il est bien plus problématique de représenter la "conjonction" des deux sous-réseaux Σ_1 et Σ_2 correspondant aux RSém des deux phrases considérées S_1 et S_2. En effet, il semble évident que, dans le cas d'une véritable conjonction de phrases, le nœud sémantique représentant le Sens de la conjonction devrait avoir pour premier et second actants les nœuds dominant les RSém des deux phrases conjointes. Ce qui veut dire que, dans notre exemple, 'mais**II.3**' devrait avoir comme premier actant 'être**II**' et comme second actant 'interdit$^{\mathbf{1}}$**1**'.

Si les RSém connectées à 'mais**II.3**' possèdent un ND, cela signifie qu'elles sont hiérarchisées. La conjonction de phrase ne peut donc être complètement formalisée que sur des RSém déjà

hiérarchisées qui ont perdu une partie de leur puissance représentationnelle. (Elles représentent sémantiquement un ensemble réduit de paraphrases.) La solution à ce problème pourrait consister en une claire séparation, au véritable niveau de représentation sémantique, des différents réseaux impliqués. Il faudrait ainsi considérer la figure IV-13, ci-dessous, comme la RSém de (15) aussi bien que de (16), qui en est une paraphrase:

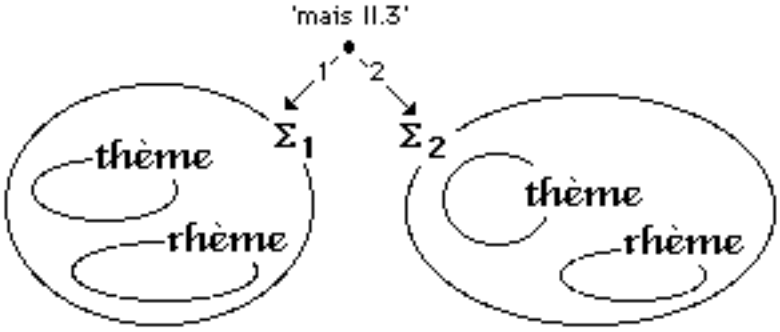


Figure IV-13
RSém avec sous-réseaux conjoints

(16) Elle s'était promenée cinq ou six fois, au bras de celui qu'elle idolâtrait en secret, ce qui avait été les seuls jours de bonheur qu'elle eût jamais connus, mais même de si simples joies lui étaient maintenant à jamais interdites.

Une fois les RSém Σ_1 et Σ_2 hiérarchisées, il devient possible de connecter ces réseaux directement par des relations actancielles sémantiques, comme l'illustre la RSémR de la phrase (15), où $\Sigma^r{}_1$ et $\Sigma^r{}_2$ correspondent aux RSémR de S_1 et S_2:

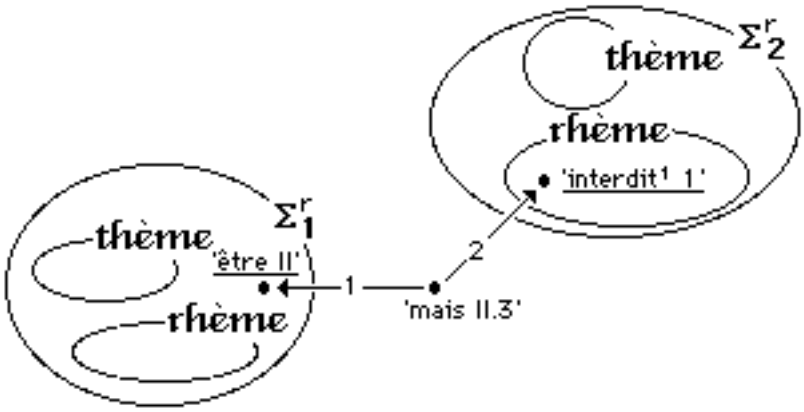


Figure IV-14
RSémR de (15)

Cela ne fait que suggérer une approche possible pour le traitement des RSém mettant en jeu plusieurs structures communicatives distinctes. L'étude de telles structures complexes sort du cadre de ce travail et nous arrêterons donc ici cette digression.

Pour conclure cette section consacrée à la notion de ND, nous tenons à souligner que, dans ce travail, nous nous sommes surtout attaché à présenter cette notion à travers sa fonction dans le processus de transition RSém ⇒ RSyntP; cette fonction est décrite de façon détaillée. En revanche, les règles de détermination du ND qui viennent d'être proposées ne sont, elles, qu'un point de départ pour une étude plus poussée et systématique. Une telle étude doit réunir plus de données sur ce que nous appelons la *topographie des RSém*: hiérarchisation des réseaux en sous-réseaux, répartition des nœuds dominants de ces sous-réseaux, liens avec les structures sémantique, communicative et rhétorique, etc.

CHAPITRE V

Hiérarchisation: parcours des réseaux sémantiques

Nous abordons maintenant la partie véritablement "procédurale" de notre exposé. Dans ce chapitre, nous allons tout d'abord limiter la notion de parcours des graphes au cas particulier des réseaux sémantiques de la TST. Nous décrirons ensuite les différentes étapes d'un parcours sémantique et donnerons quelques règles générales de contrôle de ce parcours (qui tiennent compte de la façon dont, dans la pratique, il influe sur la réalisation de la transition RSém ⇒ RSyntP). Finalement, nous établirons certains parallèles entre notre approche et les propositions faites par J. Sowa (entre autres) dans le cadre de la génération de textes à partir de graphes conceptuels.

1 CARACTERISATION DE LA NOTION DE PARCOURS APPLIQUEE AUX RSEM

La hiérarchisation d'une RSém passe par une première étape de choix d'un ND. Cette étape a été traitée en détail au chapitre précédent; il nous reste maintenant à expliquer comment la dépendance communicative proprement dite est calculée. Rappelons que nous proposons un algorithme de parcours des RSém dont le

point de départ est le ND sélectionné; cet algorithme permet de hiérarchiser la RSém de la façon suivante:

> En passant par un arc <n_1,n_2>, l'algorithme de parcours de la RSém crée une relation de dépendance communicative $n_1 \rightarrow n_2$ s'il traverse le nœud n_1 avant de traverser le nœud n_2.

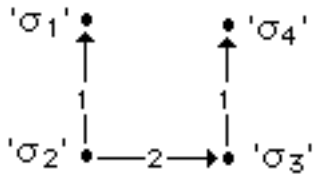
Il apparaît immédiatement qu'une telle règle s'harmonise avec la définition même de ND donnée plus haut. En effet, le ND étant le point de départ du parcours sémantique, il domine communicativement tous les autres nœuds: il est traversé par le parcours avant tous les autres nœuds.

Rappelons que le parcours d'un réseau sémantique doit passer une et une seule fois par chacun des arcs de cette représentation — cf. la définition de la notion de *parcours sémantique* en III:3. Le parcours doit toujours être construit à partir de tronçons connectés, ce que nous pouvons reformuler formellement de la façon suivante:

> Les *parcours sémantiques* sont des parcours connectés: lorsqu'un parcours sémantique a traversé un nœud n_1, il peut se poursuivre soit en traversant un nœud n_2 lié à n_1, soit, par une sorte de retour en arrière, en traversant un nœud n_3 lié à tout autre nœud n_i déjà traversé.

<u>Remarque</u>: Par défaut, lorsque nous disons qu'un nœud d'une RSém est "lié" à un autre, cela signifie "lié par une relation sémantique" (c.-à-d. actancielle).

La contrainte qui vient d'être énoncée sur les parcours sémantiques nous assure que les parcours théoriquement envisageables pour un réseau comme:



ne sont qu'au nombre de huit, c'est-à-dire:

{ ('σ_1','σ_2') , ('σ_2','σ_3') , ('σ_3','σ_4') }
{ ('σ_2','σ_1') , ('σ_2','σ_3') , ('σ_3','σ_4') }
{ ('σ_2','σ_3') , ('σ_2','σ_1') , ('σ_3','σ_4') }
{ ('σ_2','σ_3') , ('σ_3','σ_4') , ('σ_2','σ_1') }
{ ('σ_3','σ_4') , ('σ_3','σ_2') , ('σ_2','σ_1') }
{ ('σ_3','σ_2') , ('σ_3','σ_4') , ('σ_2','σ_1') }
{ ('σ_3','σ_2') , ('σ_2','σ_1') , ('σ_3','σ_4') }
{ ('σ_4','σ_3') , ('σ_3','σ_2') , ('σ_2','σ_1') }

Ainsi, pour un tel réseau, un parcours non connecté comme:

{ ('σ_1','σ_2') , **('σ_4','σ_3')** , ('σ_2','σ_3') }

ne serait pas admissible. Rappelons qu'il n'est pas nécessaire que le parcours s'effectue en suivant l'orientation de arcs de la représentation sémantique: il peut se faire dans le sens des flèches ou à contre-sens.

Même si la contrainte de connectivité réduit considérablement le nombre de parcours qui peuvent être effectués sur un même réseau, il n'en reste pas moins que, pour une structure sémantique relativement complexe, de très nombreux parcours risquent d'être envisageables. Ainsi, même pour un réseau aussi simple que:

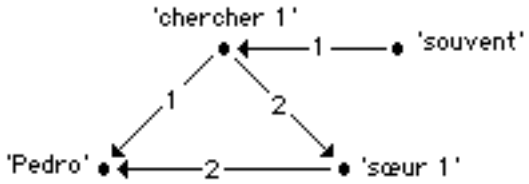


Figure V-1
Structure sémantique à parcourir

il est possible d'effectuer quatre-vingts parcours connectés différents. Notre stratégie de parcours des RSém est cependant suffisamment

élaborée pour permettre d'effectuer, de façon efficace, des choix pertinents entre les différents parcours possibles.

Deux aspects complémentaires du parcours sémantique permettent de limiter le nombre de parcours possibles.

Premièrement, le ND nous donnant le point de départ du parcours, la spécification du ND limite considérablement le choix des parcours. Ainsi, dans l'exemple de la figure V-1, le nombre des parcours possibles se distribue de la façon suivante en fonction des ND sélectionnés[1] :

'chercher**1**' ⇒ 32 parcours
'Pedro' ⇒ 20 parcours
'sœur**1**' ⇒ 20 parcours
'souvent' ⇒ 8 parcours

Deuxièmement, nous définissons en réalité le parcours d'une RSém comme une opération complexe qui peut être décomposée en deux étapes SUCCESSIVES:

1) un *parcours descendant*: c'est-à-dire, à partir d'un nœud sémantique donné, un parcours qui va suivre l'orientation (prédicat → argument) des arcs sémantiques;

2) un *parcours ascendant*: c'est-à-dire, à partir d'un nœud sémantique donné, un parcours qui ira à contre-sens de l'orientation des arcs sémantiques.

Cela établit une contrainte supplémentaire sur la stratégie de parcours sémantique qui permet encore mieux de contrôler le nombre des parcours différents menant à la hiérarchisation d'une même

---

[1] Tous ces résultats ont été obtenus en faisant fonctionner un petit programme permettant de générer de tels parcours. Nous déconseillons au lecteur scrupuleux ou méfiant de tenter d'en faire le décompte "à la main".

RSém. L'algorithme de parcours qui va être décrit ci-dessous permettrait ainsi d'effectuer un seul parcours sémantique sur le réseau de la figure V-1 (au lieu de 32), une fois le nœud 'chercher**1**' sélectionné comme ND. Ce parcours serait:

P_{V-1} = { ('chercher**1**','Pedro') , ('chercher**1**','sœur**1**') ,
('sœur**1**','Pedro') , ('chercher**1**','souvent') }

Nous allons décrire chacune des étapes du parcours sémantique, parcours descendant puis parcours ascendant, en appliquant l'ensemble des procédures au cas plus intéressant de la figure IV-5 (chapitre précédent).

Notons qu'afin de faciliter la description de notre stratégie de parcours, nous adopterons la convention terminologique suivante: on appellera *gouverneurs d'un nœud sémantique* les foncteurs dont ce nœud est un actant sémantique.

2 PARCOURS DESCENDANT

Le parcours sémantique commence par une première phase de parcours des actants sémantiques du ND, qui se fait récursivement jusqu'à ce que tous les descendants de ce nœud aient été traversés. Nous avons appelé un tel parcours *parcours descendant* car il se propage récursivement à travers le réseau en suivant l'orientation des liens de dépendance sémantique. Ce parcours prendra tout naturellement fin puisque les RSém sont des réseaux acycliques. (Rappelons que cette propriété des réseaux sémantiques a été examinée en I:2.3.) L'ordre dans lequel les dépendants sémantiques d'un nœud sont traversés n'est théoriquement pas important; dans nos exemples, nous avons pour principe de respecter l'ordre de numérotation des actants (d'abord le premier actant, puis le deuxième, etc.). Nous faisons aussi ce parcours en profondeur d'abord car une telle stratégie nous semble plus aisée à manipuler. La figure V-2, ci-dessous, reprend la RSém de la figure IV-5 en visualisant le

déroulement du parcours descendant au moyen des deux conventions suivantes:

1) les arcs parcourus sont dessinés en pointillé;
2) les nœuds traversés lors du parcours sont signalés par un numéro qui indique, du même coup, l'ordre dans lequel le parcours a été effectué.

Remarquons qu'un nœud peut être étiqueté par plusieurs numéros différents s'il a été traversé plusieurs fois durant le parcours du réseau, en empruntant des voies différentes.

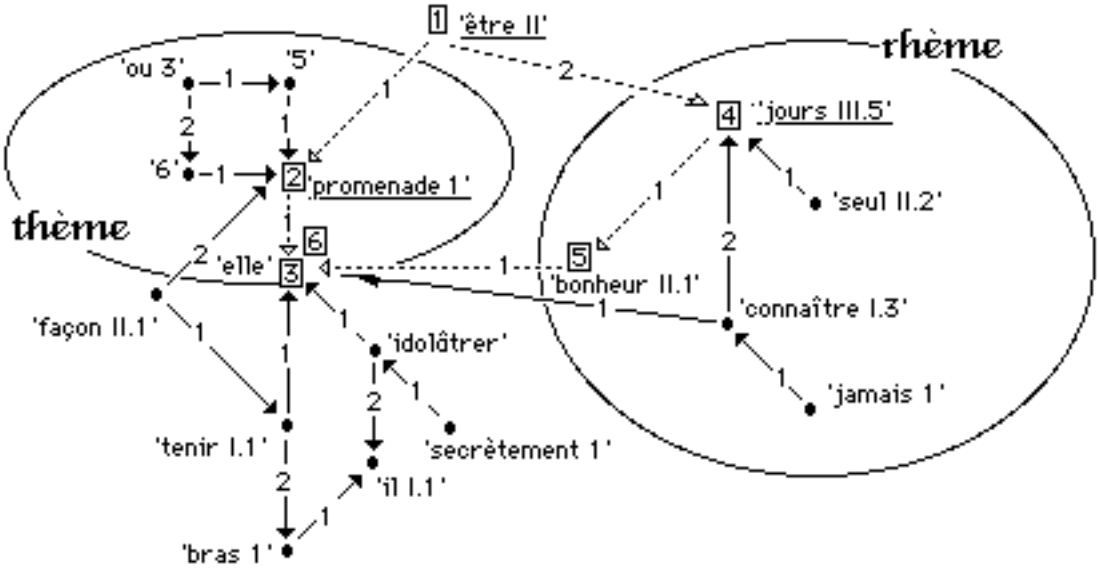


Figure V-2
Résultat du parcours descendant réalisé à partir du ND

Le parcours descendant apparaissant dans cette figure est le suivant:

P_d = {('être**II**',promenade**1**') , ('promenade**1**',elle') , ('être**II**','jours**III.5**') , ('jours**III.5**','bonheur**II.1**') , ('bonheur**II.1**','elle')}

Les arcs en pointillé correspondent, dans la pratique, à des parcours du réseau qui, ayant été effectués, ne devront plus être considérés. En effet:

Une contrainte de base sur le parcours sémantique est qu'il est impossible de passer deux fois par le même arc.

Lorsque l'on atteint un nœud qui a déjà été traversé, il faut arrêter là le parcours (pour le poursuivre, au besoin, sur un autre tronçon du réseau non encore examiné).

Le parcours descendant qui part du ND permet, dans notre exemple, de hiérarchiser communicativement les nœuds 'être**II**', 'promenade**1**', 'elle', 'jours**III.5**' et 'bonheur**II.1**'. On peut visualiser cette hiérarchie au moyen du graphe suivant:

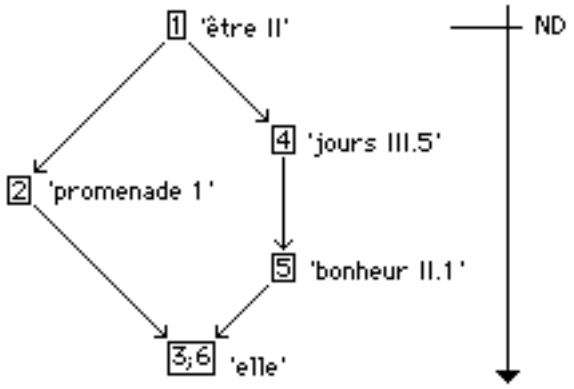


Figure V-3
Hiérarchie communicative résultant d'un parcours descendant

Dans un tel graphe, les arcs représentent un lien de dépendance communicative directe entre les nœuds. La hiérarchie construite ici est strictement orientée du haut vers le bas à partir du nœud dominant ('être**II**'). Dans cette hiérarchie, seuls comptent les liens de dépendance directement ou indirectement établis de même que l'orientation générale du système hiérarchique. Le fait, par exemple, que 'promenade**1**' soit plus directement relié au ND que 'bonheur**II.1**', ou le fait qu'ils gouvernent tous deux le même nœud 'elle', n'est pas du tout pertinent ici: aucune relation communicative n'est établie entre ces deux nœuds.

On peut se demander pourquoi il est nécessaire de spécifier ici un axe général vertical du haut vers le bas. La mention des relations de dépendance communicative deux-à-deux entre nœuds n'est-elle pas suffisante pour représenter une hiérarchie communicative au sein d'un réseau sémantique?

Il se trouve que la hiérarchie communicative est en fait un système complexe. Elle résulte des liens de dépendance communicative directs ou indirects établis entre les nœuds ET d'une orientation générale établie à partir du nœud dominant. (Rappelons que nous avons précédemment montré, en IV:1, comment cette orientation générale de la hiérarchie communicative pouvait être représentée par une métaphore physique où un réseau se trouve saisi par un de ses nœuds et soulevé vers le haut.) Ces deux paramètres de la hiérarchie pourraient être qualifiés respectivement de dépendance communicative locale et d'orientation générale de la hiérarchie communicative.

Nous dirons que, au niveau d'une hiérarchisation communicative, un nœud *domine communicativement* un autre s'il le domine directement ou indirectement par des liens de dépendance communicative locale dont aucun ne va à contre-sens de l'orientation générale de la hiérarchie.

Une conséquence importante de cela est qu'il convient de spécifier que notre définition de la notion de ND — le ND d'une RSémR est le seul nœud de cette représentation qui ne dépende communicativement d'aucun autre nœud — se base sur une notion de dépendance communicative telle que spécifiée par la hiérarchisation du réseau. Il ne s'agit donc pas d'une simple dépendance communicative locale.

Une autre conséquence est que nous essayons, dans la mesure du possible, de réserver l'emploi du terme *dominance (communicative)*, ou *dominer (communicativement)*, au cas de la dominance générale telle que définie ci-dessus; c'est-à-dire celle qui tient compte de l'orientation

générale de la hiérarchie communicative. Le terme *dépendance (communicative)*, ou *dépendre (communicativement)*, sera lui plutôt réservé aux cas de dépendance locale (directe). Il n'y a donc pas nécessairement, dans notre usage terminologique, de lien de réciprocité entre dépendance et dominance communicatives.

Il est nécessaire de présenter ainsi la hiérarchisation d'une RSém car l'on peut facilement construire des réseaux dont la hiérarchisation présentera des cas de dépendances locales fonctionnant à l'inverse de l'orientation hiérarchique générale. La RSém de la phrase (1) ci-dessous en est un exemple:

(1) *Agricol épouse la Marquise bien que Dagobert le désapprouve.*[2]
← 'Quelle est la chose que fait Agricol et qui va contre la volonté de Dagobert?'

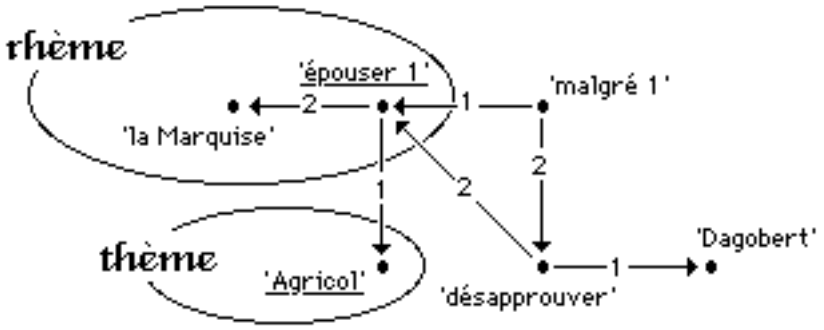


Figure V-4
RSém de
"Agricol épouse la Marquise bien que Dagobert le désapprouve."

Anticipons sur la suite de notre présentation du parcours sémantique et prenons pour acquis que notre méthode de hiérarchisation des

---

[2] Nous sélectionnons bien sûr ici l'interprétation problématique de (1) où *le* est co-référenciel à *épouse* et non à *Agricol*. La phrase (1) pourrait en fait être reformulée de façon moins élégante en: *"Agricol épouse la Marquise bien que Dagobert désapprouve cela."*

RSém nous permet de construire la hiérarchie communicative de la figure V-5, une fois 'épouser**1**' sélectionné comme ND. Cette hiérarchie est déduite du parcours suivant:

$P_{(1)}$ = { ('épouser**1**',Agricol') , ('épouser**1**','la Marquise') , ('épouser**1**','malgré**1**') , ('malgré**1**','désapprouver') , ('désapprouver','Dagobert') , ('désapprouver','épouser**1**') }

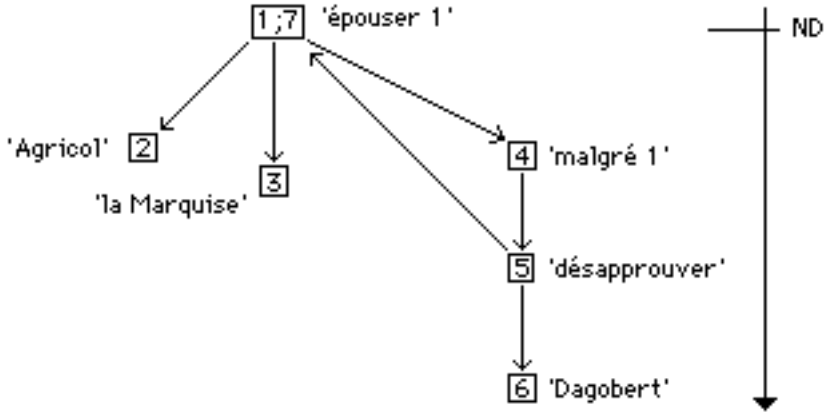


Figure V-5
Hiérarchie communicative établie sur la RSém de
"Agricol épouse la Marquise bien que Dagobert le désapprouve."

Cette hiérarchie fait apparaître une relation de dépendance communicative locale 'désapprouver' → 'épouser**1**'; pourtant, nous dirons qu'au niveau de la hiérarchie générale établie, c'est le nœud 'épouser**1**' qui domine indirectement 'désapprouver' car il le domine par une succession de liens de dépendance locale dont aucun ne s'oppose à l'orientation de la hiérarchie communicative.

Cette approche, selon plusieurs plans, de la hiérarchie communicative nous permet de rendre compte d'un phénomène assez complexe qui est reflété, lors de l'arborisation, au moyen de la construction de nœuds syntaxiques co-référenciels. Ce phénomène peut être décrit comme suit:

Au niveau d'un réseau ou d'un sous-réseau, un nœud sémantique peut dominer communicativement un autre nœud dont il dépend pourtant si leur relation est considérée à un niveau local.

Dans la RSémR de (1), le ND 'épouser**1**' domine communicativement 'désapprouver' de façon évidente: IL VIENT COMMUNICATIVEMENT AVANT; pourtant, il en dépend lorsque seul l'arc 'désapprouver'—2→'épouser**1**' est considéré.

Ces mises au point sur la hiérarchisation des RSém étant faites, nous pouvons poursuivre notre présentation du parcours sémantique avec la description de la seconde phase: le parcours ascendant.

3 PARCOURS ASCENDANT

Une fois exécuté un parcours descendant à partir du ND, il faut poursuivre la hiérarchisation de la RSém en parcourant les nœuds du réseau dont ce ND dépend sémantiquement, c'est-à-dire ses gouverneurs: comme ce parcours se fait en remontant à travers les liens de dépendance sémantique, il a été appelé *parcours ascendant*. Le parcours ascendant est lui-même une opération complexe qui peut être analysée de la façon suivante:

- soit 'X', un nœud sémantique dont 'Y' est un des gouverneurs;
- pour traverser 'Y', l'algorithme va tout d'abord atteindre ce nœud en créant une dépendance communicative entre 'X' et 'Y' qui est orientée à l'inverse de la dépendance sémantique: 'X' → 'Y';
- puis, 'Y' va à son tour être considéré, localement, comme une sorte de nœud dominant: ses actants sémantiques vont être parcourus récursivement à l'exception des actants qui, comme 'X', l'ont déjà été (parcours descendant) puis, ses gouverneurs seront parcourus les uns après les autres (parcours ascendant).

Comme pour les actants sémantiques, l'ordre dans lequel sont explorés les gouverneurs d'un nœud n'est théoriquement pas pertinent. Ici, la sélection doit vraiment être faite au hasard puisque l'on ne dispose pas de critère strict équivalent à celui de la numérotation des actants. On pourra cependant se baser ici aussi sur la numérotation des liens actanciels avec les gouverneurs pour ordonner les parcours ascendants: d'abord les gouverneurs dont le nœud considéré est un premier actant sémantique, etc. Nous reviendrons plus loin sur un certain nombre d'autres éléments de contrôle du parcours qui peuvent intervenir ici.

Remarque: Soulignons le fait que le parcours sémantique est structuré dans sa totalité de façon récursive. Il comprend un parcours descendant — qui se propage récursivement, puis un parcours ascendant — ce dernier pouvant lui-même comprendre des parcours sémantiques (avec parcours ascendants et descendants).

Une fois terminé le parcours ascendant à partir d'un nœud sémantique donné, la même opération doit être effectuée à partir de tous les actants et tous les descendants de ce nœud (qui ont déjà été traversés lors du parcours descendant). Cela va permettre de hiérarchiser complètement les réseaux en propageant le parcours de façon systématique à travers tous les liens de dépendance sémantique. Dans l'exemple que nous traitons dans cette section, le ND n'étant pas lui-même un actant sémantique (il n'a pas de gouverneur), le parcours ascendant commencera directement sur ses actants. La figure V-6 visualise le résultat du parcours ascendant P_a débutant au deuxième nœud traversé précédemment ('promenade**1**').

P_a = { ('promenade**1**','5') , ('5','ou**3**') , ('ou**3**',6') , ('6','promenade**1**') , ('promenade**1**','façon**II.1**') , ('façon**II.1**',tenir**I.1**') , ('tenir**I.1**','elle') , ('tenir**I.1**','bras**1**') , ('bras**1**','il**I.1**') , ('il**I.1**','idolâtrer') , ('idolâtrer',elle') , ('idolâtrer',secrètement**1**') }

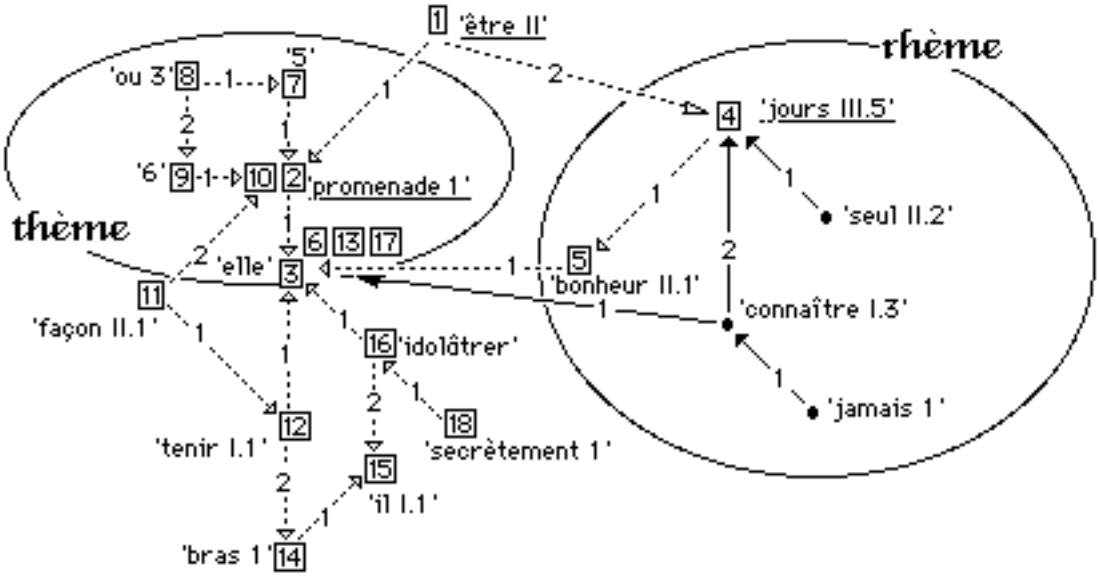


Figure V-6
Résultat du parcours ascendant
effectué à partir du nœud 'promenade**1**'

<u>Remarque</u>: Le premier tronçon de P_a, {('promenade**1**','5') , ('5','ou**3**') , ('ou**3**',6') , ('6','promenade**1**')}, aurait théoriquement pu être effectué dans un ordre différent, {('promenade**1**','6') , ('6','ou**3**') , ('ou**3**','5') , ('5','promenade**1**')}, puisque 'promenade**1**' est à la fois un premier actant de '5' et de '6'. On voit que les procédures de parcours devraient ici être contraintes pour que le premier actant de la conjonction sémantique 'ou**3**' soit traversé en priorité. Nous nous permettrons d'ignorer ici ce problème.

Le double parcours, parcours descendant et parcours ascendant, va être appliqué à toute la structure sémantique de la RSém, permettant ainsi d'effectuer une hiérarchisation complète. LA HIERARCHISATION SERA CONSIDEREE COMME ACHEVEE LORSQUE TOUS LES NOEUDS DU RESEAU AURONT ETE TRAVERSES AU MOINS UNE FOIS ET QUE TOUS LES ARCS AURONT ETE PARCOURUS UNE ET UNE SEULE FOIS. Dans nos représentations, cela se traduit par un réseau dont tous les nœuds sont étiquetés par un numéro (ou plusieurs) indiquant l'ordre de

parcours et dont tous les arcs sont en pointillé, comme l'illustre la figure ci-dessous:

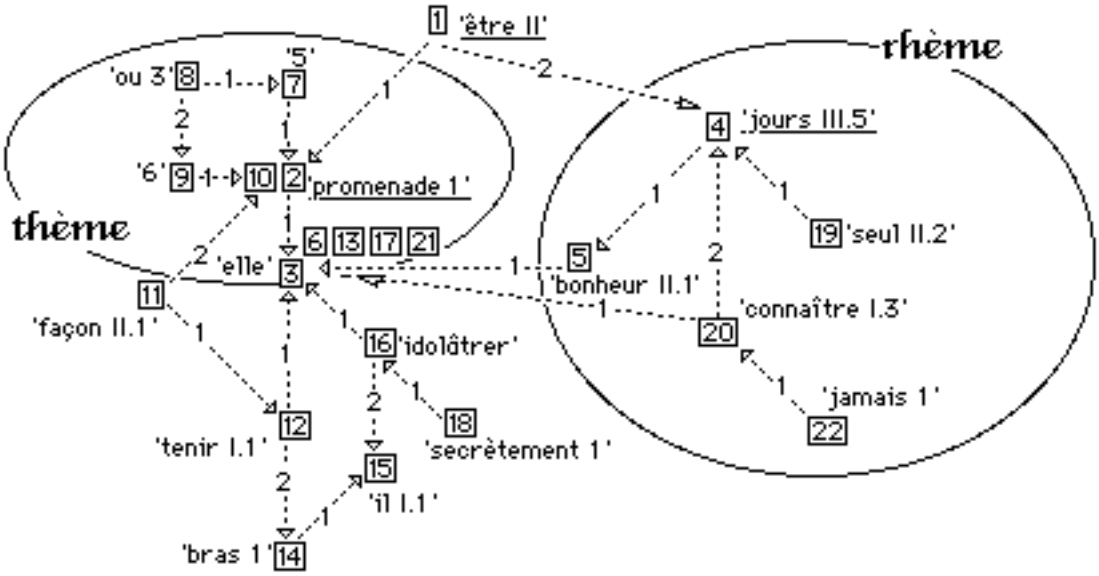


Figure V-7
RSém entièrement hiérarchisée

Nous récapitulons maintenant dans sa totalité le parcours apparaissant dans la figure V-7 afin que le lecteur puisse mieux constater comment il peut être produit en utilisant l'algorithme qui vient d'être décrit (basé sur le parcours descendant puis ascendant des nœuds de la représentation):

P = { ('être**II**',promenade**1**') , ('promenade**1**',elle') , ('être**II**','jours**III.5**') ,
('jours**III.5**','bonheur**II.1**') , ('bonheur**II.1**',elle') , ('promenade**1**','5') ,
('5','ou**3**') , ('ou**3**','6') , ('6','promenade**1**') , ('promenade**1**','façon**II.1**') ,
('façon**II.1**','tenir**I.1**') , ('tenir**I.1**','elle') , ('tenir**I.1**','bras**1**') ,
('bras**1**','il**I.1**') , ('il**I.1**','idolâtrer') , ('idolâtrer','elle') ,
('idolâtrer','secrètement**1**') , ('jours**III.5**','seul**II.2**') ,
('jours**III.5**','connaître**I.3**') , ('connaître**I.3**','elle') ,
('connaître**I.3**','jamais**1**') }

4 CONTROLE DU PARCOURS SEMANTIQUE

Nous avons mentionné plus haut que l'ordre dans lequel s'effectuait le parcours des actants et des gouverneurs d'un même nœud n'était pas pertinent dans le processus de hiérarchisation d'une RSém. Il est cependant certain que, pour un même réseau, une variation dans l'ordre de parcours peut entraîner un résultat sensiblement différent quant à la hiérarchie produite. Il apparaît que, si certaines de ces variations sont triviales, d'autres peuvent avoir des conséquences importantes sur le déroulement des opérations postérieures de grammaticalisation, lexicalisation et arborisation. Nous proposons donc d'introduire certains tests de contrôle du parcours qui vont garantir la production d'une hiérarchie communicative permettant, au bout du compte, de construire une RSémR (et donc une RSyntP) satisfaisante.

Nous procéderons en trois étapes:

1) Nous expliciterons la parenté existant entre la dépendance communicative et la dépendance syntaxique.
2) Nous montrerons ensuite que certaines hiérarchisations vont à l'encontre d'une lexicalisation et d'une arborisation satisfaisantes. Pour ce faire, nous examinerons deux cas de hiérarchisations ne permettant pas d'effectuer une transition sémantique correcte. Le lien de cause à effet entre la hiérarchisation et les opérations postérieures étant démontré, nous proposerons deux contraintes sur le parcours des RSém qui permettent de contrôler la consistance des hiérarchies communicatives produites.
3) Finalement, nous présenterons un dernier élément de contrôle du parcours sémantique fondé sur ce que nous avons appelé l'*imperméabilité des sous-réseaux* ***thème*** *et* ***rhème***.

4.1 Dépendance communicative et dépendance syntaxique

Soit le graphe représentant la hiérarchie communicative introduite dans la figure V-7:

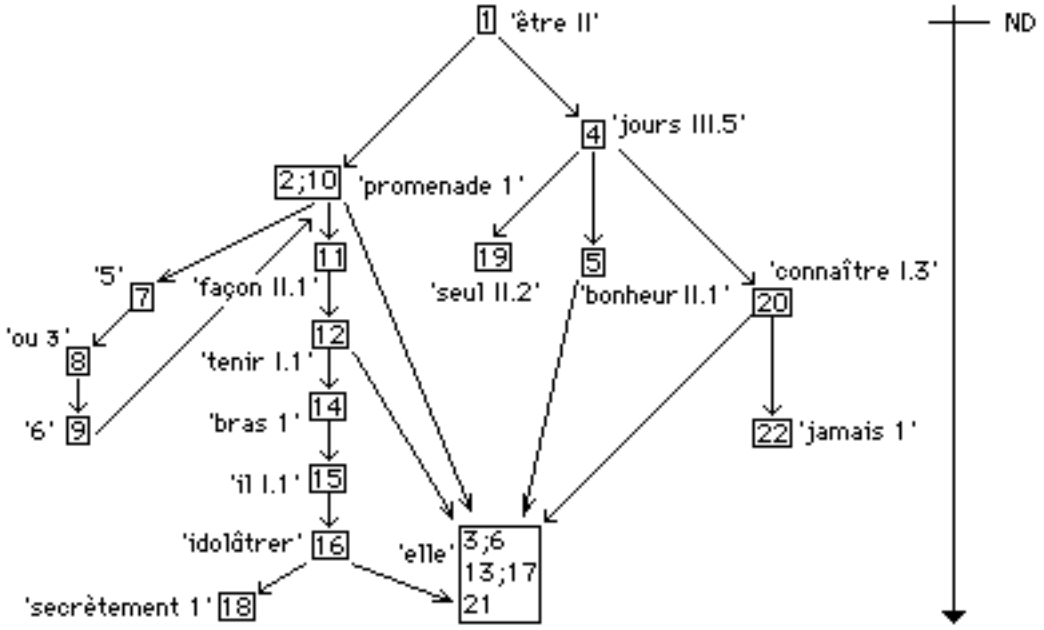


Figure V-8
Hiérarchie communicative de la figure V-7

En examinant la figure V-8, il devient clair que l'intérêt de la hiérarchisation communicative des RSém réside dans le fait que cette opération introduit dans le réseau une structure de dépendance très proche d'une dépendance syntaxique. Sans qu'il soit besoin de donner dès maintenant la RSyntP associée à notre exemple, on voit que sa structure générale va être calquée sur la dépendance communicative de V-8. Bien sûr, le système de dépendances syntaxiques d'une RSyntP correspond à une hiérarchie stricte et se représente par un arbre, ce qui n'est généralement pas le cas du graphe produit grâce au parcours du réseau. Cependant, le dédoublement des nœuds qui dépendent communicativement de plusieurs autres nœuds, avec introduction de liens de co-référence entre ces "siamois", permettrait aisément de produire une figure dont la structure serait très nettement une sorte de squelette de la structure d'un arbre

syntaxique profond. On peut même aller plus loin en dessinant en gras, dans une telle figure, les liens de dépendance communicative qui fonctionnent à l'inverse de la dépendance sémantique présente dans le réseau. Il apparaît alors instantanément que ces liens de dépendance spéciaux (qui sont ceux établis lors d'un parcours ascendant) sont en correspondance avec des dépendances syntaxiques profondes autres que les dépendances actancielles. Nous avions en effet déjà constaté, dans la partie précédente en II:4.2, que de telles configurations du niveau sémantique étaient rendues, en syntaxe profonde, par l'établissement de liens **ATTR**, **APPEND** et, dans certains cas, **COORD**. La figure V-9 présente la hiérarchie de la figure V-8 sous une forme modifiée pour faire apparaître le type de structure arborescente qui vient d'être décrit:

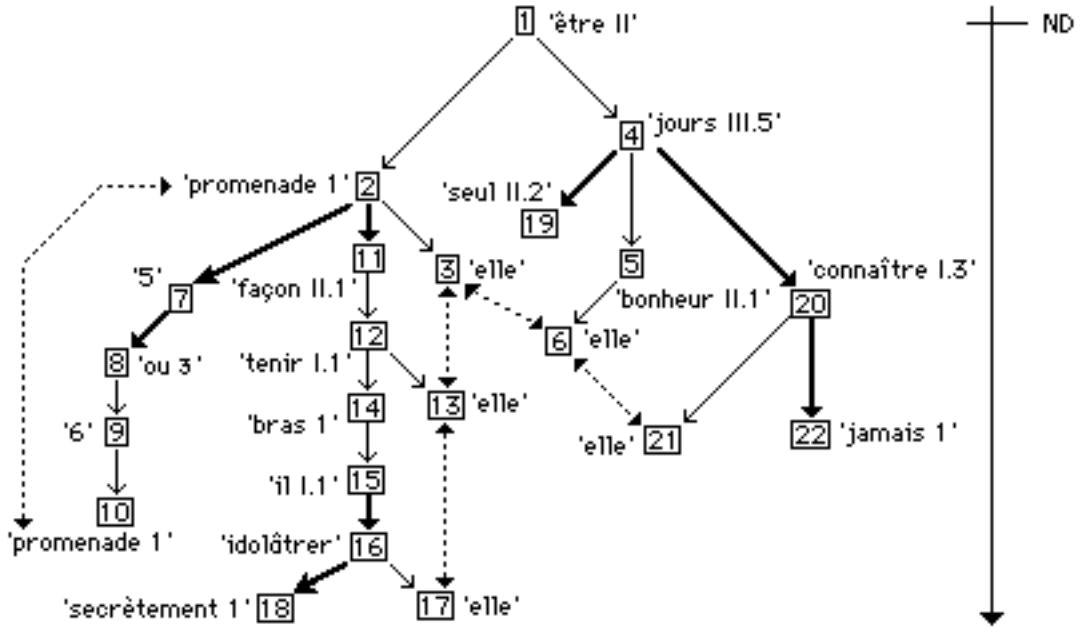
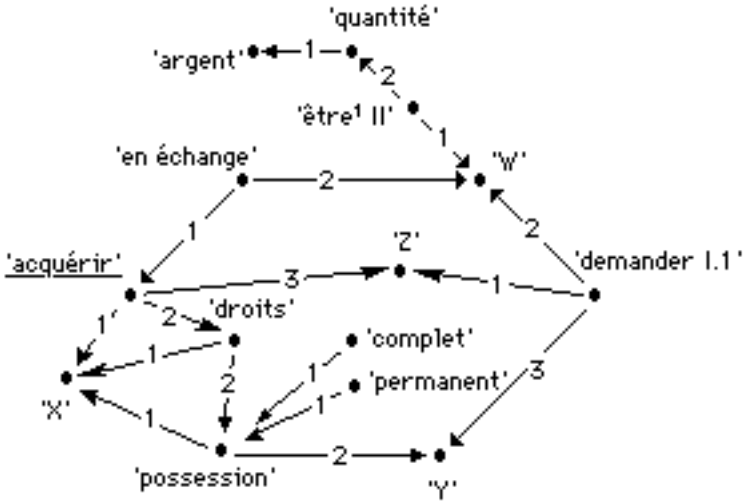


Figure V-9
Hiérarchie communicative modifiée
pour créer une structure arborescente

Ce type de figure, sans être bien sûr une représentation d'une véritable dépendance syntaxique profonde, évoque très directement le niveau cible de la transition sémantique. Nous pouvons donc conclure que:

La hiérarchie communicative contient en filigrane un système de dépendances syntaxiques profondes.

Ce qui vient d'être dit à propos des RSém s'applique aussi aux définitions de lexèmes. Dans la première partie de ce travail, nous avons montré que les règles de paraphrase sémantique doivent être appliquées sur des réseaux communicativement hiérarchisés. Dans la pratique, ceci signifie qu'à chaque partie droite d'une règle de paraphrase sémantique doit être associée une hiérarchisation communicative et que c'est la définition hiérarchisée qui doit matcher sur une partie de la RSém. Par exemple, la décomposition sémantique de ACHETER**1** (examinée en II:2.1):



pourrait se voir associer la hiérarchie communicative de la figure V-10:

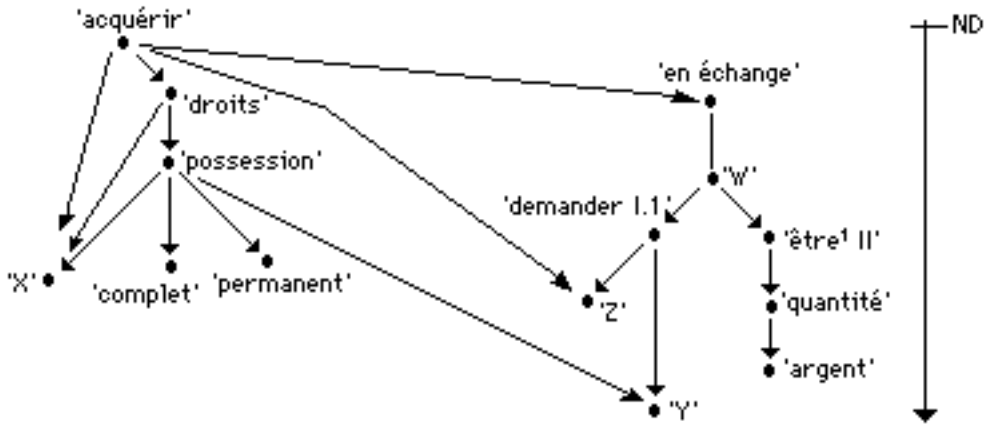


Figure V-10
Hiérarchie communicative associée à la définition de ACHETER**1**

Cette hiérarchie communicative peut être modifiée, sur le modèle de la figure V-9, pour créer une structure arborescente très proche de la structure syntaxique profonde de (2) ci-dessous; rappelons que (2) est une phrase correspondant à la définition "francisée" proposée dans le DEC.

(2) "X acquiert de Z les droits de possession complète et permanente de Y en échange d'une quantité d'argent W que Z demande pour Y"

<u>Bilan Provisoire</u>

Arrivé à ce stade de notre exposé, la logique de notre stratégie d'organisation procédurale de la transition sémantique commence à apparaître clairement. Récapitulons-en les aspects essentiels.

Nous avons tout d'abord mis à jour la nécessité de déterminer un NOUVEAU TYPE DE DEPENDANCE dans une RSém permettant de faire jouer de façon adéquate les règles de correspondance sémantique. Cette dépendance résulte de la structure communicative de la RSém et, par conséquent, a été appelée *dépendance communicative*.

Nous avons ensuite montré que cette dépendance était une notion implicitement contenue dans celle de *nœud dominant* et que la

présence d'un ND dans un réseau conditionnait une certaine hiérarchie communicative reflétée par le sous-ensemble particulier des lexicalisations et arborisations pouvant être opérées sur ce réseau. Nous avons donc décidé de faire passer la transition RSém ⇒ RSyntP par une étape de hiérarchisation communicative de la RSém débutant par la sélection d'un ND.

Nous avons montré finalement qu'il était possible d'établir un algorithme de parcours des RSém permettant de calculer, pour chaque RSém possédant un ND, une hiérarchie communicative qui est un bon canevas pour mener à bien, notamment, la lexicalisation et l'arborisation.

Les conséquences sur la lexicalisation d'une mauvaise hiérarchisation ont pu être démontrées aisément lorsque, dans la Partie **1** (II:2.1), nous avons étudié les liens existant entre la notion de dépendance communicative et la mise en jeu des règles de paraphrase sémantique. Pour l'arborisation, nous nous sommes contenté (en II:4) de montrer que la dépendance communicative devait être considérée dans l'écriture des règles de correspondance sémantique proprement dites. Cependant, le lien entre la hiérarchie communicative d'un réseau sémantique et la structure hiérarchique d'une RSyntP correspondante n'avait pas été envisagé au niveau global de la phrase.

Nous avons jusqu'à présent traité un exemple particulier de façon à produire une hiérarchisation qui soit valide. À fin de comparaison, il nous reste à montrer quelles pourraient être les conséquences d'une mauvaise hiérarchisation de la RSém: C'EST-A-DIRE QU'IL NOUS FAUT MONTRER QU'IL Y A EFFECTIVEMENT DE BONNES ET DE MAUVAISES HIERARCHISATIONS.

4.2 Conséquences d'une mauvaise hiérarchisation

Nous allons tout d'abord, dans cette sous-section, introduire un exemple particulier qui servira de base à notre discussion. La présentation de cet exemple nous conduira à effectuer quelques rappels sur la notion de *présupposition* qui sera située dans le cadre

spécifique de la TST. Nous pourrons ensuite utiliser notre exemple pour montrer les différentes incidences d'une variation dans la hiérarchisation (i) sur la lexicalisation, et (ii) sur l'arborisation.

4.2.1 Présentation d'un exemple

Nous allons reprendre ici un exemple en anglais utilisé dans Mel'čuk & Polguère (à paraître). Il s'agit de la phrase (3) ci-dessous dont la figure V-11 est la RSém.

(3) *The police persecute pilots who possess a Polish plane.*

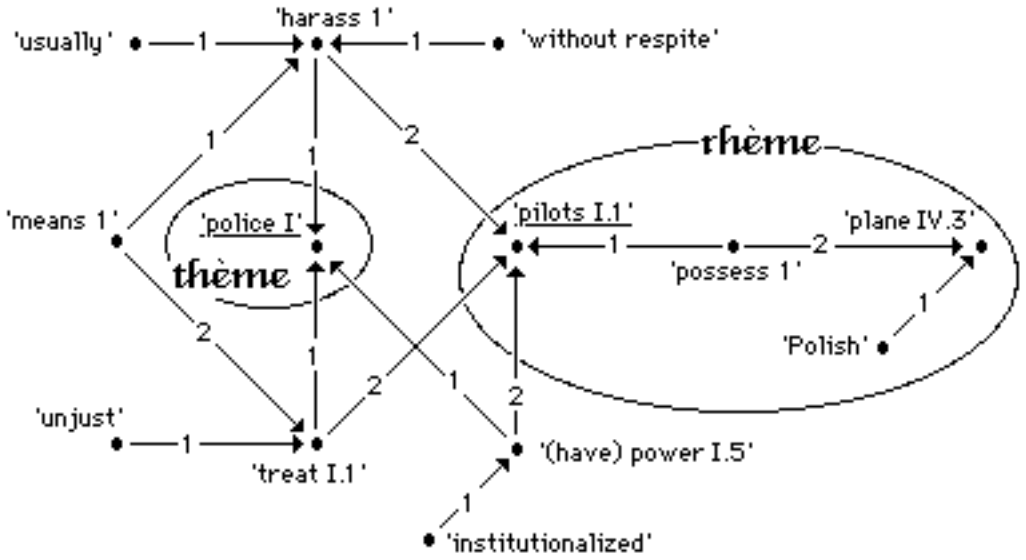


Figure V-11

RSém de *"The police persecute pilots who possess a Polish plane."* [3]

La construction de la phrase (3), à partir de ce réseau, doit passer par la sélection de 'harass**1**' comme ND. Un tel choix permet de lexicaliser une partie du réseau en 'persecute**1**' grâce à la définition du lexème PERSECUTE**1**, donnée dans la figure V-12 ci-dessous. Cette définition

---

[3] Les numéros distinctifs étiquetant les Sens de lexèmes sont ici tirés de *Longman Dictionary of Contemporary English*. Nous gardons l'exemple dans sa version originale anglaise, mais il va de soi que ce qui suit s'appliquerait aussi bien à des configurations sémantiques de phrases françaises.

introduit un type d'information dont nous n'avions pas parlé dans notre présentation des règles de paraphrase sémantique; il s'agit de la composante présuppositionnelle d'un réseau sémantique. Cette composante est indiquée ici sous forme de deux sous-réseaux (**présup. 1** et **présup. 2**) et nous allons revenir en détail sur son rôle dans la manipulation des règles de paraphrase sémantique aussi bien que des RSém en général.

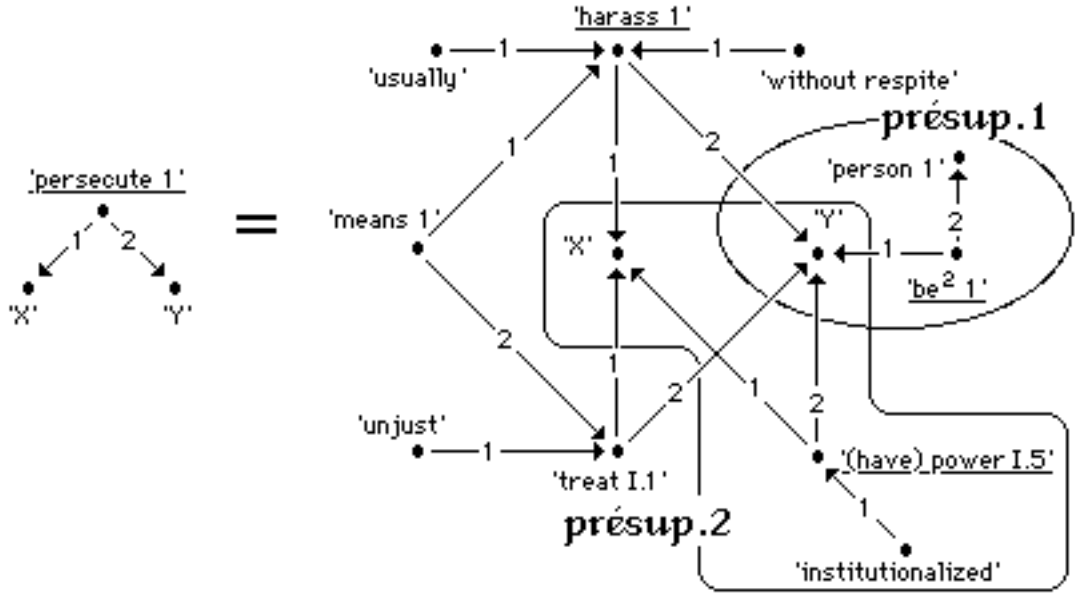


Figure V-12
Définition du lexème anglais PERSECUTE**1**

Les trois questions suivantes peuvent être formulées à propos de la notion de présupposition:

1) Qu'est-ce qui confère à certaines composantes d'une telle définition le statut de présupposition?
2) Pourquoi est-il nécessaire de faire figurer les présuppositions dans les définitions?
3) Les présuppositions doivent-elles aussi être marquées comme telles dans une RSém?

Nous allons tenter de répondre à toutes ces questions après avoir (i) fait quelques brefs rappels sur la notion de présupposition, et (ii) donné une définition provisoire de la notion de présupposition dans le cadre de la théorie et des modèles Sens-Texte.

4.2.2 La notion de présupposition

La présupposition peut être décrite comme une signification qui, bien qu'étant implicite, fait tout de même partie de la signification littérale d'un énoncé. Considérons l'exemple suivant:

(4) *Dagobert soupçonne M. Rodin d'appartenir à la Compagnie.*
présup. 1 'Dagobert pense qu'appartenir à la Compagnie est quelque chose de mal'
...

On peut tout à fait imaginer que, dans un contexte donné, l'énonciation de (4) contienne implicitement un message du type 'Dagobert veut que tu ailles trancher le cou de M. Rodin'. Cependant, ce que nous avons relevé comme étant une présupposition est une signification implicite d'un type très différent. Elle se perçoit sans qu'il faille imaginer un contexte d'énonciation donné. Elle est tout à fait linguistique et fait partie du Sens littéral de (4). Communicativement, il faut noter, comme le signale Ducrot 1972, que la perception d'une signification implicite comme la présupposition nécessite la perception de la signification explicite. Du point de vue du locuteur, qui seul nous intéresse véritablement ici, les significations implicites de l'énoncé produit correspondent à des Sens qui ne sont pas NECESSAIREMENT volontairement communiqués mais qui découlent indirectement de l'énoncé. Il existe bien sûr des emplois "pathologiques" de la langue (peut-être même les plus fréquents) où le locuteur réalise en fait une énonciation afin de communiquer avant tout le Sens qui est implicite. Mais, dans un tel cas, il fait comme s'il ne cherchait pas à communiquer ce Sens. Donc, au strict plan de la langue, nous dirons que LES PRESUPPOSITIONS, EN TANT QUE SIGNIFICATIONS IMPLICITES, SONT DES SIGNIFICATIONS QUE L'ON

N'EXPRIME PAS COMME TELLES MAIS QUI SONT AUTOMATIQUEMENT COMMUNIQUEES PAR L'ENONCE LUI-MEME. O. Ducrot utilise les termes de "signification exprimée" *vs.* "signification attestée" pour rendre compte de cette différence.

Il est très important de remarquer que "implicite" ne signifie pas ici "qui n'est pas réalisé linguistiquement". Une présupposition, bien que non explicitement communiquée, est formulée d'une façon ou d'une autre. On peut ainsi distinguer deux cas de figure:

1) Le présupposé trouve son origine dans l'emploi d'un lexème donné qui le véhicule intrinsèquement; c'est le cas illustré par l'exemple (4) ci-dessus.
2) Le présupposé peut être aussi formulé en surface mais avec un statut particulier dans la phrase. Dans ce dernier cas, on pourrait dire que LE PRESUPPOSE DE LA PHRASE SE TROUVE LA "INCOGNITO" ET C'EST EN CE SENS QU'ON LE CONSIDERE COMME UNE SIGNIFICATION IMPLICITE. Ainsi, une phrase comme *"Il parle à sa sœur."* présuppose 'Il a une sœur'. Ce Sens est directement exprimé par les mots *sa sœur*, mais il est exprimé de telle façon que LA PHRASE NE SEMBLE PAS LE COMMUNIQUER: il est formulé explicitement mais communiqué implicitement.

Il conviendra donc de distinguer le fonctionnement de la *présupposition de lexème* et celui de la *présupposition de phrase*, qui n'est pas nécessairement une présupposition de lexème.

On a souvent proposé d'utiliser la négation comme critère de mise en évidence de la présupposition. Il semblerait en effet qu'une des caractéristiques des présupposés d'un énoncé *E* est qu'ils restent valides comme présupposés d'un énoncé ~*E* (qui serait la négation de *E*), comme l'illustre (5):

(5) *Dagobert ne soupçonne pas M. Rodin d'appartenir à la Compagnie.*

présup. 1 'Dagobert pense qu'appartenir à la Compagnie est quelque chose de mal'

...

Bien que séduisant, ce critère est cependant beaucoup moins rigoureux et infaillible qu'on pourrait le croire à première vue. En effet, que veut dire "la négation de *E*"? Est-ce une négation linguistique? Mais, à ce moment, quel type de négation linguistique et sur quel élément de la phrase doit-elle porter? Est-ce une négation logique? Mais alors, à nouveau, quelle est, sur le plan logique, la portée de cette négation? Ce que l'on veut, c'est disposer d'un critère qui pourrait s'appliquer à tout type d'énoncé — aussi complexe ou aussi simple soit-il, véhiculant sémantiquement des présupposés. De plus, ce critère de la négation, même s'il peut être précisé et formalisé un peu mieux, n'est sûrement pas suffisant pour permettre d'identifier tous les présupposés d'un énoncé et rien que les présupposés. Il faut alors faire appel à des outils additionnels comme le questionnement qui, lui aussi, semble préserver la valeur des présupposés:

(6) *Est--ce que Dagobert soupçonne M. Rodin d'appartenir à la Compagnie?*

présup. 1 'Dagobert pense qu'appartenir à la Compagnie est quelque chose de mal'

...

Nous ne voulons pas nous enliser dans une étude de la présupposition qui, comme celle des notions de thème et rhème, dépasse largement le cadre de ce travail. Pour certains, les présupposés d'un énoncé *E* sont considérés d'un point de vue strictement logique comme les conditions de satisfaction de l'énonciation de *E* (cf. les débats philosophiques sur cette question illustrés par Strawson 1950, puis Strawson 1964); pour d'autres, ils font partie intégrante du sens de *E* et sont approchés à travers la notion d'acte illocutoire de présupposition (cf. Ducrot 1972 et les

travaux ultérieurs). Parmi toutes les définitions relativement linguistiques de la présupposition, nous mentionnerons celle de Mørdrup 1975:128 qui est, à notre avis, assez élégante:

> "P présuppose Q si et seulement si à chaque fois que P est affirmée, niée ou mise sous forme de question le locuteur ne peut pas nier que Q sans se contredire."

Nous contournerons le débat philosophico-linguistique entourant cette notion en reprenant à notre compte la définition ci-dessus. Nous voudrions cependant tenter de donner, pour la notion plus particulière de *présupposé de lexème* UNIQUEMENT, une caractérisation toute formelle, qui sera partielle et provisoire mais néanmoins suffisante dans le cadre de la présente discussion.

4.2.3 Ébauche de définition topographique de la *présupposition de lexème*

Nous allons tenter de donner une définition de la présupposition de lexème qui soit formulée strictement à l'aide de critères portant sur la *topographie des réseaux sémantiques*. Mentionnons cependant que ce qui est proposé ici relève plus d'une hypothèse forte, qui demanderait à être étudiée plus en profondeur, que d'une affirmation ou même d'une thèse que nous défendrions la tête sur le billot.

Notons tout d'abord que, contrairement à ce que certains ont proposé (cf. Boguslavskij 1985), il ne nous paraît pas du tout approprié d'effectuer un rapprochement entre la notion de présupposé et celle de thème. Comme on peut le constater à travers les travaux mentionnés plus haut sur les enchaînements thématiques dans le texte (cf. Iordanskaja 1989 et Daneš 1974), un des rôles essentiels du thème est de rendre compte de la cohérence textuelle: un énoncé prend généralement comme thème un élément ayant été explicité dans un énoncé précédent — notamment en tant que rhème de ce dernier. Thème et rhème, pour fonctionner comme tels, s'appuient

donc sur des composantes du Sens d'un énoncé qui sont essentiellement explicites. Or, justement, le présupposé est un élément de Sens qui est communiqué de façon implicite. Il se laisse difficilement mettre en valeur puisqu'il correspond à une sorte de communication camouflée. Ducrot 1972 utilise même cette caractéristique comme critère de spécification du présupposé qui s'ajoute aux critères de négation et d'interrogation. Il note ainsi que si l'on enchaîne deux énoncés E_1 et E_2 au moyen de conjonctions comme *donc*, *parce que*, etc. le second énoncé portera sur le posé, et non sur les présupposés, du premier énoncé; c'est ce que Ducrot appelle la "loi d'enchaînement"[4]. On peut aussi trouver dans Padučeva 1985:68 une remarque intéressante liée au fonctionnement de la présupposition dans les enchaînements d'énoncés. Si un présupposé p_1 d'un énoncé E_1, inclus dans un énoncé E, est rendu explicite dans le reste de E, il perd son caractère de présupposé. L'exemple (7) illustre ce phénomène:

(7) E_1 = *Dagobert accuse M. Rodin de cette action étrange.*
p_1 = 'Dagobert pense que cette action étrange est mauvaise'
E = *Dagobert accuse M. Rodin de cette action étrange dont il pense qu'elle est mauvaise depuis les événements de février dernier.*

On voit ici que, si p_1 est présupposé par E_1 lorsque cet énoncé est pris de façon isolée, il est difficile de le considérer comme étant PRESUPPOSE par E. Cela découle du fait que la subordonnée *"dont il pense..."* semble explicitement communiquer ce Sens qui est une présupposition du lexème ACCUSER**1**.

Nous considérerons donc que, si la spécification des présupposés fait partie de la caractérisation de la structure communicative (au sens large) d'une RSém, elle ne doit pas être confondue avec l'opposition thème/rhème dont nous avons largement

---

4 Il s'agit bien sûr ici d'enchaînement "naturel" des énoncés. La loi d'enchaînement et d'éventuels contre-exemples sont discutés en détail dans Ducrot 1972:81-89.

discuté au Chapitre IV. La principale caractéristique de l'opposition thème/rhème est, justement, de correspondre à un système d'opposition binaire qui isole deux grappes de Sens "concurrentes" devant être mises en valeur dans l'énoncé. Le présupposé, lui, ne joue pas le même rôle dans la structuration de l'énoncé puisqu'il contrôle la transition sémantique en ce sens qu'il doit "se faire le plus discret possible" aux autres niveaux de représentation[5]. Ces remarques s'appliquent tout particulièrement aux présupposés spécifiés dans la RSém afin de contrôler la lexicalisation, c'est-à-dire les présupposés qui devront être exprimés (implicitement) par un lexème. En nous appuyant sur l'appareillage formel dont nous disposons maintenant, nous pouvons même proposer cette première hypothèse:

> Les RSém peuvent comprendre une composante présuppositionnelle particulière présentant les futures *présuppositions de lexème*. Cette composante devra cependant disparaître lors de la lexicalisation et sera absente des RSémR.

Ces composantes présuppositionnelles particulières des RSém doivent donc être occultées par la lexicalisation ou même la grammaticalisation. Les définitions de lexèmes, par contre, doivent très clairement faire apparaître leur composante présuppositionnelle pour permettre, justement, une lexicalisation adéquate. Malgré le parallèle formel que nous avions établi précédemment entre les RSémR et les réseaux sémantiques des définitions, on voit donc que ces derniers se distinguent radicalement des RSémR au niveau de la structuration communicative[6].

---

5 On trouvera dans Ducrot 1972:59-62 une réfutation de l'assimilation présupposé-thème. Nous ne reprenons pas ici les arguments proposés car, s'ils se fondent sur une description assez poussée de la notion de présupposition, la notion de structure communicative y est, elle, présentée de façon trop vague.

6 Structuration dont fait partie, rappelons-le, la spécification de la composante présuppositionnelle.

Cette première particularisation "Sens-Texte" de la présupposition de lexème étant faite, nous allons tenter d'isoler ce qui, DANS LA DEFINITION D'UN LEXEME, peut être considéré comme une présupposition. Nous nous référerons ici à Iordanskaja 1986 qui, dans le cadre d'une étude sémantique des verbes promoteurs de la négation en français, utilise des définitions de type DEC indiquant explicitement une COMPOSANTE PRESUPPOSITIONNELLE et une COMPOSANTE ASSERTIVE. Le format de ces définitions est le suivant[7] :

Sens du lexème = Composante présuppositionnelle |
Composante assertive

Bien que L. Iordanskaja ne réfère qu'au seul critère de la négation pour spécifier la présupposition, nous croyons pouvoir noter que la présupposition de lexème porte sémantiquement avant tout sur les actants du lexème défini et non sur d'autres participants sémantiques de la composante assertive. Ainsi, en reprenant notre exemple (4), on remarque que la composante présuppositionnelle 'X pense que Z est quelque chose de mal', qui ferait partie de la définition de 'X soupçonne Y de Z', établit un lien sémantique entre 'X' et 'Z' qui ne met aucunement en jeu les Sens que l'on pourrait trouver dans la composante assertive de la définition de 'soupçonner' ('penser', 'avoir fait', etc.). Nous faisons donc ici l'hypothèse que ceci n'est pas un cas isolé et que les présuppositions de lexèmes servent toujours (i) à introduire une spécification sémantique des actants du lexème, ou (ii) à établir des liens sémantiques entre ces actants.

Notre hypothèse est vraisemblablement un peu trop forte et il convient de tenter de la vérifier au moins sur quelques exemples. Nous utiliserons ici quelques définitions proposées dans Iordanskaja 1986:349.

---

7 Dans les définitions du DEC français (cf. Mel'čuk *et al.* 1984 et 1988), la présupposition de lexème est indiquée par des tournures syntaxiques comme 'étant...', 'croyant...', 'qui croit que...'. On peut se reporter à II:2.1 pour quelques commentaires sur la transcription "francisée" des réseaux sémantiques des définitions.

Soit les deux verbes, PRIER et ORDONNER, qui ne semblent se distinguer (sémantiquement) que par la composante présuppositionnelle de leur définition (indiquée ici en gras)[8] :

(8) *X prie Z de Y-er* = '**X croit que si Z fait l'action Y cela serait une faveur de la part de Z envers X; |** X communique à Z, avec le but d'inciter Z à faire Y, que X veut que Z fasse Y'

(9) *X ordonne à Z de Y-er* = '**X a un pouvoir sur Z; |** X communique à Z, avec le but d'inciter Z à faire Y, que X veut que Z fasse Y'

La composante assertive de ces définitions peut être représentée graphiquement par le sous-réseau suivant:

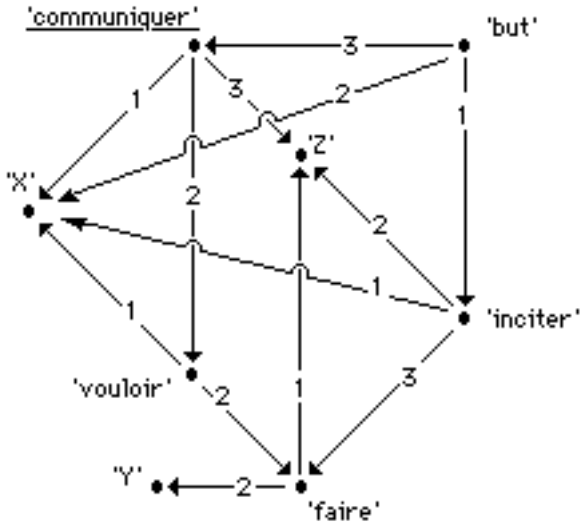


Figure V-13
Représentation graphique de la composante assertive
des définitions de PRIER et ORDONNER

et leur composante présuppositionnelle par:

---

8 Ces définitions, telles que formulées, ne tiennent pas compte d'une différence vraisemblablement majeure entre le Sens de PRIER et celui de ORDONNER; il s'agit de la façon dont quelque chose est communiqué (avec un certain ton, en manifestant une certaine autorité, etc.). Nous laisserons de côté ce problème qui n'interfère pas avec ce que nous voulons dire sur la présupposition dans ces définitions.

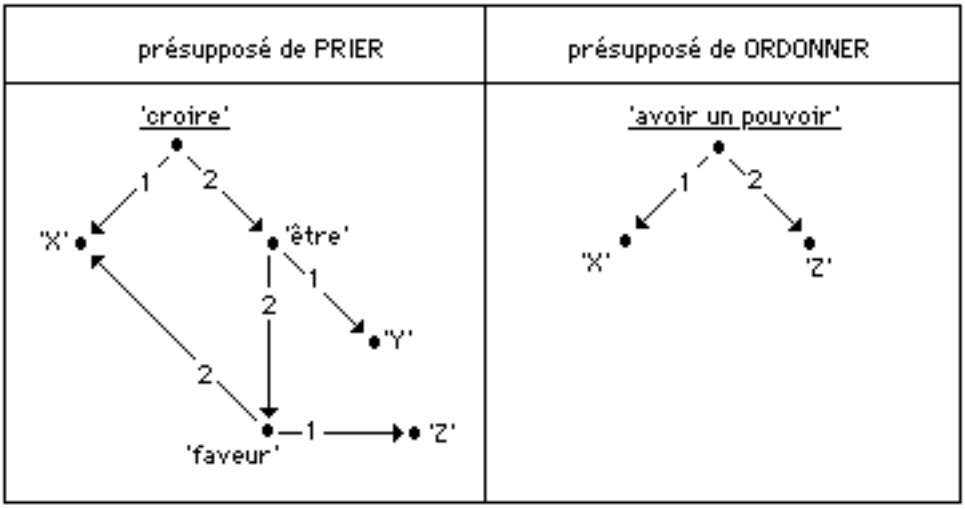


Figure V-14
Représentation graphique de la composante présuppositionnelle des définitions de PRIER et ORDONNER

Remarque: Nous avons allégé la présupposition de PRIER en ne tenant pas compte de la composante 'si Z fait Y' qui découle en fait "logico-sémantiquement" du reste du réseau (si X a pour but d'inciter Z à faire Y, c'est qu'au moment de la communication X pense que Y n'a pas encore eu lieu; Y sera donc une faveur de Z envers X si Z accomplit effectivement Y).

Les sous-réseaux présuppositionnels de la figure V-14 se connectent à la composante assertive (figure V-13) exclusivement par l'intermédiaire des actants sémantiques du verbe défini: X, Y et Z dans le cas de PRIER et X et Z dans le cas de ORDONNER. Cela est donc en accord avec notre hypothèse qui veut que la composante présuppositionnelle ne porte sur aucun élément asserté de la définition, si ce n'est les actants sémantiques eux-mêmes.

On trouve cependant dans Iordanskaja 1986 certaines définitions qui, de par leur formulation, semblent contredire notre hypothèse; cf., par exemple (page 349 et page 362):

(10) *X demande pardon à Z pour Y* = '**X croit avoir causé un mal à Z par son action Y;** | X communique à Z, avec le but de compenser ce mal, que X regrette que X ait fait Y et que X veut que Z n'ait plus de sentiments désagréables à cause de Y'

(11) *X permet à Z de Y-er* = '**X sait ou croit que Z veut Y-er et que Z ne veut pas Y-er contre la volonté de X;** | X communique à Z que l'action Y de Z ne serait pas contraire à la volonté de X'

Dans (10), la définition de ⌜DEMANDER PARDON⌝ établit très clairement un lien sémantique entre la composante présuppositionnelle et la composante assertive par l'intermédiaire de 'mal': 'X croit avoir causé un mal...' et 'avec le but de compenser ce mal'. Nous pensons cependant que cela est dû à un problème de formulation. Notons tout d'abord que, selon cette définition, 'X demande pardon à Z pour Y' implique que 'X a fait Y'. Il serait peut-être plus juste de considérer que la proposition présente ici est plutôt du type 'X a une responsabilité dans Y' dans la mesure où l'on peut toujours demander pardon à quelqu'un pour une action que l'on n'a pas commise mais dans l'exécution de laquelle on se sent impliqué à divers titres:

(12) *Mes chats ne sont vraiment pas très bien éduqués; je vous demande pardon pour ce qu'ils ont fait à votre poisson rouge.*

Cependant, même en oubliant ce problème et en considérant que X est effectivement l'auteur de Y, nous pensons que dans (10) 'X croit avoir causé un mal à Z par son action Y' n'est pas un présupposé DIRECT de ⌜DEMANDER PARDON⌝ mais est en réalité un présupposé HERITE DE LA COMPOSANTE ASSERTIVE DU SENS DE CE PHRASEME. 'X croit avoir causé un mal...' est en effet un présupposé de 'dans le but de compenser Y' (que nous préférons à 'dans le but de compenser ce mal'); si X a le but de compenser Y cela présuppose que Y est ou a causé, selon X, quelque chose de mal. De même, si X regrette d'avoir fait Y, cela présuppose aussi que X pense que Y est ou a causé quelque chose de mal. Le présupposé de (10) est donc valide en tant que présupposé HERITE d'un Sens de la composante assertive. Un des

présupposés DIRECTS à spécifier dans la définition de ⌜DEMANDER PARDON⌝ serait peut-être dans ce cas 'X pense que Z est informé de Y'.

Notons que ce problème de la distinction entre présupposé direct de la définition du lexème et présupposé hérité de la composante assertive de sa définition peut être soulevé à partir d'autres études sémantiques que celles fondées sur la TST. On trouve ainsi dans Fillmore 1971:282 la description sémantique suivante du lexème anglais ACCUSE[9] :

ACCUSE [Judge , Defendant , Situation (of)]
Meaning: SAY [Judge , ‘ X ’ , Addressee]
X = RESPONSIBLE [Situation , Defendant]
Presupposition: BAD [Situation]

Selon nous, le présupposé est très clairement ici un présupposé de RESPONSIBLE, ou plutôt d'un lexème RESPONSIBLE**4** (cf. la numérotation du dictionnaire Longman 1978); c'est celui que l'on retrouve dans une phrase comme (13a) — noter, dans (13b), qu'il est difficile d'employer un tel lexème avec un complément d'objet ayant des connotations positives[10] :

(13) a. *You are responsible for that car crash.*
b. ? *You are responsible for that wonderful ceremony.*

Le présupposé relevé par Fillmore pourrait donc en réalité être décrit seulement en tant que présupposé de ce lexème RESPONSIBLE**4** (présupposé hérité par ACCUSE).

---

9 Une telle description est fondée sur ce que Fillmore appelle la "structure de rôle" dont la présentation ne nous concerne pas ici. Pour plus de détails, on pourra se référer directement à Fillmore 1971.

10 (13b) est acceptable sans problème si l'on considère un autre Sens de *responsible*, du type 'être en charge de'. Ce Sens correspond au lexème RESPONSIBLE**1** (toujours selon la numérotation du Longman).

Considérons maintenant le cas de la définition (11). Un problème surgit à cause du recours au bloc de Sens 'la volonté de X' dans les composantes présuppositionnelle et assertive. Une telle formulation semble impliquer que l'on parle d'un seul et même nœud sémantique alors qu'il est bien évident que l'on réfère à "deux volontés" distinctes et même contradictoires de X. On peut donc reformuler la définition de la façon suivante pour mettre en évidence le fait que seuls les actants X, Y et Z se retrouvent à la fois dans les composantes présuppositionnelle et assertive:

(11') *X permet à Z de Y-er* = '**X sait ou croit que Z veut Y-er et que Z ne veut pas Y-er si X s'opposait à Y;** | X communique à Z que X ne s'oppose pas à l'action Y de Z'

Dans une telle formulation, on retrouve le Sens 's'opposer' dans le présupposé et l'assertion mais il est clair ici que l'on est en présence de deux occurrences distinctes du même Sens — deux nœuds sémantiques distincts. Il était beaucoup plus difficile de percevoir une telle distinction dans la formulation (11). La représentation graphique permet, dans de tels cas, de faire clairement apparaître les différentes occurrences d'un même Sens.

Nous retiendrons donc de l'examen de ces deux contre-exemples apparents la consigne suivante:

Pour juger de la relative séparation formelle des composantes présuppositionnelle et assertive d'une définition sémantique, il faut tenir compte des deux facteurs suivants: (i) la composante présuppositionnelle est parfois elle-même héritée de (présupposée par) la composante assertive, (ii) deux mentions d'un même Sens, l'une présupposée, l'autre posée, doivent correspondre à deux occurrences distinctes de ce Sens.

Si l'examen de ces quelques cas de définitions ne prouve pas que notre hypothèse soit parfaitement juste, il semble en tout cas n'apporter que des confirmations de notre proposition initiale. Nous pouvons donc maintenant reformuler de façon formelle la définition provisoire que nous proposons pour la composante présuppositionnelle de la définition d'un lexème:

Définition provisoire de la notion de *présupposition d'un lexème*:

La présupposition d'un lexème *L* est la composante de la définition de ce lexème qui reste inchangée dans la RSém de *non L* et qui ne se rapporte à aucun Sens asserté par *L*.

Une telle définition peut sembler circulaire: ce qui est présupposé, c'est ce qui n'est pas asserté. Cependant, dans la pratique, il est souvent plus facile d'isoler des éléments de la définition qui sont évidemment assertés; comme 'communiquer' dans la définition de PRIER — cf. (8) ci-dessus. À partir de là, nous avons un critère formel relativement fiable pour dire que, dans une telle définition, une composante sémantique comme 'avec le but d'inciter Z à faire Y' ne fait pas partie de la présupposition de PRIER car elle se rattache sémantiquement à 'communiquer': 'but' possède 'communiquer' comme argument. Notons que justement, dans ce cas, le seul critère de la négation est peu opérant pour donner à la composante 'avec le but d'inciter Z à faire Y' un statut d'assertion et non de présupposition.

Nous pouvons maintenant refermer cette longue parenthèse, visant à introduire la notion de présupposition dans le cadre de la TST, et revenir à notre exemple de la figure V-11, phrase (3), afin d'examiner certains cas de mauvaise hiérarchisation communicative. On pourra noter que les sous-réseaux présuppositionnels **présup.1** et **présup.2**, qui apparaissent dans la définition de PERSECUTE**1**

(figure V-12), correspondent bien à la notion de présupposition qui vient d'être définie. De plus, comme tous les sous-réseaux que nous manipulons, ils possèdent un nœud dominant (désigné au moyen du soulignement).

4.2.4 Deux cas d'échec dans la hiérarchisation

Dans la figure V-11, le choix de 'harass**1**' comme ND permet, à travers la hiérarchie communicative qu'il conditionne, d'effectuer une arborisation adéquate.

'harass**1**' n'est pourtant pas le seul ND apparemment valide. Nos règles de détermination du ND nous permettent de choisir, par exemple, 'treat**I.1**', qui se trouve lui aussi sur un chemin communicatif. Un tel choix donnerait lieu à la production d'une phrase comme (14) qui est une paraphrase raisonnable de (3):

(14) *The police unjustly treat pilots who possess a Polish plane and over who they have institutionalized power, this treatment being carried out by means of usual incessant harassment.*

Cette phrase n'est bien sûr pas très élégante mais elle semble "parler tout à fait de la même chose" que (3).

Supposons maintenant que nous sélectionnions comme ND le nœud '(have) power**I.5**', qui se trouve semble-t-il lui aussi sur un chemin communicatif. Un tel ND va permettre d'effectuer le parcours sémantique P_1* qui conditionne la hiérarchisation communicative de la figure V-15, dont nous montrerons qu'elle n'est pas valide. Nous utilisons ici le formalisme arborescent de représentation de la hiérarchie communicative (introduit précédemment dans la figure V-9) afin de faciliter, pour le lecteur, l'association avec une hiérarchie syntaxique profonde donnée. Il convient de noter que, dans une telle représentation, le sous-arbre de dépendance communicative ayant pour sommet 'possess**1**' pourrait être rattaché à n'importe lequel des nœuds co-référents 'pilots**I.1**'.

P_1* = {('(have) power**I.5**','police**I**') , ('(have) power**I.5**','pilots**I.1**') , ('(have) power**I.5**','institutionalized') , ('police**1**','harass**1**') , ('harass**1**','pilots**I.1**') , ('harass**1**','usually') , ('harass**1**','without respite') , ('harass**1**','means**1**') , ('means**1**','treat**I.1**') , ('treat**I.1**','police**I**') , ('treat**I.1**','pilots**I.1**') , ('treat**I.1**','unjust') , ('pilots**I.1**','possess**1**') , ('possess**1**','plane**IV.3**') , ('plane**IV.3**','Polish')}

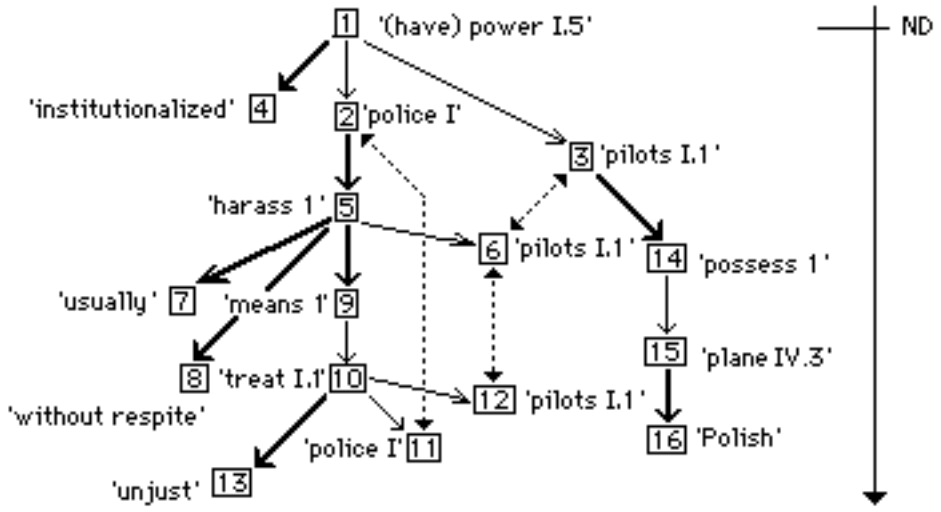


Figure V-15
Hiérarchisation communicative invalide de la RSém V-11
établie à partir du ND '(have) power**I.5**'

Une telle hiérarchie communicative correspond à un énoncé comme (15) qui, contrairement à (14), ne peut absolument pas être considéré comme une paraphrase, même inélégante, de (3):

(15) *The police, who usually harass without respite, and by means of unjust treatment, pilots who possess a Polish plane, have institutionalized power over those pilots.*

La hiérarchie communicative établie grâce au parcours sémantique P_1* est donc invalide car elle conduit à des Textes qui n'expriment pas véritablement le contenu sémantique de la RSém de départ. Cette anomalie tient au fait que nous avons sélectionné ici comme ND une composante de la RSém qui n'a pas d'autre raison d'être que de permettre d'effectuer une lexicalisation grâce à la définition de

PERSECUTE**1**. Une telle composante a été en fait introduite pour permettre de "satisfaire" les présupposés de ce lexème. Cette information est donc incluse dans la RSém pour être exprimée implicitement; cependant, sa promotion au statut de ND en fait nécessairement, au niveau du Texte, une information explicite.

Nous devons tirer de cela trois conclusions:

1) Tout d'abord, c'est par un artifice tout à fait théorique que nous nous permettons ici d'introduire dans les RSém des sous-réseaux dont nous savons très bien qu'ils correspondent à une information qui doit être présupposée par un lexème (elle va être rendue implicite à la faveur d'une lexicalisation). Nous reviendrons sur ce problème dans la Conclusion de cette étude en expliquant comment la notion de présupposition de lexème pourrait intervenir de façon un peu moins artificielle et *ad hoc* pour effectuer la transition sémantique dans un véritable cadre de génération de texte.

2) Ensuite, notre RSém de la figure V-11 n'est donc pas complète, en tant que RSém de (3), si le sous-réseau 'police**I** have institutionalized power**I.5** over pilots**I.1**' n'y est pas signalé comme étant une composante présuppositionnelle.

3) Finalement, la plus importante conclusion sera que LE ND D'UNE RSEM NE PEUT FAIRE PARTIE D'UNE COMPOSANTE PRESUPPOSITIONNELLE DE CETTE REPRESENTATION. Pour ce faire, on peut compléter comme suit la définition de la notion de *chemin communicatif*, donnée au chapitre précédent (en IV:3.2.2):

Définition plus complète de la notion de *chemin communicatif*:

> Il existe un *chemin communicatif* entre les nœuds sémantiques n_1 et n_2 si l'on peut trouver une succession de liens actanciels unissant ces deux nœuds, sans que le chemin qui est ainsi spécifié pour relier un nœud à l'autre ne traverse deux nœuds entretenant déjà une relation de dépendance communicative dans le réseau, NI NE TRAVERSE AUCUN NOEUD APPARTENANT à UNE COMPOSANTE PRESUPPOSITIONNELLE.

La mauvaise hiérarchisation que nous venons d'examiner provient du fait que la spécification initiale de nos règles de choix du ND était incomplète. Elle ne tenait pas compte de la possibilité d'avoir dans la RSém des sous-réseaux qui correspondent en réalité à des composantes présuppositionnelles. Voici donc décrit et expliqué un premier cas d'échec dans la hiérarchisation d'une RSém.

Supposons maintenant que nous sélectionnions, EN TRANSGRESSANT LE SYSTEME DE REGLES QUE NOUS AVONS PROPOSE, le nœud 'possess**1**' comme ND. Un tel choix conditionne un parcours tout différent du réseau sémantique — P_2* — parcours produisant la hiérarchie communicative de la figure V-16 (représentée sous une forme arborescente).

P_2* = {('possess**1**','pilots**I.1**') , ('possess**1**','plane**IV.3**') , ('pilots**I.1**','harass**1**') ,
('harass**1**','police**1**') , ('harass**1**','usually') , ('harass**1**','without respite'),
('harass**1**','means**1**') , ('means**1**','treat**I.1**') , ('treat**I.1**','police**I**') ,
('treat**I.1**','pilots**I.1**') , ('treat**I.1**','unjust') , ('police**I**','(have) power**I.5**') ,
('(have) power**I.5**','pilots**I.1**') ,
('(have) power**I.5**','institutionalized') , ('plane**IV.3**','Polish') }

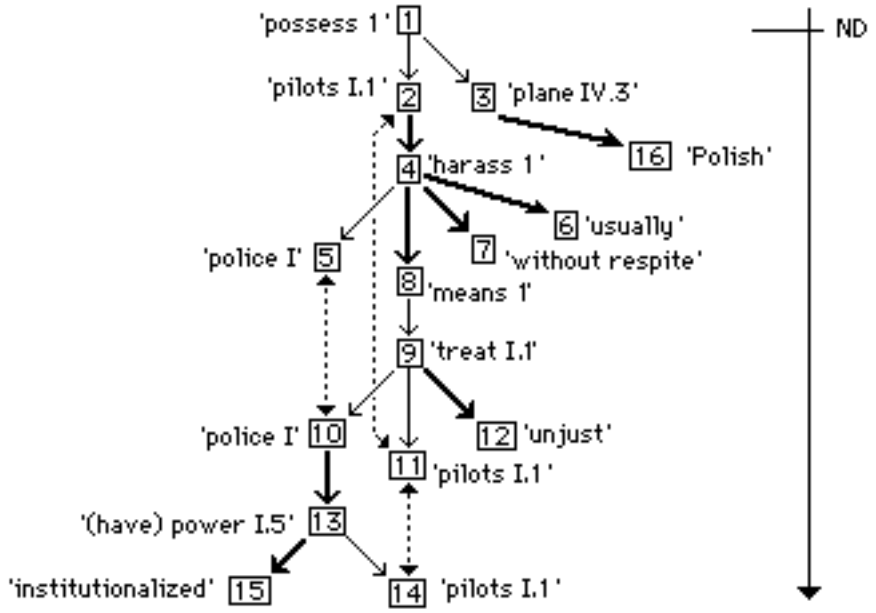


Figure V-16
Hiérarchisation communicative invalide de la RSém V-11
établie à partir du ND 'possess**1**'

Cette hiérarchie correspond à une phrase comme (16) dont nous allons montrer qu'elle n'est pas une paraphrase valide de (3):

(16) *Pilots who are persecuted by police possess a Polish plane.*

Il est certain que, communicativement, les phrases (3) et (16) ne sont pas perçues comme équivalentes. (3) dit quelque chose à propos de la police alors que (16) dit quelque chose à propos d'un certain groupe de pilotes. En utilisant la technique du questionnement pour identifier les rhèmes possibles, on voit que ces deux phrases ne semblent pas répondre aux mêmes questions:

(3') ← 'Who do the police persecute?'
(16') ← 'What do the pilots persecuted by police possess?'

La RSém de (16) aurait donc une structure communicative différente de celle de (3); son thème étant, grosso modo, 'pilots**I.1** who are harassed**1** ... by police**I**' et son rhème 'Polish plane**IV.3**'.

Cette différence est certainement importante puisque, la structure communicative faisant partie intégrante de la RSém, on peut considérer que deux RSém divergeant par leur structure communicative ne correspondent pas à deux mêmes Sens. (3) et (16) ne manifestent cependant pas seulement une différence de type communicatif. En effet, (16) ne dénote pas la même situation que (3).

Cette différence dans la dénotation apparaîtra plus clairement si l'on a recours à la logique. Les approches logiciennes de la sémantique linguistique (au sens large, non Sens-Texte) modéliseraient cette différence par deux types distincts d'implication apparaissant dans les expressions logiques associées à ces phrases. Soient (3'') et (16'') nos tentatives de représentation logique du contenu dénotationnel de, respectivement, (3) et (16):

(3'') $\forall \mathbf{x}$ ((pilot($\mathbf{x}$) $\wedge$
($\exists \mathbf{y}$ plane($\mathbf{y}$) $\wedge$ polish($\mathbf{y}$) $\wedge$ possess($\mathbf{x}$, $\mathbf{y}$)))
$\Rightarrow$ persecute(Police , $\mathbf{x}$))

(16'') $\forall \mathbf{x}$ (pilot($\mathbf{x}$) $\wedge$ persecute(Police , $\mathbf{x}$))
$\Rightarrow$ ($\exists \mathbf{y}$ plane($\mathbf{y}$) $\wedge$ polish($\mathbf{y}$) $\wedge$ possess($\mathbf{x}$, $\mathbf{y}$))

Ces formules ne prétendent nullement à une quelconque validité théorique dans le cadre d'une modélisation logique de la sémantique. Nous ne tenons pas compte ici, notamment, du problème de la portée des quantificateurs universel et existentiel. Notons que, sur ce plan, les phrases (3) et (16) sont ambiguës puisqu'elles correspondent soit à une situation où l'on considère un avion polonais particulier possédé par tous les pilotes ($\forall$ dans le champ de $\exists$) soit un avion différent possédé par chaque pilote ($\exists$ dans le champ de $\forall$). Dans nos formules, nous ignorons ce problème en sélectionnant une de ces quantifications — la seconde. On constate immédiatement que les deux implications logiques mises en jeu par de telles formules ne sont pas du tout équivalentes et que, donc, les valeurs de vérité de (3'') et (16'') ne coïncident pas. Pour mieux faire apparaître ces distinctions, nous visualisons, dans la figure V-17, les interprétations

dénotationnelles en termes de structures ensemblistes de (3'') et (16'') — c'est-à-dire de (3) et (16).

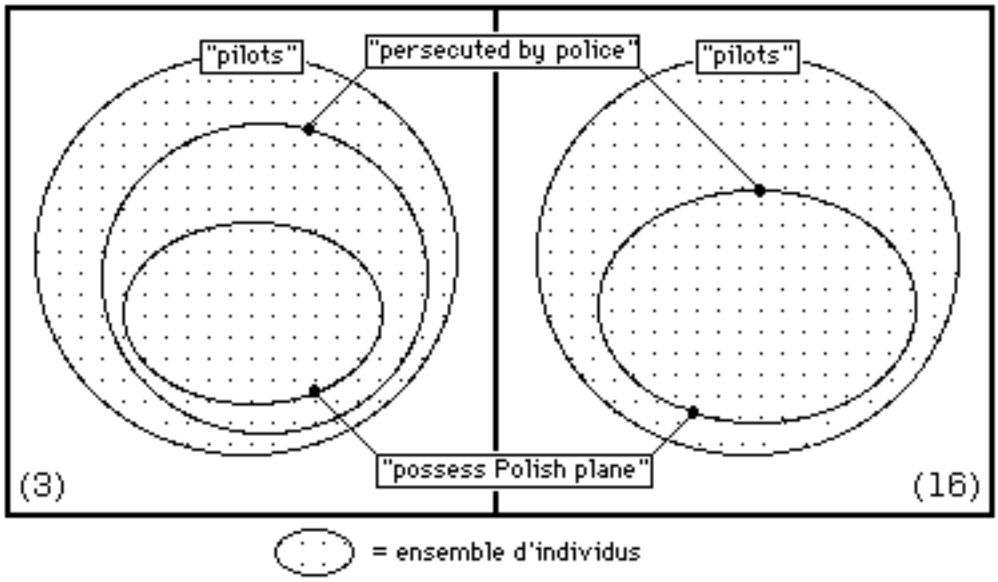


Figure V-17
Interprétations dénotationnelles ensemblistes de (3) et (16)

Dans le cas de (3), la police peut tout à fait persécuter d'autres pilotes que ceux ayant le malheur de posséder un avion polonais — c'est l'interprétation que nous avons adoptée pour (3) dans la figure V-17 en utilisant l'inclusion de deux ensembles. Dans le cas de (16), au contraire, les pilotes persécutés par la police sont nécessairement ceux et seulement ceux possédant un avion polonais — ce qui est rendu, dans la figure V-17, par deux ensembles équivalents d'individus.

Résumons ce que nous venons d'observer. Nous venons de montrer qu'en transgressant nos lois de détermination du ND nous pouvons être amené à construire une hiérarchie communicative correspondant à un Texte n'ayant pas le Sens désiré. Un tel Texte se verrait associé une RSém ayant une structure communicative tout à fait différente de celle considérée initialement. Cependant, une autre conséquence importante de cette mauvaise hiérarchisation est de mettre en évidence le fait que deux phrases qui diffèrent, selon

l'approche Sens-Texte, par la structure communicative de leur RSém, peuvent du même coup avoir deux dénotations différentes. Nous pouvons donc légitimement nous poser la question suivante:

> QUEL EST LE LIEN ENTRE LA STRUCTURE COMMUNICATIVE D'UNE RSEM ET L'INTERPRETATION DENOTATIONNELLE DE CETTE RSEM?

Pour répondre à cette question, nous devons, une fois encore, nous situer délibérément dans une perspective synthétique plutôt qu'analytique. Il nous semble tout d'abord que toutes les paraphrases langagières de (3) permises par un MST devraient dénoter la même situation que (3). Il en va de même pour (16). Cela revient à dire que l'information présente dans la RSém doit suffire à refléter indirectement la différence dénotationnelle entre (3) et (16). La composante Représentation Conceptuelle ⇔ Texte d'un système de génération doit donc être équipée des règles et procédures permettant de calculer quelle RSém va être apte à refléter une structure de dénotation donnée. Cela se traduit essentiellement par une connaissance de l'effet produit par la combinaison des quantificateurs linguistiques mais aussi, comme nous venons de le montrer, par une connaissance de l'effet de l'interaction avec une structure communicative donnée. Bien que nous ne puissions pas proposer ici de règles de ce type, nous pouvons, après examen de ce deuxième cas d'échec dans la hiérarchisation, tirer la conclusion suivante:

> La structure communicative interagit avec les quantificateurs de la structure sémantique pour refléter une dénotation donnée. Dans un cadre de transition RConcept ⇒ Sens, il faut donc tenir compte de cette interaction.

Dans cette conclusion, nous insistons bien sur le fait que c'est l'interaction quantificateurs-structure communicative qui reflète une dénotation. Il est ainsi toujours possible d'utiliser une structure communicative donnée, comme celle de la RSém de (16), pour dénoter

une situation comme celle dénotée par (3), pour peu que l'on fasse varier les quantificateurs utilisés. (17) est un exemple de phrase qui pourrait être ainsi produite:

(17) *Some of the pilots who are persecuted by the police possess a Polish plane.*

Nous n'irons pas plus loin dans ce type de considération puisque cela dépasse largement le cadre de l'étude des procédures de transition sémantique.

Nous venons donc d'examiner deux cas d'échec dans la hiérarchisation. Le premier nous a conduit à préciser le rôle de la présupposition dans la structuration communicative (au sens large) des RSém. Nous avons ainsi proposé de raffiner la notion de chemin communicatif pour restreindre l'ensemble des ND potentiellement valides dans une RSém comportant des composantes présuppositionnelles. Le second cas d'échec nous a permis de démontrer l'importance de respecter, lors de la hiérarchisation, les contraintes imposées aux procédures par la structure communicative. Nous avons aussi "soulevé un nouveau lièvre" en montrant que la structure communicative interagissait avec les quantificateurs de la structure sémantique pour refléter, au niveau linguistique, une dénotation donnée[11].

Nous allons maintenant, pour conclure la présentation des contrôles additionnels imposés aux procédures de hiérarchisation, introduire une notion supplémentaire que nous avons appelée l'*imperméabilité des sous-réseaux* ***thème*** *et* ***rhème***.

---

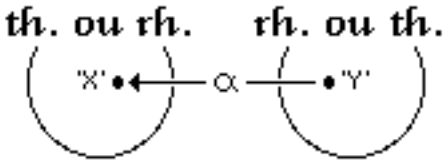
[11] Ceci correspond à la façon "Sens-Texte" d'aborder certains phénomènes souvent étudiés selon une approche logicienne de la description linguistique (cf. Dowty *et al.* 1981).

4.3 Imperméabilité des sous-réseaux **thème** et **rhème**

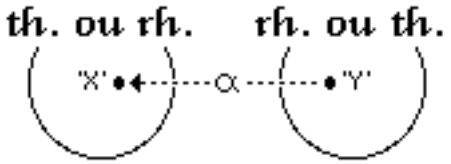
Un Texte doit pouvoir rendre compte de la dualité existant entre la composante thématique et la composante rhématique du Sens qu'il véhicule. Pour modéliser cela au niveau de la transition sémantique, il faut considérer que les sous-réseaux **thème** et **rhème** sont hiérarchisés de façon séparée. Pour les procédures de parcours des RSém, cela signifie que le parcours du sous-réseau **thème** ne doit pas mener au parcours du sous-réseau **rhème** et vice versa. C'est ce que nous avons appelé l'*imperméabilité des sous-réseaux **thème** et **rhème***, notion que l'on peut définir de la façon suivante:

> Les sous-réseaux **thème** et **rhème** sont en quelque sorte imperméables l'un par rapport à l'autre lors de la hiérarchisation. En conséquence, le seul parcours possible pour chacun d'entre eux doit se faire à partir du ND par le biais d'un chemin communicatif sur lequel ce ND se situe.

L'imperméabilité des sous-réseaux **thème** et **rhème** permet donc d'établir une contrainte supplémentaire sur le parcours sémantique. Pour spécifier de façon formelle une telle contrainte, nous dirons que LE PARCOURS SEMANTIQUE EST BLOQUE DANS LES DEUX CAS DE FIGURE SUIVANTS:



- lorsqu'à partir d'un nœud 'X', appartenant au **thème** ou au **rhème**, on pourrait effectuer un parcours ascendant vers un nœud 'Y' appartenant, lui, à l'autre sous-réseau communicatif.



- lorsqu'à partir d'un nœud 'X', appartenant au sous-réseau **thème** ou **rhème**, on pourrait effectuer un parcours de ce sous-réseau mais que 'X' a été atteint par un parcours descendant à partir d'un nœud 'Y' appartenant, lui, à l'autre sous-réseau communicatif.

Nous voulons, par une telle contrainte, éviter qu'un sous-réseau communicatif ne soit parcouru, et donc hiérarchisé, à partir d'un autre sous-réseau communicatif sans que soit mis en évidence le rôle de transition que doit jouer le ND.

L'exemple de la figure IV-5, que nous avons déjà utilisé pour expliciter le fonctionnement du parcours sémantique, va nous permettre d'illustrer notre dernière contrainte sur ce parcours. On note dans la figure V-6 que le parcours ascendant effectué à partir du nœud 'promenade**1**' n'a pas mené au parcours du sous-réseau **rhème**. Or, cela aurait pu se produire si nous avions traversé 'connaître**I.3**' à partir de 'elle'. Cependant, le nœud 'connaître**I.3**' appartenant au rhème, nous ne pouvons le traverser à partir de 'elle' qui, lui, est un nœud sémantique du thème. Dans le cas où le parcours du rhème se fait avant celui du thème, 'elle' peut être atteint par un parcours descendant à partir de 'bonheur**II.1**' ou connaître**I.3**', mais ce parcours descendant ne peut se poursuivre par un parcours ascendant à l'intérieur du thème.

Nous laissons au lecteur le soin de tenter, sur la figure IV-5, des parcours qui ne respecteraient pas cette contrainte. Il pourra constater que les différentes hiérarchisations communicatives ainsi construites sont en contradiction avec les hiérarchies syntaxiques que l'on voudrait voir associées à une telle RSém. Une nouvelle fois, on peut noter que notre technique de parcours des RSém cherche à modéliser une certaine lecture de ces représentations qui respecte

l'information véhiculée par leurs structures sémantique ET communicative.

5 PARCOURS DES GRAPHES CONCEPTUELS PROPOSE DANS SOWA 1984

Pour terminer ce chapitre sur le parcours des réseaux, il est utile d'élargir la perspective envisagée jusqu'ici et d'examiner quelques propositions faites dans un ouvrage relativement récent pour traiter le même type de problème. Il s'agit de la génération de texte à partir de graphes conceptuels, telle que présentée dans Sowa 1984[12]. Nous avons évoqué précédemment le formalisme des graphes conceptuels élaboré par J. Sowa (cf. I:3). Rappelons qu'il est une alternative au formalisme de la logique permettant de représenter les connaissances afin de construire des "modèles" et de raisonner sur ces représentations et modèles.

Une petite partie de l'ouvrage de J. Sowa (Chapitre 5:4) est consacrée à la génération de texte à partir des graphes conceptuels — nous parlerons quant à nous ici de synthèse de Texte. Plus spécifiquement, Sowa s'attache à montrer comment on peut établir, sur un graphe donné, différents parcours dont chacun est associé à la construction d'une phrase exprimant linguistiquement le contenu conceptuel du graphe. Sowa appelle de tels parcours *chemins énonciatifs* ("utterance paths"). On voit immédiatement le lien entre notre notion de parcours des RSém et celle de chemin énonciatif. Il est même évident que la notion de ND, liée à celle de parcours sémantique, va elle aussi se retrouver d'une certaine façon dans l'approche de Sowa. Bien que nous ne pensions pas que l'auteur présente une véritable "théorie générale" de la correspondance Graphe Conceptuel ⇔ Texte, ses propositions sur la question sont suffisamment précises et logiquement organisées pour qu'il faille ici les considérer relativement

---

[12] Certaines des idées présentées ici semblent avoir été initialement proposées dans Sowa 1968, mais vraisemblablement sous une forme moins développée. Comme, de plus, ce texte n'a pas été publié, nous nous baserons ici uniquement sur Sowa 1984.

en détail. Pour mieux cerner toute la parenté existant entre notre traitement de la transition RSém ⇒ RSyntP et le traitement proposé par Sowa, citons les propres termes de ce dernier (p. 233):

> "... the same graph can be expressed in many different sentences, depending on the starting point and direction of the utterance path. Yet not all word orders are possible: the utterance path visits each node a limited number of times, and a concept can only be uttered as a word at one of those visits."

On doit vraisemblablement à McNeill 1979 la première présentation de la notion de chemin énonciatif, qui, en ce sens, n'a pas été élaborée par Sowa[13]. Nous ne nous basons cependant pas ici sur les travaux de D. McNeill pour des raisons en partie opposées à celles qui nous ont fait choisir les thèses de Sowa 1984 comme point de comparaison:

1) Les travaux de McNeill, limités à une approche psychologique de la production de la parole, sont aujourd'hui beaucoup moins connus et exploités que ceux de Sowa, dont la portée est bien plus générale.
2) Nous avons vu en I:3 qu'un des reproches que l'on peut faire aux représentations conceptuelles proposées par Sowa est qu'elles établissent des hiérarchies entre les concepts qui semblent souvent motivées par les hiérarchies syntaxiques des Textes correspondants. La critique s'applique encore plus directement à McNeill qui, lui, structure explicitement ses représentations en fonction de ce qu'il appelle les "séquences d'encodage", c'est-à-dire, en fait, l'ordre linéaire (voir, par exemple, le graphe proposé dans McNeill 1979:101 pour *Luther*

---

13 On pourrait bien sûr retracer la genèse de cette notion dans, par exemple, les travaux beaucoup plus spécialisés sur la théorie des graphes orientés; cf. Harary, Norman & Cartwright 1965.

pours the coffee)[14]. Or, justement, la finalité d'un modèle linguistique Sens-Texte est de montrer comment on peut passer d'une représentation sémantique, multidimensionnelle par nature, à la représentation unidimensionnelle (linéaire) du Texte à travers toutes les règles "arbitraires" de la langue. Nous ne pouvons donc pas nous inspirer d'une approche qui tente d'ignorer, par des détours formels, une caractéristique fondamentale du fait linguistique[15].

3) Les aspects de McNeill 1979 véritablement importants pour notre discussion se retrouvent présentés dans Sowa 1984. L'avantage de ce dernier ouvrage est qu'il attache plus d'importance à la validation de la notion de chemin énonciatif en fonction de considérations grammaticales qu'en fonction de considérations psychologiques. Il est donc nécessairement plus lié à nos préoccupations présentes.

Compte tenu du lien existant entre ce que nous allons présenter ici et notre propre approche, on pourrait se demander pourquoi cette partie de l'exposé arrive en conclusion et non en introduction de notre examen de la notion de parcours des RSém. Il nous semble cependant préférable de procéder de la sorte pour au moins deux raisons. Premièrement, notre approche de la transition sémantique n'a pas été initialement inspirée par les propositions de Sowa, ou de McNeill, mais s'est imposée d'elle-même à partir de l'examen du problème posé. Deuxièmement (et c'est une conséquence du premier argument), nos idées ne peuvent être comprises comme une adaptation des thèses de Sowa 1984; au contraire, ces dernières

---

14 Une conséquence de cela est que le parcours énonciatif, contrairement à notre parcours des RSém, suivra toujours le sens des relations entre les concepts (c'est-à-dire, le sens des flèches dans les graphes).

15 Cette critique un peu sévère des travaux de McNeill ne vaut que d'un point de vue linguistique. Il se pourrait, par ailleurs, que ses thèses soient psychologiquement tout à fait valides. Dans ce cas, elles correspondraient à un "plan" d'étude du phénomène langagier qui se superposerait au plan d'étude strictement linguistique, sans en être un prolongement au niveau conceptuel.

permettent, a posteriori, de faire ressortir la généralité "procédurale" de certaines notions que nous avons développées.

Nous allons maintenant présenter, en le commentant, le traitement que fait Sowa 1984 d'un exemple de graphe conceptuel dans un cadre de synthèse de Texte.

La figure V-18, ci-dessous, est un graphe conceptuel sur lequel nous avons fait figurer un chemin énonciatif: chaque étape du chemin est représentée par un numéro encadré par des pointillés. Seuls les concepts (et non les relations) sont considérés comme des étapes dans le chemin énonciatif qui correspond à la production de la phrase suivante:

(18) *Blithe babies with fat bellies drink fresh milk in new bottles.*

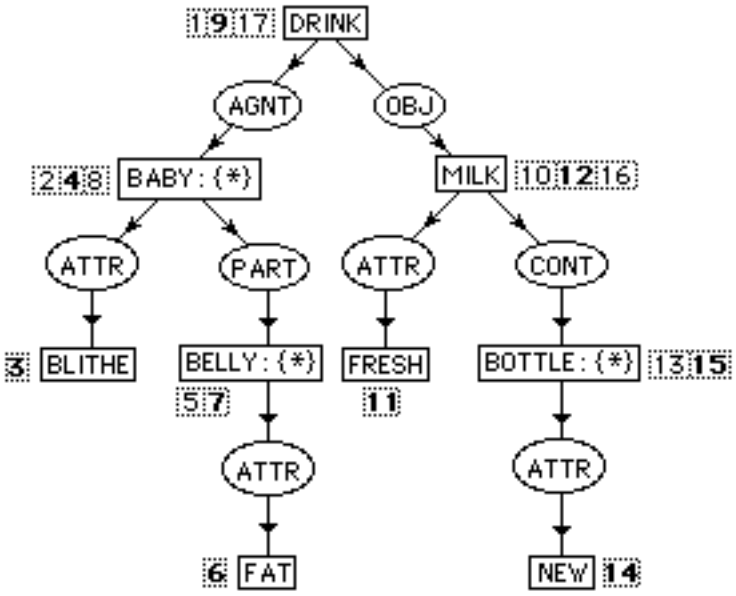


Figure V-18
Graphe conceptuel avec chemin énonciatif (d'après Sowa 1984)

Le chemin énonciatif commence par ce que Sowa appelle le "prédicat principal", ici DRINK, et doit passer à travers tous les nœuds du graphe au moins une fois. Il existe un lien direct entre l'accès à un nœud lors du cheminement énonciatif et l'énonciation

d'un mot de la langue correspondant dans la phrase que l'on produit. Ainsi, lorsque des nœuds "terminaux" comme FAT ou NEW seront atteints, il faudra automatiquement déclencher l'énonciation de *fat* et *new*. Les nœuds non terminaux peuvent, eux, déclencher la réalisation linguistique au cours de l'un des passages, sans que ce soit nécessairement lors du premier ou du dernier passage. Dans la figure V-18, nous avons numéroté en gras les passages correspondant à une réalisation linguistique pour la synthèse de la phrase (18). On peut noter que les relations présentes dans le graphe conceptuel seront utiles aussi bien pour établir le chemin énonciatif que pour introduire dans la phrase certains lexèmes comme *with* ou *in*. L'algorithme construisant le chemin énonciatif a accès à la grammaire de la langue dans laquelle on veut synthétiser le Texte. Les connaissances grammaticales permettent de contrôler l'ordre de passage à travers les arcs et le choix du passage qui déclenchera la réalisation linguistique d'un nœud conceptuel. Il existe aussi un ensemble de règles générales de contrôle de l'algorithme sur lequel nous ne nous attarderons pas ici.

On retrouve donc chez Sowa, outre la notion de parcours (ou chemin énonciatif), la notion de prédicat principal qui pourrait correspondre grosso modo à notre ND des RSémR. Cependant, Sowa ne parle pas véritablement du problème du choix de ce prédicat principal et semble présupposer que le graphe conceptuel à traiter possédera de façon évidente un tel prédicat. La structure générale du graphe proposé en exemple est en tout cas déjà suffisamment "syntaxique" pour que ce type de problème ne se pose pas ici.

Résumons donc la stratégie de Sowa qui vient d'être présentée. Pour produire un Texte, on dispose d'un graphe conceptuel qui correspond à un certain contenu "sémantique" que le Texte doit exprimer. Un algorithme assez général de parcours de ce graphe va permettre, tout en associant un lexème aux concepts et relations lorsqu'ils sont atteints, de produire la suite linéarisée des mots composant une phrase. Cet algorithme est contrôlé par des paramètres reflétant certaines caractéristiques très générales du

langage cible (linéarisation de type Sujet-Verbe-Objet, Sujet-Objet-Verbe, etc.). Il est aussi contrôlé par les règles grammaticales et lexicales qui permettent, de plus, d'ajouter l'information purement linguistique requise pour construire une phrase correcte. On voit donc que Sowa veut que le parcours du graphe reflète directement l'ordre linéaire des lexèmes dans la phrase. Il faut alors constater que, malgré le recours à des règles de grammaire que Sowa reconnaît comme nécessaire, son approche ignore en fait la plupart des notions qui pour nous sont fondamentales dans la description de la correspondance Sens ⇔ Texte: la notion dont l'absence se manifeste de la façon la plus évidente étant peut-être celle de dépendance syntaxique.

Arrivé à ce point de notre exposé, nous devons brièvement expliquer pourquoi, selon nous, le type d'approche proposé par Sowa doit être écarté lorsque l'on veut modéliser le processus de synthèse linguistique. Nous sommes convaincu que c'est une véritable ERREUR METHODOLOGIQUE que de tenter d'associer directement une suite de mots à une représentation conceptuelle par un parcours de cette représentation. Nous pensons que le parcours d'une représentation multidimensionnelle du contenu de la phrase, qui détermine ce que nous avons choisi d'appeler une hiérarchie communicative, doit au bout du compte permettre de mettre à jour une hiérarchie syntaxique. Les dépendances syntaxiques seront ensuite associées à des linéarisations, selon les règles de la grammaire. Tous ces phénomènes et toutes les caractéristiques d'une langue donnée qui permettent de les mettre en jeu doivent être considérés séparément car ils manifestent une autonomie logique. Dans les faits, il est vrai qu'il existe un lien entre ordre linéaire et relations sémantiques, mais ce lien est, même dans une langue comme l'anglais, si indirect et si complexe à définir que l'on ignore la réelle économie du système linguistique en tentant de le matérialiser directement par quelque procédé que ce soit (comme par un ordre de parcours d'une représentation conceptuelle ou sémantique).

Il nous semble intéressant que, en dépit de postulats inadéquats sur la langue, Sowa parvienne tout de même à mettre en évidence certaines notions qui ont un véritable intérêt quant à l'activation procédurale d'un modèle linguistique. Une telle constatation semble plaider un faveur d'un point de vue énoncé dans différents écrits sur la TST (cf. Mel'čuk à paraître) et voulant que seuls les aspects purement déclaratifs de la construction d'un modèle d'une langue soient véritablement linguistiques. Si les procédures que nous cherchons à définir peuvent en effet être décomposées en micro-procédures ayant une validité quel que soit le type de "raisonnement" opéré sur des formalismes, nous pensons pourtant, quant à nous, que le système procédural global n'est pas général. Il est destiné à la manipulation de données et de représentations linguistiques et doit donc s'articuler autour de notions développées à cette seule fin, qui sont elles aussi linguistiques. Un tel système n'a aucune raison d'être en dehors du travail pour lequel il a été conçu et nous n'avons aucun scrupule à le caractériser comme LINGUISTIQUE. Que certains aspects des procédures développées ici se retrouvent utilisés dans d'autres disciplines, pour d'autres fins, ne doit pas surprendre. Après tout, personne ne s'étonne du fait que le formalisme de graphe puisse être utilisé aussi bien pour représenter le contenu sémantique d'un Texte, les stratégies possibles en résolution de plans, un câblage électrique, etc. Il semblerait absurde d'utiliser cette versatilité du formalisme de graphe comme un argument pour rejeter la sémantique de l'étude linguistique.

Malgré toutes ces critiques, il faut reconnaître que la stratégie globale proposée par Sowa repose exactement sur le même principe que nous avons nous-même énoncé plus haut. À savoir que, considérant une représentation sémantique initiale sous forme de réseau, la production des différentes paraphrases possibles permettant d'exprimer le Sens encodé dans cette représentation sera modélisée par différents parcours pouvant être accomplis à travers la structure du réseau. Nous considérons donc cette similitude de vues sur les aspects strictement procéduraux du problème de la production

de Texte à partir de réseaux comme un argument en faveur de notre modélisation de la transition RSém ⇒ RSyntP.

CHAPITRE VI

Arborisation: activation des règles de correspondance sémantique proprement dites

Ce dernier chapitre présente les procédures d'*arborisation*, c'est-à-dire les procédures permettant d'activer les *règles de correspondance sémantique proprement dites* dans un cadre de transition RSém ⇒ RSyntP.

Nous avons montré que l'arborisation s'effectuait à partir d'une RSémR qui est une RSém ayant été traitée par les trois opérations de *hiérarchisation*, *grammaticalisation* et *lexicalisation*. La transition considérée ici est donc plus précisément une transition RSémR ⇒ RSyntP.

Nous avons aussi décrit en détail, dans les chapitres précédents, comment s'effectuait la hiérarchisation et quel était le statut formel de la RSémR. On trouve, dans la Partie **1** ainsi que dans la présente Partie, quelques idées nouvelles quant à la façon dont devraient être utilisées les règles de paraphrase sémantique pour effectuer la lexicalisation. Les deux opérations de lexicalisation et grammaticalisation ne sont cependant pas étudiées en détail au niveau des procédures et nous passons directement à la dernière des

opérations constituant la transition sémantique. Ce chapitre sur l'arborisation est structuré de la façon suivante:

- la première section explique pourquoi et comment l'algorithme de hiérarchisation est utilisé dans l'arborisation pour ordonner l'application des règles de correspondance sémantique;
- la seconde partie montre comment la technique proposée fonctionne dans la pratique, à partir d'un exemple de RSémR relativement complexe.

Cette description de l'opération d'arborisation est courte et, nous l'espérons, d'un abord très simple. En effet, les notions auxquelles il va être fait appel maintenant ont toutes été présentées et discutées dans les chapitres précédents. Nous ne faisons ici qu'utiliser les outils déclaratifs et procéduraux qui définissent la composante sémantique d'un MST.

1 UTILISATION DE L'ALGORITHME DE HIERARCHISATION DANS LE CONTROLE DE L'ARBORISATION

1.1 Justification de la méthode proposée

Nous avons décrit au chapitre précédent le lien direct qui existe entre la hiérarchie communicative établie sur une RSém et la hiérarchie syntaxique contenue dans une RSyntP associée à cette RSém. L'algorithme de parcours sémantique est conçu de telle façon qu'il construit la hiérarchie communicative du haut vers le bas à l'aide d'une stratégie mixte: parcours descendants et parcours ascendants alternés. C'est le même algorithme qui va être utilisé pour sélectionner une à une les dépendances sémantico-communicatives de la RSémR et leur associer une dépendance syntaxique profonde. Nous pouvons donc formuler la règle centrale de contrôle de la transition sémantique:

La RSyntP, comme la hiérarchie communicative, va être construite du haut vers le bas en suivant la progression d'un parcours sémantique mené à travers la RSémR.

S'il n'était pas nécessaire d'effectuer les opérations de lexicalisation et de grammaticalisation, un seul parcours sémantique suffirait pour hiérarchiser communicativement la RSém tout en construisant la RSyntP. Le fait que des règles de paraphrase sémantique et des règles grammatico-sémantiques (i) s'appliquent sur un même niveau de représentation sémantique, et (ii) fonctionnent sur une RSém hiérarchisée, implique qu'il faille effectuer plusieurs parcours sémantiques.

Le dernier parcours a une double fonction:

- hiérarchiser l'ultime version de la RSém en associant une relation de dépendance communicative à chaque relation de dépendance sémantique — condition *sine qua non* de l'application des règles de correspondance sémantique (cf. II:4);
- contrôler la construction de l'arbre syntaxique profond.

Deux remarques doivent être faites à propos du nombre de parcours sémantiques nécessaires à la transition sémantique:

1) Il serait souhaitable que la lexicalisation et la grammaticalisation s'effectuent tout en conservant la trace de la hiérarchie communicative établie sur la RSém. Ainsi, lorsqu'un Sens de lexème est substitué à un sous-réseau sémantique (la définition du lexème), il doit hériter des liens de dépendance communicative que le nœud dominant du sous-réseau (le Sens générique) entretenait avec les autres composantes de la RSém. En cas de substitution inverse, le sous-réseau HIéRARCHISé est inséré dans la RSém et son nœud dominant hérite de la "place" dans la hiérarchie communicative qu'occupait le Sens décomposé. Ces règles sont cependant très générales et

passent outre de nombreux problèmes posés par ce type de substitution, problèmes qui devraient être résolus par une étude dédiée aux opérations de lexicalisation et de grammaticalisation[1].

2) Les options dans le parcours sémantique correspondent à plusieurs hiérarchisations possibles de la RSém et donc à la production de plusieurs RSyntP paraphrastiques. Si le but de la transition est de construire toutes les paraphrases à partir d'une RSém, il faudra tenter plusieurs parcours visant à la production de transitions "concurrentes". EN THEORIE, il faut explorer toutes les transitions possibles. DANS LA PLUPART DES APPLICATIONS EN GENERATION DE TEXTE, on se contentera d'utiliser des heuristiques permettant de sélectionner la transition menant à la RSyntP la plus "prometteuse".

Une mise au point s'impose ici quant à un contre-sens fréquemment rencontré à propos de la notion de *pouvoir paraphrastique des MST*. Cette notion ne réfère pas NECESSAIREMENT, dans un contexte d'application, aux possibilités qu'offre un MST de produire automatiquement un grand nombre de paraphrases à partir d'une même RSém. Dans bien des applications véritables, produire plusieurs paraphrases semble relever plutôt du "jeu pour linguistes" et ne présente strictement aucun intérêt pratique.

L'approche paraphrastique a au moins, cependant, deux avantages.

Tout d'abord, si le pouvoir paraphrastique permet bien sûr de synthétiser sur demande des paraphrases en faisant faire des "retours en arrière" au synthétiseur linguistique, il permet aussi et surtout de construire un synthétiseur véritablement puissant. La puissance d'un synthétiseur est en rapport avec l'imprévisibilité de son comportement. Ainsi, on ne peut pas, en examinant une RSém,

---

[1] Faute de techniques sophistiquées pour gérer les substitutions dans la RSém tout en préservant une hiérarchie communicative, il est toujours possible — et c'est ce qui a été implémenté dans GOSSIP — d'effectuer une nouvelle hiérarchisation chaque fois que la configuration de la RSém est modifiée. Cette solution n'est bien sûr pas élégante.

dire immédiatement quel Texte va lui être associé: même en connaissant le MST utilisé et même en connaissant les différentes procédures qui contrôlent l'activation de ce MST. Bien trop de choix peuvent intervenir et chacun de ces choix est tributaire des micro-structures (sémantiques, syntaxiques, morphologiques, etc.) qui doivent être traitées, des "micro-échecs" pouvant intervenir dans leur traitement et, finalement, des heuristiques permettant de trancher en cas de choix véritable.

Finalement, un autre avantage de l'approche paraphrastique est de permettre d'ignorer en grande partie le problème des culs-de-sac, déjà mentionné plus haut. Plus le pouvoir paraphrastique d'un modèle fonctionnel[2] est grand, moins l'activation synthétique de ce modèle a de chances de nous mener à des culs-de-sac incontournables et plus le traitement séquentiel de la synthèse — du niveau le plus profond vers le niveau le plus superficiel, a de chances de réussir.

Le fondement paraphrastique de la description linguistique Sens-Texte ne vise donc pas seulement à satisfaire les aspirations purement théoriques des linguistes. Il garantit, dans la pratique, la fonctionnalité même des modèles construits: c'est-à-dire, leur capacité à mettre en relation un Sens et un Texte.

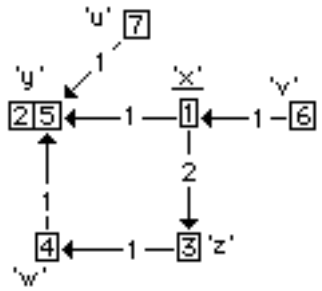
1.2 Fonctionnement des procédures

Maintenant qu'a été établi le lien entre parcours sémantique et arborisation, voyons comment se construit véritablement un arbre syntaxique profond.

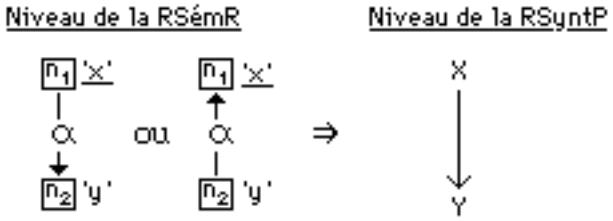
Soit la structure sémantique d'une RSémR sur laquelle figure un parcours sémantique:

---

[2] Au sens de FONCTIONNEL**II**, cf. Introduction.



Les Sens étiquetant les nœuds ne nous intéressent pas ici; seules comptent les relations actancielles, ainsi que l'ordre dans lequel une telle structure est parcourue par l'algorithme décrit au chapitre précédent. Lorsque cet algorithme est utilisé pour l'arborisation, chaque arc est considéré successivement dans l'ordre du parcours, puis est traduit en une micro-structure syntaxique profonde grâce aux règles de correspondance sémantique de la grammaire. Si nous utilisons le symbolisme suivant pour représenter les transitions opérées:



nous pouvons visualiser la construction progressive de la RSyntP associée à la RSémR considérée:

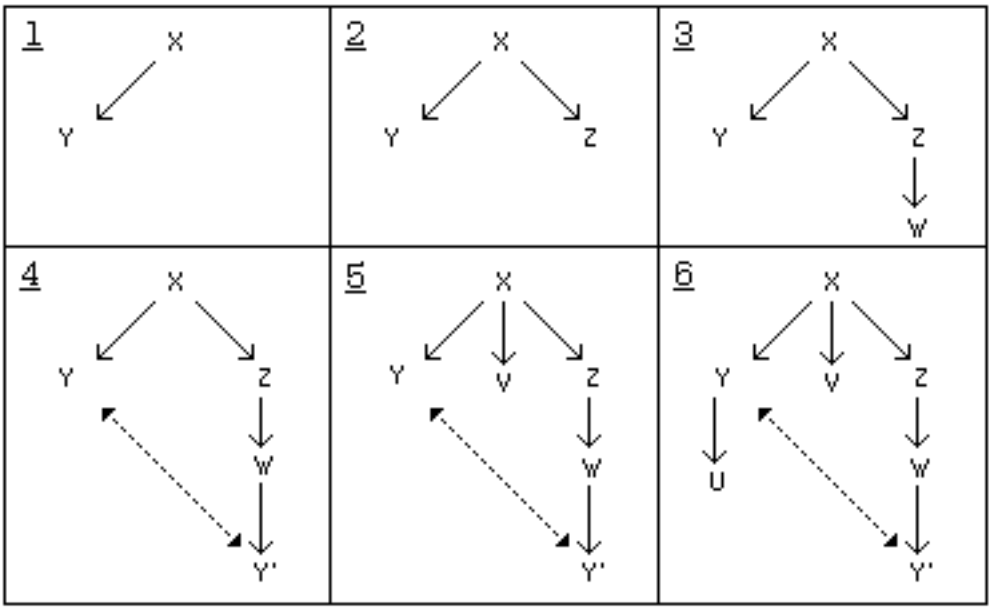


Figure VI-1
Construction progressive d'une RSyntP

Bien sûr, les structure syntaxiques profondes du types X→Y peuvent elles-mêmes être relativement complexes, selon la nature exacte de la configuration sémantique traitée.

En observant l'arbre apparaissant à l'étape 4 de la figure VI-1, on remarque le fait suivant:

> Si un même nœud sémantique est traversé plusieurs fois, cela entraîne la création de liens de co-référence dans l'arbre syntaxique profond.

Les liens de co-référence peuvent aussi trouver leur origine dans les règles de correspondance sémantique elles-mêmes; plusieurs règles proposées dans le Chapitre II et dans son Annexe (Partie **1**) illustrent ce cas de figure.

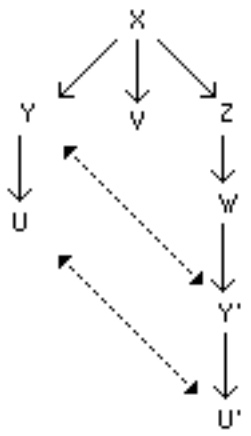
Nous avons signalé en I:2.2 que les représentations linéaires des réseaux sémantiques devaient contenir des nœuds sémantiques étiquetés non seulement par des Sens de lexèmes mais aussi par des numéros de référence. Ces numéros permettent de manipuler linéairement des réseaux présentant plusieurs occurrences d'un

même Sens, sans que la formule utilisée ne soit ambiguë. Un arbre syntaxique est a priori beaucoup plus aisé à représenter linéairement qu'un réseau sémantique. Ainsi, l'arbre de l'étape 3, figure VI-1, pourrait être représenté par une formule comme: X (Y , Z (W)). Cependant, dans le cas où il faut faire apparaître des liens de co-référence dans l'arbre, on doit à nouveau avoir recours aux numéros de référence utilisés pour linéariser le réseau sémantique; comme le montre le cas de l'étape 4, figure VI-1: X (Y_i , Z (W (Y'_i))). Dans notre implémentation de la transition sémantique, nous avons donc choisi d'attribuer un numéro de référence à chaque nœud de la RSém et d'attribuer aux nœuds des arbres syntaxiques le même numéro de référence que le Sens dont ils tirent leur origine. Grâce à ce procédé très simple, la structure anaphorique est automatiquement encodée dans les RSyntP.

Une autre caractéristique intéressante de l'arborisation présentée dans la figure VI-1 est que les liens de co-référence nous servent aussi à transférer les dépendances syntaxiques. En effet, il faut interpréter dans la figure VI-1 la double flèche pointillée liant Y à Y' comme signifiant deux choses:

- Y et Y' sont des nœuds co-référentiels — ils correspondent à un unique nœud sémantique 'y';
- toutes les configurations dépendantes de Y sont potentiellement dépendantes de Y', et vice versa.

Le second point est très important si l'on veut bien interpréter les RSyntP que construisent nos procédures d'arborisation. En effet, dans la figure VI-1, on pourrait aussi bien considérer que l'on a une relation Y'→U plutôt que Y→U. La représentation finale (étape 6) est en fait, selon notre interprétation, formellement équivalente à la représentation suivante:



Y→U et Y'→U' étant des configurations strictement parallèles. Cette utilisation des liens de co-référence permet de simplifier le processus de construction des RSyntP tout en évitant de surcharger ces représentations. C'est aux règles et procédures de transition RSyntP ⇒ RSyntS qu'il appartient de déterminer comment doit être exprimée en surface l'information présente dans la RSyntP; elles doivent notamment choisir, selon les configurations syntaxique et communicative, la meilleure façon de connecter un sous-arbre lorsque plusieurs nœuds co-référentiels peuvent en être le sommet. Les trois exemples ci-dessous illustrent ce type de choix[3] :

(1) *...un chien stupide qui aboie pendant que la caravane passe en ignorant ce chien*

(2) *...un chien stupide qui aboie pendant que la caravane passe en l'ignorant*

(3) *...un chien qui aboie pendant que la caravane passe en ignorant ce chien stupide*

Le traitement des options disponibles pour la production de (1), (2) ou (3) est un problème intéressant mais qui relève d'une étape postérieure à l'arborisation et sort donc du cadre de cette étude.

---

[3] Nous utilisons ici volontairement des syntagmes nominaux et non des phrases complètes afin de neutraliser certaines variations communicatives qui pourraient fausser le caractère paraphrastique que nous voulons donner à nos exemples.

1.3 Structure communicative de la RSyntP

La technique d'activation des règles de correspondance sémantique qui vient d'être présentée ne permet que de produire la structure syntaxique profonde et la structure anaphorique de la RSyntP. Il nous reste à montrer comment la structure communicative syntaxique profonde est construite à partir de la structure communicative sémantique.

On pourrait penser que la RSyntP hérite sa structure communicative directement de la RSém. Le procédé qui semble la plus logique est en effet, durant la construction de l'arbre syntaxique, de simplement attribuer à chaque nœud de la RSyntP les caractéristiques communicatives du nœud de la RSémR dont il tire son origine. Cette approche est globalement bonne. Il faut cependant tenir compte de certains cas particuliers de configurations sémantiques, qui pourraient introduire des incohérences communicatives au niveau syntaxique profond si la traduction de la structure communicative sémantique était faite de façon trop rigide. Le problème envisagé ici peut être illustré par la figure suivante, qui montre la production d'une structure communicative syntaxique profonde non valide:

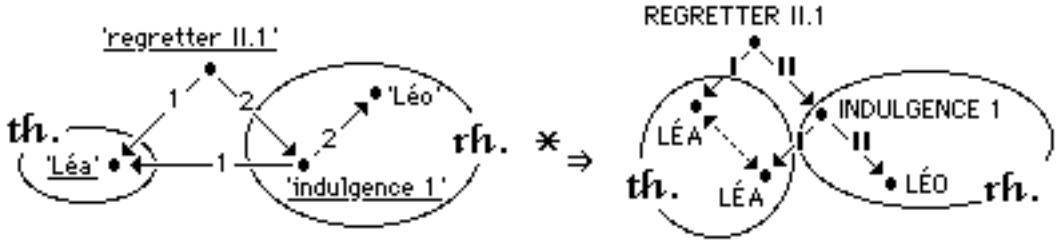


Figure VI-2
Arborisation produisant une structure communicative syntaxique profonde non valide

La RSémR figurant ci-dessus correspond à une phrase comme (4):

(4) *Léa regrette son indulgence envers Léo.*
← 'Qu'est-ce que regrette Léa?'

Il nous semble que la structure communicative syntaxique profonde résultant de la transition de la figure VI-2 ne s'harmonise pas avec la structure syntaxique profonde proprement dite. L'opposition thème-rhème, dans la RSyntP considérée ici, doit jouer entre les deux actants syntaxiques profonds de REGRETTER**II.1**; l'occurrence de LÉA comme actant de INDULGENCE**1** ne doit donc pas faire partie du thème. Cette occurrence de LÉA doit hériter de la valeur communicative de INDULGENCE**1**, c'est-à-dire faire partie du rhème syntaxique profond. Nous considérerons que la RSyntP qui doit être associée à la RSémR de la figure VI-2 est la suivante:

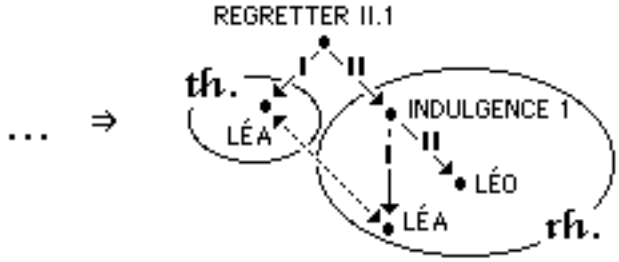


Figure VI-3
Arborisation produisant une structure communicative syntaxique profonde valide

La solution au problème soulevé ici réside dans l'utilisation d'une contrainte que nous avons appliquée au parcours sémantique afin de respecter ce que nous avons appelé, en V:4.3, l'*imperméabilité des sous-réseaux* ***thème*** *et* ***rhème***. Selon cette contrainte, les sous-réseaux communicatifs doivent être parcourus à partir du ND et du chemin communicatif sur lequel ce ND se situe. Lorsqu'un nœud d'un sous-réseau communicatif est atteint d'une autre façon, le parcours du sous-réseau ne peut se poursuivre. Dans un cas comme celui de la figure VI-2, seul le nœud syntaxique profond construit lors du parcours sémantique du thème doit être considéré comme élément du thème syntaxique profond. L'autre occurrence co-référentielle du

même lexème n'est pas marquée communicativement car elle ne résulte pas de la phase de parcours du thème sémantique.

À la suite de ces observations, il est possible de formuler une règle assez générale de construction de la structure communicative syntaxique profonde:

> Un nœud de la RSyntP sera marqué comme faisant partie du thème syntaxique ou du rhème syntaxique si, dans le cadre du processus d'arborisation, il trouve son origine dans le parcours sémantique d'un sous-réseau **thème** ou **rhème**.

On peut ajouter ici une règle de bonne formation de la structure communicative syntaxique profonde:

> Les actants syntaxiques profonds d'un lexème héritent automatiquement du marquage communicatif de ce lexème.

Cette dernière règle a permis, dans notre exemple, de ne pas laisser à l'extérieur du rhème syntaxique la seconde occurrence du lexème LÉA.

2 TRAITEMENT D'UN EXEMPLE: "*Ces cinq ou six promenades, faites au bras de celui qu'elle idolâtrait en secret,...*"

Nous allons maintenant démontrer le fonctionnement de l'arborisation proposé ci-dessus à partir du traitement d'un exemple non trivial. Nous reprenons ici la citation d'E. Sue, déjà utilisée dans les chapitres précédents:

(5) *Ces cinq ou six promenades, faites au bras de celui qu'elle idolâtrait en secret, avaient été les seuls jours de bonheur qu'elle eût jamais connus.*

Nous avons déjà proposé (figure IV-5) une RSém pour cette phrase et montré comment une telle RSém pouvait être hiérarchisée. Cette RSém, une fois hiérarchisée, peut être considérée dans ce cas précis comme une RSémR dans la mesure où:

- aucune décomposition lexicale ne figurait dans la RSém — il n'y a donc pas de lexicalisation à opérer pour produire le Texte cible;
- pour simplifier la structure du réseau, le contenu sémantique grammaticalisable n'était pas intégré à la RSém — il n'y a donc pas non plus de grammaticalisation à opérer.

Pour faciliter la lecture de cette section, il est utile de faire figurer à nouveau ici la RSémR de (5) — c'est-à-dire, sa RSém hiérarchisée, grammaticalisée et lexicalisée:

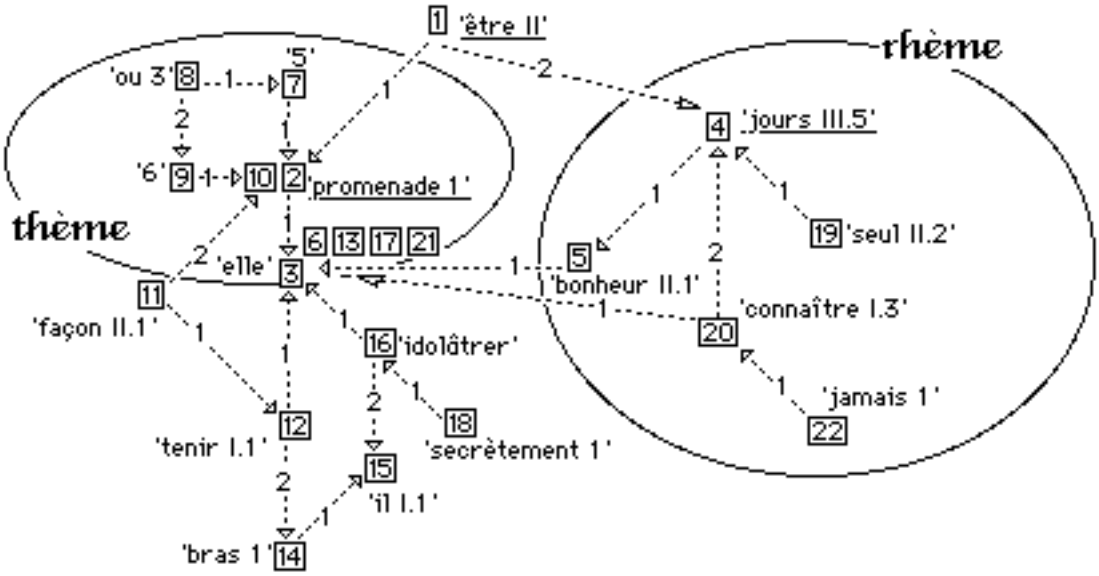


Figure VI-4 (rappel figure V-7)
RSémR de (5)

C'est cette représentation que nous allons maintenant arboriser afin de construire la RSyntP de (5).

Le processus de parcours du réseau, et donc d'arborisation, va être décomposé en cinq étapes qui permettent de décrire de façon satisfaisante la construction progressive de l'arbre syntaxique profond. Chacune de ces étapes, à l'exception de l'étape initiale, est conditionnée par la topographie particulière de la RSémR considérée. Le processus d'arborisation n'est pas, par nature, décomposable en un nombre fixe de phases successives. Si l'on considère une décomposition maximale de l'arborisation, il est possible de dire que ce processus se décompose en autant de phases successives qu'il y a d'arcs dans la RSémR. On peut alors conclure qu'à chaque étape du parcours sémantique de la RSémR correspond une étape dans l'arborisation. Il y a dans l'exemple que nous traitons ici vingt et une étapes minimales à considérer; nous ne mettrons cependant en relief que celles qui posent un problème original quant à la correspondance RSém ⇔ RSyntP mise en jeu.

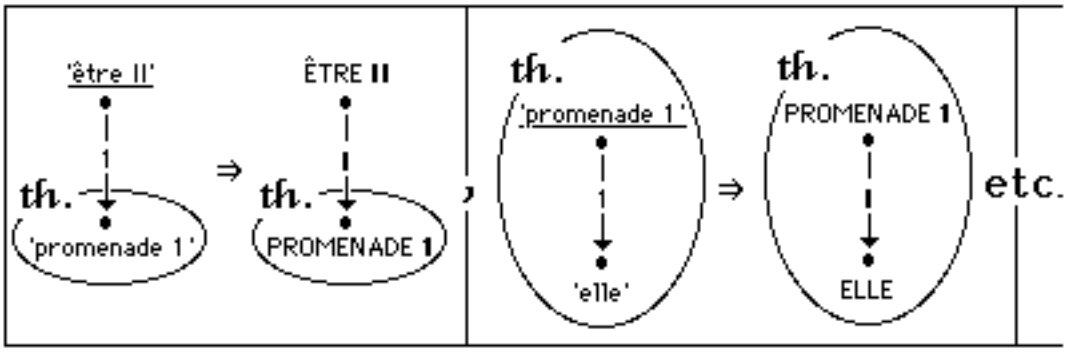
a) Parcours descendant à partir du ND 'être**II**'

Tout parcours sémantique commence par un parcours descendant à partir du ND sélectionné. On peut donc isoler, dans toutes les arborisations, une première étape consistant à:

- construire le sommet de l'arbre syntaxique profond, qui correspond au ND de la RSémR — ici, 'être**II**';
- construire récursivement toutes les relations actancielles syntaxiques dépendant du sommet de l'arbre.

Nous avons vu que le parcours descendant crée des configurations sémantiques où dépendances sémantique et communicative ont la même orientation. De plus, ce type de configuration sémantique est la plupart du temps mis en correspondance, par les règles de correspondance sémantique, avec des configurations syntaxiques profondes basées sur les dépendances actancielles **I**, **II**, **III**, **IV**, **V** et **VI** (cf. la présentation des règles de correspondance sémantique en

II:4). Dans notre exemple, la première étape du parcours sémantique va nous permettre, grâce aux règles de la grammaire, d'effectuer des arborisations du type:



Ces arborisations "atomiques" sont faites une à une, selon la progression du parcours sémantique. Elle permettent de construire du haut vers le bas, par rattachements successifs, le sous-arbre de la figure VI-5 qui est en quelque sorte l'épine dorsale de la RSyntP de (5). Il correspond grosso modo à la RSyntP de la phrase *"Ces promenades étaient des jours de bonheur"*.

Afin d'alléger les représentations syntaxiques données dans cette section nous utilisons la convention d'écriture suivante: les nœuds de l'arbre syntaxique profond contenus dans le **thème** sont marqués par un soulignement simple (par ex. PROMENADE**1**), ceux contenus dans le **rhème** sont marqués par un soulignement double (par ex. JOURS**III.5**).

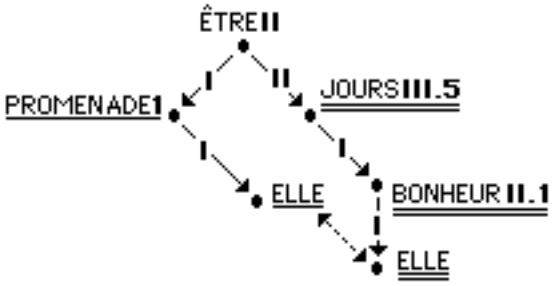


Figure VI-5
RSyntP de *"Ces promenades étaient des jours de bonheur."*

Pour pouvoir construire une telle figure, nous avons mis en pratique les techniques présentées dans la première section. On peut ainsi constater la création du nœud ELLE, co-référentiel à ELLE, qui trouve son origine dans le fait qu'un même nœud sémantique ('elle') est traversé plusieurs fois. Conformément à ce qui a été établi plus haut, seul le premier parcours de 'elle' confère au nœud syntaxique correspondant une valeur thématique dans la mesure où il est effectué pendant un parcours du thème sémantique. La seconde fois, ce nœud est atteint de l'extérieur du **thème** et sans utiliser le chemin communicatif sur lequel se trouve le ND. Conformément à la règle de bonne formation donnée plus haut, ELLE hérite sa valeur rhématique de BONHEUR**II.1** dont il est un actant syntaxique.

b) Parcours ascendant à l'intérieur du **thème** à partir du nœud thématique 'promenade**1**'

Ce deuxième parcours va provoquer des arborisations à partir de configurations sémantiques manifestant des dépendances sémantique et communicative opposées. Ceci explique l'apparition des relations **ATTR** et **COORD** (cf. II:4.2). La figure VI-6 donne le sous-arbre syntaxique profond construit lors de ce parcours ascendant effectué dans le **thème**; ce sous-arbre correspond à la RSyntP de *"cinq ou six promenades"*.

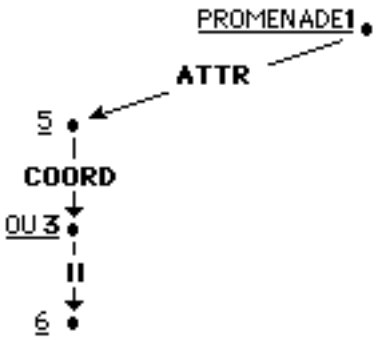


Figure VI-6
RSyntP de *"cinq ou six promenades"*

Tous les nœuds de ce sous-arbre sont des éléments du **thème** syntaxique profond puisqu'ils sont construits à partir de nœuds sémantiques thématisés et puisque cette construction tire son origine d'un parcours effectué à l'intérieur du **thème** sémantique.

Remarque: Il est nécessaire de disposer, dans la grammaire, d'une règle de correspondance sémantique permettant qu'une relation comme '6'—1→'promenade**1**' — le soulignement indique ici la dominance communicative locale — reçoive une traduction "vide" au niveau syntaxique profond.

c) Parcours ascendant à l'extérieur du **thème** à partir du nœud thématique 'promenade**1**'

Un autre parcours ascendant à partir de 'promenade**1**' peut être isolé. Il s'effectue à l'extérieur du **thème** et permet la construction de la RSyntP d'un syntagme comme:

(6) *des promenades faites en tenant son bras*

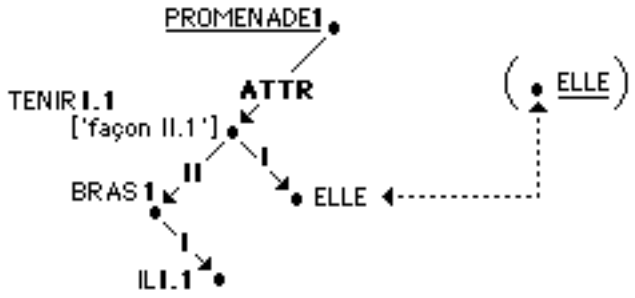


Figure VI-7
RSyntP de *"des promenades faites en tenant son bras"*

Remarque: La production d'un nœud syntaxique augmenté d'un trait de type sémantique — comme TENIR**I.1**['façon**II.1**'] — a été présentée dans l'Annexe du Chapitre II (Partie **1**); cf. la règle de correspondance sémantique proprement dite numérotée "VI".

La figure VI-7 ne fait pas apparaître les lexèmes qui sont véritablement mis en jeu dans le syntagme nominal qui lui est associé. Nous avons en quelque sorte ici la RSyntP d'un syntagme incorrect comme:

(7) **des promenades en tenant son bras*

Pour pouvoir effectuer la correspondance avec le niveau de représentation syntaxique de surface, il sera nécessaire ici de faire jouer des règles d'AJUSTEMENT de la RSyntP — comme celle présentée ci-dessous qui met en jeu la *fonction lexicale* $Oper_1$:

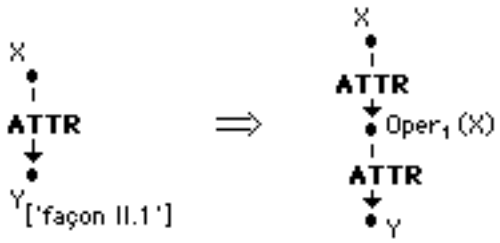


Figure VI-8
Exemple de règle d'ajustement de la RSyntP

Cette règle permet de prendre en compte, en plus du cas particulier de notre exemple, des syntagmes comme (8a) et (8b):

(8) a- *des coups*[X] *donnés*[$Oper_1$(X)] *avec acharnement*[Z]
b- *des pleurs*[X] *versés*[$Oper_1$(X)] *en se cachant*[Z] *les yeux*

Ce type de règle jouant exclusivement au niveau de la RSyntP, il n'a pas été inclus dans le champ de la présente étude.

d) Parcours ascendant à partir du nœud 'il**I.1**'

Le parcours entamé ci-dessus trouve sa conclusion dans un parcours ascendant effectué à partir du nœud 'il**I.1**'. On obtient ainsi une structure **ATTR**ibutive qui sera à l'origine de la subordonnée

relative *"(celui) qu'elle idolâtrait en secret"* dans les étapes postérieures du traitement.

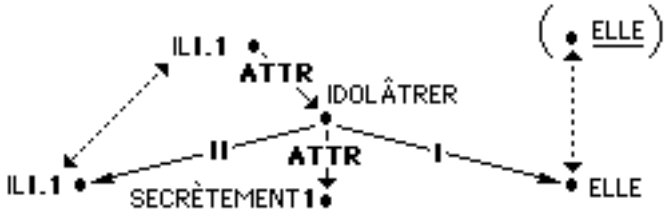


Figure VI-9
RSyntP de *"celui qu'elle idolâtrait en secret"*

Nous constatons ici l'apparition d'un lien de co-référence entre deux nœuds IL**I.1** qui tire son origine, non plus de plusieurs passages successifs par un même nœud lors du parcours sémantique, mais de l'activation de certaines règles particulières de correspondance sémantique. Ces règles ont été abondamment illustrées en II:4 et, surtout, dans l'Annexe du Chapitre II.

e) Parcours ascendant à l'intérieur du **rhème** à partir du nœud 'jours**III.5**'

Nous condensons en une seule étape tout le parcours ascendant effectué à l'intérieur du **rhème** car il ne pose pas de problème particulier par rapport aux cas qui ont déjà été examinés. La représentation obtenue ici correspond à *"les seuls jours qu'elle eût jamais connus"*.

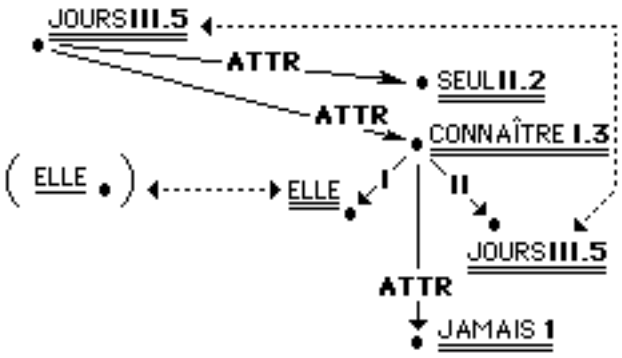


Figure VI-10
RSyntP de *"les seuls jours qu'elle eût jamais connus"*

La RSémR de la figure VI-4 est maintenant entièrement arborisée. Les cinq sous-arbres qui viennent d'être construits peuvent être connectés les uns aux autres pour former la RSyntP complète de (5):

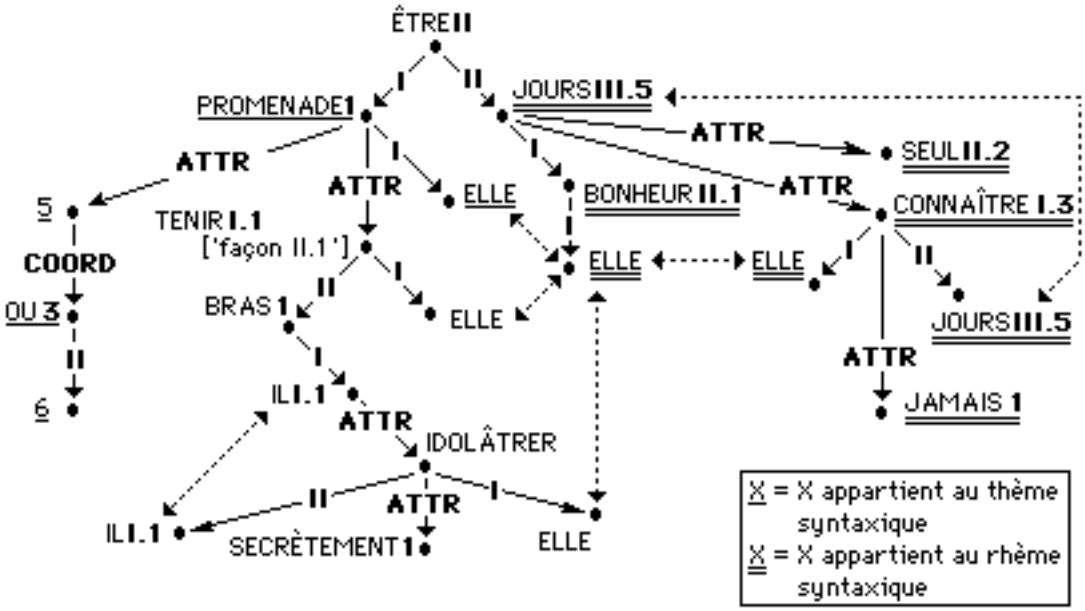


Figure VI-11
RSyntP complète de (5)

Pour que cette représentation soit vraiment complète, il faudrait notamment y faire figurer tous les marqueurs grammatico-sémantiques du type passé/présent, singulier/pluriel, défini/indéfini, etc. qui ne

sont pas pris en considération pour cet exemple (l'intervention des grammèmes dans la transition sémantique a été évoquée dans la Partie **1** en II:3.3). Suivant le modèle de la transition sémantique proposé dans cette étude, les marqueurs grammaticaux résultant de la grammaticalisation sont associés aux nœuds de la RSémR pour être ensuite transmis, lors de l'arborisation, aux nœuds syntaxiques profonds correspondants. La présence de marqueurs grammaticaux dans la RSémR pourrait être considérée comme un argument supplémentaire pour la caractérisation de cette représentation intermédiaire en tant que "RSém de Surface" (cf. la discussion en III:3).

Toutes les notions qui ont été étudiées dans les chapitres précédents interviennent plus ou moins directement dans le processus d'arborisation, notamment:

- les notions de dépendance sémantique, dépendance communicative et dépendance syntaxique — car elles sont mises en jeu dans les règles de correspondance sémantique proprement dites;
- les notions de ND et de parcours sémantique — car elles sont à la base des procédures de contrôle de l'activation de ces mêmes règles de correspondance;
- la notion de RSémR — car elle définit le domaine d'application de l'opération d'arborisation.

Ces notions n'ont pas été développées ou raffinées ici pour le plaisir de "théoriser", mais bien pour tenter de modéliser d'une façon CLAIRE et PRECISE le processus extrêmement complexe de transition sémantique.

Ceci termine notre présentation de l'arborisation et termine aussi cet exposé sur la transition sémantique. Nous allons maintenant conclure en récapitulant le contenu de la présente recherche et en proposant de possibles continuations.

CONCLUSION

L'étude que nous concluons maintenant visait à décrire l'activation de la composante sémantique des *Modèles Sens-Texte* (MST). Il s'agissait donc de définir un modèle dynamique d'utilisation des règles sémantiques de type Sens-Texte afin de mettre en œuvre la transition entre *Représentation Sémantique* et *Représentation Syntaxique Profonde* (RSém ⇒ RSyntP). Pour ce faire, nous avons tout d'abord redéfini le système de règles sémantiques déclaratives proposé pour les MST. Quatre types majeurs de règles ont été isolés:

1) les *règles de bonne formation*, qui définissent le formalisme des réseaux sémantiques et des arbres syntaxiques profonds;
2) les *règles de paraphrase sémantique*, qui sont basées sur les définitions lexicales et gèrent le niveau de décomposition des significations mises en jeu dans la RSém — dans un cadre de transition RSém ⇒ RSyntP, elles permettent de lexicaliser, c'est-à-dire de sélectionner les lexèmes profonds qui figureront effectivement dans la RSyntP;
3) les *règles grammatico-sémantiques*, qui mettent en correspondance des significations et des marqueurs grammaticaux de la langue — dans un cadre de transition RSém ⇒ RSyntP, elles isolent dans la RSém les significations réalisées par des moyens grammaticaux (cf. le temps verbal en français);
4) les *règles de correspondance sémantique proprement dites*, qui établissent les correspondances entre des configurations

sémantiques minimales et des configurations syntaxiques profondes.

Ce dernier type de règle est une innovation puisque, auparavant, les seules correspondances RSém ⇔ RSyntP envisagées par les MST étaient décrites au moyen des définitions lexicales elles-mêmes: mise en correspondance d'un sous-réseau sémantique (une définition) et d'un lexème (un nœud de la RSyntP). Nous avons démontré que cette approche n'était pas fonctionnelle car elle était en contradiction avec d'autres principes descriptifs de la *Théorie Sens-Texte* (TST). Nous avons donc appliqué au niveau de la correspondance RSém ⇔ RSyntP le principe qui est à la base de la construction des autres composantes des MST:

écriture d'un ensemble de règles déclaratives spécialisées prenant en charge la correspondance des micro-structures mises en jeu par deux niveaux de représentation adjacents.

Nous avons démontré que les configurations sémantiques prises en compte par les règles sémantiques d'un MST étaient basées sur un double système de dépendance:

- la *dépendance sémantique*, qui correspond à une relation prédicat-argument et est symbolisée au moyen d'un arc étiqueté par une relation actancielle sémantique;
- la *dépendance communicative*, qui correspond à un nouveau type de relation entre deux nœuds sémantiques et est symbolisée par le soulignement d'un nœud appelé *nœud dominant*.

La notion de dépendance communicative a été développée à partir de celle de *Sens générique* d'une définition lexicale. Nous avons montré que le Sens générique est en quelque sorte un Sens présenté comme premier, essentiel, par rapport au reste de la définition d'un lexème. Il

a ensuite été possible de montrer qu'il existait un lien entre la structure communicative des RSém et la notion de nœud générique ou dominant. Nous avons développé la notion de dépendance communicative afin de rendre compte de ce type de lien. Cette notion semble présenter un intérêt véritablement théorique puisqu'elle ne trouve pas seulement son application dans le cadre de l'activation procédurale des MST. Elle intervient aussi (i) au niveau du lexique, dans l'écriture des définitions lexicales, (ii) au niveau de la grammaire, dans l'écriture des règles grammatico-sémantiques et des règles de correspondance sémantique proprement dites.

La description des différents types de règles sémantiques nous a conduit à isoler trois opérations qui doivent être modélisées dans la transition sémantique:

1) la *lexicalisation*, application des règles lexico-sémantiques;
2) la *grammaticalisation*, application des règles grammatico-sémantiques;
3) l'*arborisation*, application des règles de correspondance sémantique proprement dites.

Ces trois opérations sont elles-mêmes dépendantes d'une quatrième que nous avons appelée la *hiérarchisation communicative* ou, plus simplement, la *hiérarchisation*. Elle consiste à calculer dans une RSém, à partir de l'analyse des structures sémantique et communicative, des liens de dépendance communicative entre les nœuds de cette représentation.

Il est vite devenu clair, dans le cours de cette recherche, que l'opération de hiérarchisation était la clé de l'activation des règles sémantiques Sens-Texte. Nous avons donc concentré notre étude du modèle dynamique sur les procédures permettant de calculer les liens de dépendance communicative à l'intérieur d'une RSém. La hiérarchisation a été décomposée en deux étapes:

1) la détermination d'un *nœud dominant* (ND) à partir de l'analyse des structures sémantique et communicative des RSém;
2) le parcours méthodique de la RSém à partir du ND.

Ce parcours, appelé *parcours sémantique*, établit les liens de dépendance communicative par étapes successives en traversant un à un tous les nœuds de la RSém. La pertinence de la hiérarchisation dépend donc de la pertinence des règles de choix du ND et des règles permettant de contrôler le parcours sémantique. Toutes ces règles sont basées sur l'analyse de ce que nous avons appelé la *topographie des réseaux sémantiques*.

Une dernière notion importante mise en jeu dans ce travail est celle de *Représentation Sémantique Réduite* (RSémR). La RSémR est un niveau intermédiaire de représentation postulé pour permettre de mieux mettre en œuvre la transition sémantique. Elle est formellement une sorte de RSém "traitée" pour être en relation directe avec une RSyntP donnée. Pour reprendre en la déformant une définition Sens-Texte, nous dirons que:

> Une RSémR est en quelque sorte une RSém qui a perdu sa raison d'être: qui ne représente plus un invariant de paraphrase mais une variante spécifique de paraphrase.

C'est la représentation sous forme de réseau d'une hiérarchie syntaxique profonde. Nous avons pu constater que ce type de représentation a été postulé dans d'autres recherches récentes sur les MST (cf. Beringer 1988 et St-Germain 1988). Selon notre approche, la RSémR est une RSém qui a été hiérarchisée, lexicalisée et grammaticalisée. Elle est la représentation formelle sur laquelle s'applique l'opération d'arborisation.

Voici donc, résumé brièvement, le contenu de cette recherche. Un aspect très intéressant de la transition sémantique qui n'a pu être

traité de façon satisfaisante est la lexicalisation. Nous voudrions, pour conclure définitivement, faire quelques remarques à ce propos.

La paraphrase prise en charge par les définitions de lexèmes ne modélise pas nécessairement une opération linguistique "naturelle". Il est aisé de montrer que le remplacement d'un lexème par sa définition, dans une phrase donnée, préserve peut-être le contenu sémantique exprimé mais enfreint aussi certaines lois de la communication. Ainsi, sorti d'un contexte d'étude linguistique, personne n'admettra que la phrase (2) est une paraphrase acceptable de (1), si "acceptable" signifie "qui pourrait être employée telle quelle dans un même contexte d'énonciation":

(1) *Iznogoud croit que je t'ai acheté ce tapis pour 14 maravédis.*
(2) *Iznogoud croit que j'ai acquis auprès de toi les droits de possession complète et permanente sur ce tapis en échange des 14 maravédis que tu demandes pour ce tapis.*

(2) ne devrait pas, pour ce qui est des usages linguistiques, être employée à la place de (1); pourtant, au niveau de la description linguistique, un MST doit pouvoir mettre ces deux phrases en relation de paraphrase.

Examinons maintenant deux autres paraphrases approximatives de (1):

(3) *Iznogoud croit que j'ai réalisé auprès de toi l'achat de ce tapis pour 14 maravédis.*
(4) *Iznogoud croit que tu m'a vendu ce tapis pour 14 maravédis.*

Ces deux paraphrases, basées sur une nominalisation dans le cas de (3) et sur l'utilisation d'un conversif dans le cas de (4), semblent, elles, assez naturelles. La lexicalisation n'est-elle donc pas une opération naturelle?

Les exemples de lexicalisation donnés dans la littérature Sens-Texte sont en général fondés sur une réduction d'une RSém initiale.

Ainsi, pour produire une phrase correcte, il ne semble jamais nécessaire de décomposer un Sens présent dans la RSém; il faut plutôt lexicaliser au maximum. D'une certaine façon, la lexicalisation apparaît ainsi comme une sorte d'artifice théorique et non comme la modélisation d'un véritable phénomène langagier. Pourquoi, alors, parler d'une opération de lexicalisation et comment justifier le recours aux règles de paraphrase sémantique?

Pour répondre à ces questions, remarquons tout d'abord que l'activité consistant à décomposer un Sens de lexème en Sens plus primitifs le constituant est fondée sur la capacité qu'a un locuteur idéal d'une langue donnée de construire une définition plus ou moins approximative d'un lexème par décomposition de son Sens. Un locuteur idéal doit donc être capable de construire une définition; cependant, lorsqu'il fait cela, ce n'est pas dans un contexte normal de communication mais bien EN PARLANT DE LA LANGUE ELLE-MEME. Par exemple, nous pourrons enseigner à un étranger le Sens du lexème ACHETER**1** en construisant une définition du type 'acquérir de qqn. les droits de possession complète et permanente de qqch. en échange d'une quantité d'argent que cette personne demande pour cette chose'. Pourtant, si nous n'employons pas dans une conversation le mot *acheter*, nous ne dirons pas pour autant *"acquérir les droits de possession..."* .

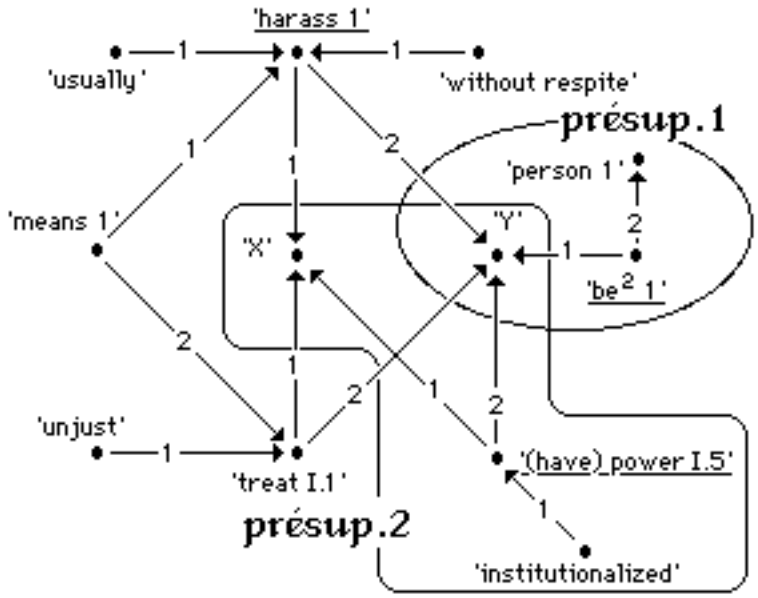
> La définition d'un lexème d'une langue **L** modélise donc une certaine connaissance linguistique que possède un locuteur idéal de **L**, mais l'utilisation de cette définition en tant que règle de paraphrase sémantique présente avant tout un intérêt théorique. La lexicalisation est une opération présupposée par les MST qui ne sert pas à établir des relations entre phrases dans le cadre d'une utilisation normale de la langue.

Notons aussi qu'une définition permet de mieux comprendre un lexème. En ce sens, dans un contexte de compréhension automatique

de texte, les définitions pourraient permettre d'interpréter le Sens de certains lexèmes par le biais des Sens plus primitifs le composant, lorsque aucune procédure directe d'interprétation n'a été prévue pour lui.

Finalement, en génération de texte (pour rester dans le domaine des applications), la lexicalisation utilise les présupposés d'une définition comme des contraintes impliquant une communication avec (i) d'autres données linguistiques, ou (ii) les modules conceptuels/de planification — ceci, en fonction de la nature des présupposés.

Soit la décomposition sémantique du lexème anglais PERSECUTE**1**, étudiée au Chapitre V:



Il est clair que **présup. 1** et **présup. 2** sont deux présupposés de nature très différente. DANS UN CONTEXTE DE GENERATION DE TEXTE, ils ne seront pas traités de la même façon. Il faut en effet considérer que ces présupposés peuvent être absents de la RSém sur laquelle doit s'effectuer la lexicalisation. Lorsque l'on cherche à matcher la définition sur une RSém, ils sont interprétés comme des conditions à

satisfaire pour que la lexicalisation ait lieu, chacun d'entre eux déclenchant un ensemble de procédures:

présup.1 va déclencher une analyse linguistique du Sens représenté ici par la variable 'Y'. Si, par exemple, 'Y' = 'pilot**I.1**', on voit que le Sens générique de la définition de 'pilot**I.1**' est 'person**1**' et que la "condition **présup.1**" est satisfaite.

L'interprétation de **présup.2** est plus complexe. Un système de génération de texte doit en effet avoir recours à des connaissances extra-linguistiques pour vérifier une proposition sémantique comme 'X has institutionalized power**I.5** over Y'. Il faudra vérifier (en fonction des données sur le Monde de référence) que l'on peut effectivement considérer que ce que dénote 'X' a un pouvoir institutionnel sur ce que dénote 'Y'. (Pour reprendre l'exemple que nous avons traité au Chapitre V, il est clair que ce type d'information ne se trouvera pas, au lexique, dans les définitions de PILOT**I.1** et POLICE**I**.)

Il est donc difficile d'isoler, dans le cas de la lexicalisation, un niveau de traitement purement linguistique. Les expérimentations sur cette opération doivent, à notre avis, s'insérer dans un contexte de recherche plus vaste que la simple activation des règles d'un MST.

La recherche que nous avons présentée ici s'est limitée à un champ d'étude plus spécifiquement linguistique. Nous pensons que la méthode générale proposée pour l'activation de la transition sémantique, ainsi que les notions qui l'accompagnent, aideront à une meilleure compréhension du processus complexe de synthèse linguistique. Nous pensons notamment que se trouvent ici certains éléments de réponse à la question que nous nous sommes toujours posée en examinant les différents travaux effectués en génération de texte:

D'où vient la syntaxe de la phrase?

Il existe en quelque sorte un consensus pour étiqueter les phénomènes et dire, par exemple, que la différence entre:

"Un chat chasse un chien chauve."
"Un chien qui est chassé par un chat est chauve."

réside en une différence de structure communicative. Mais comment une grammaire peut-elle prendre en compte une telle différence? Nous pensons avoir proposé, à partir de la modélisation Sens-Texte, une méthode formelle et intuitivement justifiée de contrôle de la production des énoncés en fonction d'informations sur leur structure communicative.

INDEX

analyse de texte **6**
arborisation **30**, **31**, **70**, **114**, **154**, **156**, **173**, **268**, **291**
BNF **59**
catégorie **95**
chemin **163**, **193**
chemin communicatif **192**, **193**, **251**, **252**
chemin énonciatif **260**
composante d'un MST **5**, **24**
composante morphologique de surface **24**
composante morphologique profonde **24**
composante phonologique profonde **24**
composante sémantique **24**
composante syntaxique de surface **24**
composante syntaxique profonde **24**
compréhension de texte **2**, **6**
concept **39**
DEC **76**
definiendum **77**
definiens **77**
dépendance communicative **84**, **114**, **121**, **133**, **177**, **191**, **290**
dépendance sémantique **121**, **290**
diathèse **86**, **89**
Dictionnaire Explicatif et Combinatoire **76**
dominer communicativement **221**

dynamisme communicatif **188**, **194**
fonction lexicale **61**, **64**, **88**, **152**, **192**, **285**
génération de texte **2**, **6**
gouverneur d'un nœud sémantique **218**
grammaticalisation **30**, **33**, **70**, **106**, **154**, **155**, **268**, **291**
grammème **95**
graphe conceptuel **47**
graphe orienté connecté **45**
hiérarchie communicative **121**, **171**, **173**
hiérarchie stricte **172**
hiérarchisation **30**, **31**, **70**, **122**, **154**, **157**, **190**, **268**, **291**
imperméabilité des sous-réseaux **thème** et **rhème** **228**, **257**, **258**, **278**
informatique linguistique **1**, **59**
Intelligence Artificielle **7**
interprétation de texte **6**
langagier **39**
lexème profond **61**
lexémisation **77**, **86**, **89**
lexicalisation **33**, **70**, **77**, **173**, **291**
lexicalisation au sens propre **30**, **77**, **86**, **89**, **154**
linguistique informatique **1**, **2**
linguistique procédurale **1**
linguistique systémique **17**
méthode heuristique **31**
méthode procédurale (ou algorithmique) **31**
Modèle Sens-Texte **4**, **289**
MST **4**
niveau conceptuel **65**
nœud dominant **32**, **46**, **81**, **158**, **165**, **171**, **177**, **190**, **290**, **292**
paraphrase **39**
paraphrase cognitive **38**, **202**
paraphrase langagière **39**, **202**
parcours **163**, **191**
parcours ascendant **217**

parcours descendant **217**
parcours sémantique **31**, **32**, **162**, **164**, **171**, **215**, **292**
pouvoir paraphrastique des MST **271**
présupposition **44**, **233**
présupposition de lexème **237**, **239**, **241**, **248**, **239**
présupposition de phrase **237**
règle de correspondance sémantique **113**
règle de correspondance sémantique proprement dite **29**, **70**, **113**, **134**, **156**, **169**, **268**, **289**
règle de paraphrase lexicale **153**
règle de paraphrase syntaxique **153**
règle de paraphrase sémantique **29**, **70**, **76**, **154**, **289**
règle grammatico-sémantique **29**, **70**, **94**, **155**, **168**, **289**
règle lexico-sémantique **86**
règle sémantique de bonne formation **29**, **70**, **198**, **289**
Représentation Morphologique de Surface **20**
Représentation Morphologique Profonde **20**
Représentation Phonologique de Surface **20**, **40**
Représentation Phonologique Profonde **20**
Représentation Sémantique **5**, **20**, **289**
Représentation Sémantique Réduite **93**, **165**, **292**
Représentation Syntaxique de Surface **20**
Représentation Syntaxique Profonde **5**, **20**, **289**
réseau **45**
réseau de systèmes **21**
réseau sémantique réduit **32**
rhème **44**, **189**
rhème second **128**
RPhonS **40**
RSém **5**, **39**
RSémR **93**, **160**, **165**
RSyntP **5**
Sens générique **79**, **81**, **290**
signe linguistique **24**, **25**
signifiant **25**

signification dérivationnelle **95**
signification flexionnelle **96**
signification grammaticale **94**
signification lexicale **94**
signifié **25**
structure communicative **44**
syntaxe de dépendance **57**
synthèse de texte **6**
système **20**, **21**
Théorie Sens-Texte **4**, **290**
thème **44**, **189**
topographie **34**
topographie des RSém **33**, **71**, **171**, **190**, **213**, **239**, **292**
trait **20**
Traitement Automatique de la Langue Naturelle **2**
transition **24**, **26**
transition sémantique **26**
transition (synthétique) RSém ⇒ RSyntP **26**
translation **101**, **102**
traverser (un nœud sémantique) **163**
TST **4**
union linguistique **24**, **25**

BIBLIOGRAPHIE

Alonso, C.H. (1984) *Gramática Functional del Espa4nol*. Biblioteca Románica Hispánica. Gredos, Madrid.

Appelt, D.E. (1985) *Planning English Sentences.* Cambridge University Press, Cambridge, G.B.

Apresjan, J.D. (1980) Tipy informacii dlja poverxnostno-semanti2ceskogo komponenta modeli Smysl ⇔ Tekst [Types d'informations de la composante sémantique de surface du modèle Sens ⇔ Texte]. *Wiener Slawistischer Almanach*, Sonderband 1, Vienne.

Apresjan, J.D., Boguslavskij, I.M. & Iomdin, L.L. (1984 et 1985) *Lingvističeskoe obespečenie sistemy francuzsko-russkogo avtomatičeskogo perevoda ETAP-1 (III, VI)* [Module linguistique du système ETAP-I de traduction automatique français-russe (III, VI)]. Publications de l'Institut de la langue russe de l'Académie des Sciences de l'URSS, no. 155 et 166, Moscou.

Apresjan, J.D., Boguslavskij, I.M., Iomdin, L.L., Lazurskij, A.V., Pertsov, N.V., Sannikov, V.Z. & Cinman, L.L. (1989) *Lingvističeskoje obespečenie sistemy ETAP-2* [Module linguistique du système ETAP-2]. Nauka, Moscou.

Beringer, H. (1988) *Traitement automatique des ambiguïtés structurales du français utilisant la Théorie Sens-Texte.* Thèse de Doctorat, Université de Paris 6.

Boguslavskij, I.M. (1985) *Issledovanija po sintaksičeskoj semantike* [Études de sémantique syntaxique]. Nauka, Moscou.

Bourbeau, L., Carcagno, D., Kittredge, R.I. & Polguère, A. (1989) English and French Marine Forecasts Generated by the FoG-89 System. Corpus compilé par ORA-Montréal pour Environnement Canada.

Bourbeau, L., Carcagno, D., Goldberg, E., Kittredge, R.I. & Polguère, A. (à paraître) Bilingual Generation of Weather Forecasts in an Operations Environment. COLING-90, Helsinki.

Boyer, M. (1984) Génération de phrases à partir de réseaux sémantiques. Document de travail 152, Département d'Informatique et de Recherche Opérationnelle, Université de Montréal.

Boyer, M. & Lapalme, G. (1985) Generating Paraphrases from Meaning-Text Semantic Networks. *Computational Intelligence*, I:3-4, 103-117.

Clocksin, W.F. & Mellish, C.S. (1981) *Programming in Prolog*. Springer-Verlag, Berlin-Heidelberg-New York.

Carcagno, D. & Iordanskaja, L. (1989) Content determination and text structuring in GOSSIP. *Extended Abstracts presented at the Second European Natural Language Generation Workshop*, 6-8 avril 1989, Edimburgh, 15-22.

Daneš, F. (1974) Functional Sentence Perspective and the Organization of the Text. In F.Daneš (ed.): *Papers on Functional Sentence Perspective*. Academia, Prague, 106-128.

Danlos, L. (1985) *Génération automatique de textes en langues naturelles.* Masson, Paris.

Davidse, K. (1987) M.A.K. Halliday's Functional Grammar and the Prague School. In R.Dirven & V.Fried (eds.): *Functionalism in Linguistics*. LLSEE, XX, John Benjamins, Amsterdam/Philadelphia, 39-79.

Décary, M. (1986) Un éditeur spécialisé pour le Dictionnaire Explicatif et Combinatoire du Français Contemporain. Mémoire de maîtrise, Département d'Informatique et de Recherche Opérationnelle, Université de Montréal.

Denning, D. (1985) Requirements and Model for IDES - A Real-Time Intrusion-Detection Expert System. Final Report SRI Project 6169.

Dik, S.C. (1978) *Functional Grammar*. North Holland, Amsterdam.

Dowty, D.R., Wall, R.E. & Peters, S. (1981) *Introduction to Montague Semantics*. Reidel, Dordrecht.

Ducrot, O. (1972) *Dire et ne pas dire*. Deuxième édition, janvier 1980, Hermann, Paris.

Fillmore, C.J. (1971) Verbs of judging: an exercise in semantic description. In C.J.Fillmore and D.T.Langendoen (eds.): *Studies in Linguistic Semantics*, Holt, Rinehart and Winston, New-York *et al.*, 272-289.

Firbas, J. (1964) On Defining the Theme in Functional Sentence Analysis. *Travaux Linguistiques de Prague*, I, University of Alabama Press, 267-280.

Firbas, J. (1965) A Note on Transition Proper in Functional Sentence Analysis. *Philologica Pragensia*, VIII, Prague, 170-176.

Firbas, J. (1974) Some Aspects of the Czechoslovak Approach to Problems of Functional Sentence Perspective. In F.Daneš (ed.): *Papers on Functional Sentence Perspective*. Academia, Prague, 11-37.

Firbas, J. (1987) On the Delimitation of the Theme in Functional Sentence Perspective. In R.Dirven & V.Fried (eds.): *Functionalism in Linguistics*. LLSEE, XX, John Benjamins, Amsterdam/Philadelphia, 137-156.

Firth, J.R. (1957) A synopsis of linguistic theory, 1930-55. In F.R.Palmer (ed.): *Selected Papers of J.R. Firth.* Indiana University Press, Bloomington & London, 168-205.

Frederiksen, C.H. (1975) Representing Logical and Semantic Structure of Knowledge Acquired from Discourse. *Cognitive Psychology*, VII, 371-458.

Furukawa, N. (1986) *L'article et le problème de la référence en français*. France Tosho, Tokyo.

Garrett, M.F. (1975) The analysis of sentence production. In G.Brown (ed.): *Psychology of learning and motivation*, IX, Academic Press, New York.

Gebruers, R. (1987) S.C. Dik's Functional Grammar: A Pilgrimage to Prague? In R.Dirven & V.Fried (eds.): *Functionalism in Linguistics*. LLSEE, XX, John Benjamins, Amsterdam/Philadelphia, 101-134.

Grevisse, M. (1975) *Le Bon Usage.* Dixième édition revue, Duculot, Gembloux, Belgique.

Gladkij, A.V. (1985) *Sintaksičeskie struktury estestvennogo jazyka v avtomatizirovannyx sistemax obščenija* [Structures syntaxiques des langues naturelles dans les systèmes automatisés de communication]. Nauka, Moscou.

Guillaume, G. (1969) *Langage et science du langage*. Nizet, Paris.

Halliday, M.A.K (1961) Categories of the Theory of Grammar. *Word*, XVII, 241-292.

Halliday, M.A.K. (1974) The Place of "Functional Sentence Perspective" in the System of Linguistic Description. In F.Daneš (ed.): *Papers on Functional Sentence Perspective*. Academia, Prague, 106-128.

Halliday, M.A.K. (1978) *Language as Social Semiotic.* Edward Arnold, London.

Halliday, M.A.K. (1985) *An Introduction to Functional Grammar.* Edward Arnold, London.

Halliday, M.A.K. & Martin, J.R. (eds.) (1981) *Readings in Systemic Linguistics*. Batsford Academic, London.

Harary, F., Norman, R.Z. & Cartwright, D. (1965) *Structural models: An introduction to the theory of directed graphs.* John Wiley & Sons, New York.

Hawkins, J.A. (1978) *Definiteness and Indefiniteness*. Croom Helm, London.

Herman, P., Sabah, G. & Vilnat, A. (1988) A question-answering system for the French Yellow Pages. *Computational Intelligence*, IV:2, 192-204.

Iordanskaja, L. (1986) Propriétés sémantiques des verbes promoteurs de la négation en français. *Lingvisticæ Investigationes*, X:2, 345-380.

Iordanskaja, L. (1989) Communicative structure and its use during text generation. Rapport interne ORA-Montréal.

Iordanskaja, L., Kittredge, R.I. & Polguère, A. (1988) Implementing a Meaning-Text Model for Language Generation. Supplément à *Proceedings of the 12th. International Conference on Computational Linguistics (COLING-88)*, 22-27 août 1988, Budapest.

Iordanskaja, L., & Polguère, A. (1988) Semantic Processing for Text Generation. *Proceedings of the First International Computer Science Conference-88*, 19-21 décembre 1988, Hong Kong, 310-318.

Jacobs, P.S. (1987) Knowledge-Intensive Natural Language Generation. *Artificial Intelligence*, XXXIII, 325-378.

Kittredge, R.I. & Lerhberger, J. (eds.) (1982) *Sublanguage: Studies of language in restricted semantic domains.* de Gruyter, Berlin.

Krámský, J. (1972) *The Article and the Concept of Definiteness in Language*. Mouton, The Hague.

Lamb, S.M. (1966) *Outline of Stratificational Grammar.* Georgetown University Press, Washington.

Landais, N. (1850) *Grammaire générale des grammaires françaises.* Didier, Paris.

Mann, W.C & Matthiessen, C.M.I.M (1983) Nigel: A Systemic Grammar for Text Generation. Information Sciences Institute RR-83-105, University of Southern California, Marina del Rey, CA.

Mathesius, V. (1939) O tak zvaném aktuálním členění větném. *Slovo a Slovcesnost*, 5:171-174. [Traduction anglaise: On the Information-Bearing Structure of the Sentence, In: S.Kuno (ed.) (1975), *Harvard Studies in Syntax and Semantics*: 467-480.]

McDonald, D.D. (1980) *Natural Language Production as a Process of Decision-Making under Constraints.* Thèse de Ph.D., M.I.T, Cambridge, Mass.

McKeown, K. (1982) *Generating Natural Language Text in Response to Questions about Database Structure.* Thèse de Ph.D., University of Pennsylvania, Philadelphia, Pa.

McLeod, J.C. (à paraître) The Forecast Production Assistant. À paraître dans *Proceedings of the Sixth International Conference on Interactive Information Processing Systems for Meteorology, Oceanography and Hydrology*, février 1990, Anaheim, CA.

McNeill, D. (1979) *The Conceptual Basis of Language.* Lawrence Erlbaum, Hillsdale, New Jersey.

Mel'čuk, I.A. (1974) *Opyt teorii lingvističeskix modelej "Smysl ⇔ Tekst"* [Vers une Théorie des Modèles Linguistiques Sens-Texte]. Nauka, Moscou.

Mel'čuk, I.A. (1978) Théorie du langage, théorie de traduction. *Meta*, XXIII:4, 271-302.

Mel'čuk, I.A. (1981a) Meaning-Text Models: A Recent Trend in Soviet Linguistics. *Annual Review of Anthropology*, X, 27-62.

Mel'čuk, I.A. (1981b) Types de Dépendance Syntagmatique entre les Mots-Formes d'une Phrase. *Bulletin de la Société de Linguistique de Paris*, LXXVI:1, 1-59.

Mel'čuk, I.A. (1982) *Towards a Language of Linguistics*. Wilhelm Fink Verlag, Munich.

Mel'čuk, I.A. (1988a) *Dependency Syntax: Theory and Practice*. State University of New York Press, New York.

Mel'čuk, I.A. (1988b) Paraphrase et lexique dans la théorie Sens-Texte. *Lexique*, VI, Presses Universitaires de Lille, Lille, 13-54.

Mel'čuk, I.A. (1988c) Paraphrase et lexique dans la théorie Sens-Texte (1ère. partie). *Cahiers de lexicologie*, LII, 1988-1, Didier, Paris, 5-50.

Mel'čuk, I.A. (1988d) Paraphrase et lexique dans la théorie Sens-Texte (2e. partie). *Cahiers de lexicologie*, LIII, 1988-2, Didier, Paris, 5-53.

Mel'čuk, I.A. (1989) Semantic Primitives from the Viewpoint of the Meaning-Text Linguistic Theory. *Quaderni di Semantica*, X:1, 65-102.

Mel'čuk, I.A. (à paraître) *Cours de morphologie générale (théorique et descriptive).* Presses de l'Université de Montréal — Editions Peeters.

Mel'čuk, I.A., Arbatchewsky-Jumarie, N., Elnitsky, L., Iordanskaja, L. & Lessard, A. (1984) *Dictionnaire explicatif et combinatoire du français contemporain - Recherches lexico-sémantiques I.* Presses de l'Université de Montréal, Montréal.

Mel'čuk, I.A., Arbatchewsky-Jumarie, N., Dagenais, L., Elnitsky, L., Iordanskaja, L., Lefebvre, M.-N. & Mantha, S. (1988) *Dictionnaire explicatif et combinatoire du français contemporain - Recherches lexico-sémantiques II.* Presses de l'Université de Montréal, Montréal.

Mel'čuk, I.A. & Pertsov, N.V. (1987) *Surface Syntax of English (A Formal Model within the Meaning-Text Framework).* John Benjamins, Amsterdam/Philadelphia.

Mel'čuk, I.A. & Polguère, A. (1987) A Formal Lexicon in the Meaning-Text Theory (or How to Do Lexica with Words). *Computational Linguistics*, XII:3-4, 261-275.

Mel'čuk, I.A. & Polguère, A. (à paraître) Aspects of the Implementation of the Meaning-Text Model for English Text Generation. Présenté à "The Dynamic Text Conference", Université de Toronto, 5-10 juin 1989.

Montague, R. (1973) The proper treatment of quantification in ordinary English. In R.H. Thomason (ed.) (1974): *Formal philosophy - selected papers of Richard Montague*. Yale University Press, Yale.

Mørdrup, O. (1975) Présuppositions, implications et verbes français. *Revue Romane*, X:1, 125-155.

Morgan, J.L. & Sellner, M.B. (1980) Discourse and Linguistic Theory. In R.J.Spiro, B.C.Bruce & W.F.Brewer (eds.): *Theoretical Issues in Reading Comprehension*, Lawrence Erlbaum, Hillsdale, New Jersey.

Padučeva, E.V. (1985) *Vyskazyvanie i ego sootnesënnost' s dejstvitel'-nost'ju* [L'énoncé et sa corrélation avec la réalité]. Nauka, Moscou.

Patten, T. (1988) *Systemic text generation as problem solving.* Cambridge University Press, Cambridge, G.B.

Penman Natural Language Group (1988) The NIGEL Manual (Draft). Information Sciences Institute, University of Southern California, Marina del Rey, CA.

Pereira, F. (1983) Logic for Natural Language Analysis. SRI International Technical Note 275.

Polguère, A. (1988a) GOSSIP Project - The Meaning-Text Component of GOSSIP (Description of the programs). Rapport technique ORA-Montréal.

Polguère, A. (1988b) A Dependency Grammar for Marine Weather Forecasts. Deliverable II - Technical Report for Environment Canada, ORA-Montréal.

Polguère, A. (1989) Grammatical and lexical formalisms in FoG-89. Part of the system documentation for Final Report for Environment Canada, ORA-Montréal.

Quillian, M.R. (1968) Semantic Memory. In M.Minsky (ed.): *Semantic Information Processing*. The MIT Press, Cambridge, Mass., 227-270.

Schank, R.C. (1975) *Conceptual Information Processing*. North-Holland.

Schank, R.C. (1987) What Is AI, Anyway? *AI Magazine*, Winter 1987, 59-65.

Schubert, K. (1987) *METATAXIS - Contrastive dependency syntax for machine translation.* Foris Publications, Dordrecht/Providence.

Sgall, P. (1987) Prague Functionalism and Topic *vs.* Focus. In R.Dirven & V.Fried (eds.): *Functionalism in Linguistics*. LLSEE, XX, John Benjamins, Amsterdam/Philadelphia, 169-189.

Sowa, J.F. (1968) *Conceptual Structures: A Model for Language.* [Manuscrit non publié, référence trouvée dans Sowa 1984.]

Sowa, J.F. (1984) *Conceptual Structures: Information Processing in Mind and Machine.* Addison-Wesley, Reading, Mass.

Sowa, J.F. (à paraître) Multi-Domain Semantic Theory. Brouillon daté du 28 novembre 1988.

St-Germain, J. (1988) *Étude sémantico-syntaxique de la causation et de la causativité dans les verbes français.* Mémoire de maîtrise, Université de Montréal, Montréal.

Strawson, P.F. (1950) On referring. *Mind*, LIX:235, 320-344.

Strawson, P.F. (1964) Indentifying reference and truth-values. *Theoria*, XXX, 96-118.

Tesnière, L. (1959) *Éléments de syntaxe structurale*. Klincksieck, Paris.

Winograd, T. (1983) *Language as a Cognitive Process – Volume I: Syntax*. Addison-Wesley, Reading, Mass.

Dictionnaires

Le Petit Robert (1981). Le Robert, Paris.

Longman Dictionary of Contemporary English (1978). Longman, Harlow & London.

Œuvres littéraires mentionnées

Hergé (1953) *Objectif Lune.* Casterman, Tournai, Belgique.

Rostand, E. (1985) *Cyrano de Bergerac.* "Classiques Larousse", Librairie Larousse, Paris.

Sue, E. (1845) *Le juif errant.* Robert Laffont, Paris (Réédition 1983).